Advances in Civil Engineering

Machine Learning-Aided Design of Concrete Mixtures

Lead Guest Editor: Hakan Basarir
Guest Editors: Junfei Zhang, Yuantian Sun, Baris Binici, Amir H. Alavi, and Mohamed Elchalakani

Machine Learning-Aided Design of Concrete Mixtures

Advances in Civil Engineering

Machine Learning-Aided Design of Concrete Mixtures

Lead Guest Editor: Hakan Basarir
Guest Editors: Junfei Zhang, Yuantian Sun, Baris Binici, Amir H. Alavi, and Mohamed Elchalakani

Contents

Hindawi
Advances in Civil Engineering
Volume 2021, Article ID 6629466, 10 pages
https://doi.org/10.1155/2021/6629466

Research Article

Estimating the Compressive Strength of Cement-Based Materials with Mining Waste Using Support Vector Machine, Decision Tree, and Random Forest Models

Hongxia Ma,[1] Jiandong Liu,[2] Jia Zhang,[2] and Jiandong Huang[2]

[1]*Jiangsu Province Xuzhou Technician Institute, Xuzhou 221116, China*
[2]*School of Mines, China University of Mining and Technology, Xuzhou 221116, China*

Correspondence should be addressed to Jiandong Huang; huang@cumt.edu.cn

Received 29 October 2020; Accepted 2 August 2021; Published 7 August 2021

Academic Editor: Junfei Zhang

To estimate the compressive strength of cement-based materials with mining waste, the dataset based on a series of experimental studies was constructed. The support vector machine (SVM), decision tree (DT), and random forest (RF) models were developed and compared. The beetle antennae search (BAS) algorithm was employed to tune the hyperparameters of the developed machine learning models. The predictive performances of the three models were compared by the evaluation of the values of correlation coefficient (R) and root mean square error (RMSE). The results showed that the BAS algorithm can effectively tune these artificial intelligence models. The SVM model can obtain the minimum RMSE, while the BAS algorithm is inefficient in DT and RF models. The SVM, DT, and RF models can be used to predict the compressive strength of cement-based materials using solid mining waste as aggregate effectively and accurately, with high R values and lower RMSE values. The RF algorithm can obtain the highest value of R and the lowest value of RMSE, demonstrating the highest accuracy. The solid mining waste to cement ratio is the most important variable to affect the compressive strength. Curing time was also an important parameter in the compressive strength of cemented materials, followed by the water-solid ratio of mining waste and fine sand ratio.

1. Introduction

The mining waste includes the stripping and waste rock (including coal gangue) produced in the process of mining, as well as the tailings discarded in the process of mineral dressing [1–6]. A large amount of mining solid waste as well as the complicated treatment has become one of the issues that should be addressed for environmental protection [4, 7–12]. The massive accumulation of mining waste can cause pollution of the land, causing disasters such as landslides and mudslides. The debris and tailings formed by the weathering of waste rock are either washed by water into the water body, dissolved and penetrated into the groundwater, or blown into the atmosphere by the wind, polluting the environment with water and gas as the medium [13–16]. These mining waste materials not only take up a lot of land but also directly pollute the environment and threaten the safety of people's lives, since some of these wastes contain highly toxic elements, such as arsenic and cadmium, and radioactive elements, which are harmful to human health [1, 2, 4, 7–9]. Tailings have the characteristics of fine particles, small weight, large surface area, being easy to run away in contact with water, and being easy to fly when exposed to wind. Therefore, those solid mining wastes are a potential hazard to the air, water bodies, farmland, and villages [4, 13, 17–20].

To address such issue of solid mining waste, researchers have started various approaches to recycling and reproduction of those construction materials using the mining waste, and one of the most representative cases is the widely used cement-based materials which can be produced environmentally friendly if the main aggregates can be replaced by the mining waste [21–31]. In general, cement and water should be added to cement-based composites to increase

strength and fluidity. After mixing with water, the slurry can be hardened in the water, and the sand, stone, and other materials can be firmly cemented together [32–40]. Due to its important role as a building material, it has received more and more attention in recent years, especially in the research objectives of strength and durability [41–49].

As mentioned above, compressive strength is a key index for evaluating the construction process and workability of cement-based composites, so most studies focus on experiments and revealing mechanisms [50–57]. Typically, the strength of cement-based materials is tested in the laboratory and the practical civil engineer finds the optimized proportion of the constituent materials in the composition [58–66]. However, experimental research is time-consuming and expensive and is not suitable for a large number of laboratory tests [41, 42, 67–72]. Some researchers have applied empirical formulas to predict the strength of cement-based materials using the so-called mathematical regression, in which some of the empirical formulas are useful, but most of them are limited by the input and output dataset. Because of the low accuracy of the empirical formula, its scope of application has barely been realized. Hence, to accurately predict the compressive strength of cement-based materials, simpler and more reliable models should be proposed.

Using artificial intelligence methods to evaluate and predict the mechanical properties of cement-based materials is a hot spot in the research field of cement-based composites [35, 39, 63]. Besides, some researchers have applied machine learning methods to evaluate the performance of cement-based composites. For example, artificial neural networks (ANN) and other evolutionary optimization algorithms have been used to predict the compressive strength of cement-based materials. Jamal et al. analyzed the potential to use multiple linear regression and adaptive neurofuzzy reasoning systems for the estimation of the strength of recycled aggregate concrete [73]. Khademi and Jamal used the same method in a follow-up study to predict the compressive strength of the recycled aggregate concrete after curing for 28 days [74]. The measured results of the above-mentioned method were consistent with the predicted ones, indicating that the machine learning techniques are feasible for the performance prediction of the cemented-based materials. However, limited studies can support accurate and efficient artificial intelligence techniques that specifically target the compressive strength of the cemented-based materials that use the mining waste as the aggregates. Also, it should be noted that the above machine learning techniques have been successfully adapted to the prediction of the concrete materials, but these studies still have the limitations of uncertainty, being time-consuming, and low efficiency. Therefore, it is necessary to propose more efficient and simple machine learning techniques to evaluate the compressive strength of the cemented-based materials that use the mining waste as the aggregates. Besides, more intelligent algorithms such as random forest, support vector machine, and decision tree are also applied in this field. It should be noted that most machine learning models can make predictions accurately, but there is limited research on the impact of different models on the accuracy of predictions. The above algorithms have different sensitivities to datasets with different attributes. Therefore, it is necessary to use different models for predictive compression and to find the optimal model for future applications.

In this study, various experimental tests were conducted to construct the dataset considering four key factors, that is, solid mining waste to cement ratio, fine sand ratio, solid mining waste to water ratio, and curing time. Then three machine learning models, that is, SVM, DT, and RF, were used for the prediction of the strength of cement-based materials using the mining waste as the aggregates, and the hyperparameters of these models were tuned by beetle antennae search (BAS) algorithm. Last, the prediction results were compared and analyzed, which give the best models in compressive strength prediction of the cement-based materials. Also, the importance of influencing variables was conducted and the most sensitive factor was found. The study provides the best models for strength prediction of cement-based materials, which is an important guideline in this research field.

2. Methodology

2.1. Dataset Collection. In the present study, the compressive strength dataset used for the prediction of cement-based materials was collected from the previous studies by the authors. The tailings were employed as the solid mining waste for the replacement of the aggregates. Four parameters confirmed to be the influenced variables for the compressive strength of the cement-based materials were determined as the input: the ratio of the solid mining waste to the cement, fine sand ratio, the ratio of the solid mining waste to water, and the curing time. Table 1 presents the composition of the dataset.

The solid mining waste to the cement ratio was determined as 4, 6, 8, and 10, respectively. The fine sand ratio concerning the total mixture was selected as 0, 0.1, 0.15, and 0.2. The solid mining waste to water ratio was determined as 0.68, 0.7, and 0.72, respectively. The concrete samples were cured for 7, 28, and 60 days. Table 2 shows the statistical analysis of these variables in the dataset.

The Pearson correlation coefficient was employed in this study to evaluate the relationship between the input parameters. The Pearson correlation coefficient between two input variables is defined as the quotient of covariance and standard deviation between two variables, as shown in the following equation:

$$\rho_{XY} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E\left[\left(X-\mu_X\right)\left(Y-\mu_Y\right)\right]}{\sigma_X \sigma_Y}. \quad (1)$$

Correspondingly, the correlation of the collected dataset was analyzed by SPSS software in the present study. Figure 1 gives the correlation matrix between the input variables (the solid mining waste to cement ratio, fine sand ratio, solid mining waste to water ratio, and curing time).

It can be found from Figure 1 that the correlation between the two same variables is 1 on the diagonal from the bottom left

Table 1: Composition of the dataset.

Solid mining waste to cement ratio	Fine sand ratio	Solid mining waste to water ratio	Curing time
4, 6, 8, and 10	0, 0.1, 0.15, 0.2	0.68, 0.70, 0.72	7, 28, 60 days

Table 2: Statistical analysis of these variables in the dataset.

Dataset	Variables	Mean	Std
Training dataset (the number of data points is 100)	Solid mining waste to cement ratio (%)	70.14	1.61
	Fine sand ratio (%)	11.15	7.24
	Solid mining waste to water ratio (%)	6.92	2.16
	Curing time (d)	33.53	21.14
Testing dataset (the number of data points is 44)	Solid mining waste to cement ratio	69.68	1.65
	Fine sand ratio	11.48	7.73
	Solid mining waste to water ratio	7.18	2.39
	Curing time	27.43	22.64

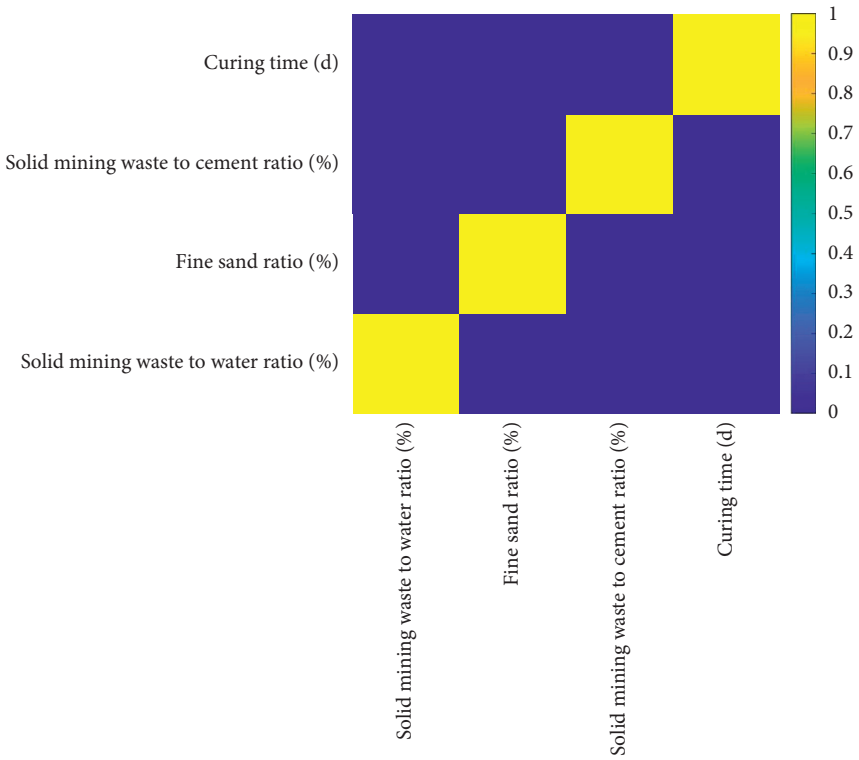


Figure 1: Correlation matrix of the input matrix.

to the top right and the correlation coefficient of the part above the diagonal is symmetric with the correlation coefficient of the part below the diagonal. The correlation coefficients between the different variables are relatively low (most values are close to 0.1). This indicates that these input variables are independent of each other, so they can be used as input variables for intelligent prediction of compressive strength of the concrete samples without causing multicollinearity issues.

2.2. Artificial Intelligence Algorithms. In the present study, three artificial intelligence algorithms, SVM, DT, and RF, were employed to predict the compressive strength of concrete samples with tailings as aggregate.

Support vector regression (SVR) seeks an estimation indicator function, which can be used to classify test samples [60]. By extending the problem from seeking indication function estimation to seeking real-valued function estimation, a support vector machine (SVM) for function estimation (regression) can be obtained [75–77]. SVM effectively solves the problems of a small number of samples, high dimension, and nonlinearity [78, 79]. However, as a new machine learning algorithm, there are still some areas to be improved, and the selection of its parameters (including error ε, error penalty factor C, and kernel function parameters y) is one of the problems to be improved [60, 80, 81]. The kernel function parameter Y affects the complexity of the distribution of sample data in the high-

dimensional feature space. The change of kernel parameters implicitly changes the mapping function, thus changing the dimension of the sample space. For an indicator function set, if there are H samples that can be separated by all possible forms of two to H of the function set, then the function set is said to be able to shatter H samples. The Vapnik-Chervonenkis (VC) dimension of a set of functions is the maximum number of samples it can shatter. If there are functions that can shatter any number of samples, then the VC dimension of the set of functions is infinite, and the VC dimension of a bounded real function can be defined by converting it to an indicator function with a certain threshold.

The DT model is a decision analysis method that obtains the probability that the expected value of net present value is greater than or equal to zero, evaluates the project risk, and judges its feasibility based on the known probability of occurrence of various situations by forming a decision tree. The DT model is a graphical method that directly uses probability analysis. In machine learning, the DT algorithm can be used as a prediction model, which represents a mapping relationship between object attributes and object values.

RF is an algorithm that integrates multiple trees through the idea of ensemble learning. Its basic unit is the decision tree, and its essence belongs to the ensemble learning method of machine learning. In fact, from an intuitive point of view, each decision tree is a classifier. Therefore, for input with n samples, n trees will have n classification results. The RF algorithm integrates all the classification voting results and designates the category with the most votes as the final output. In the RF algorithm, k feature columns are randomly selected from the dataset with a total of m features (where k is less than or equal to m), and then a decision tree is established based on these k features. This process should be repeated n times, and n decision trees can be built from the k properties through different random combinations. Finally, the random variables are passed to each decision tree to predict the outcome.

In this study, the input variables of the model are the ratio of solid waste to cement, the ratio of fine sand, the ratio of solid waste to water, and the curing time, and the output is the compressive strength of concrete samples. These machine learning algorithms are widely used to address the regression problems in the engineering area such as the prediction of the concrete performance, rock performance, and rockburst. However, these models have not been compared to find the optimized one, especially regarding the concrete samples with tailings as aggregate. Therefore, it is necessary to compare them and find out the best model to predict the compressive strength of concrete specimens with mining waste as aggregate.

2.3. Evaluation of the Predictive Performance. The evaluation method is crucial for the performance of machine learning models because it can help researchers find the optimal algorithm for application. In this study, the dataset is divided into two parts: the training set (70%, including 100 data points) and the testing set (30%, including 44 data points). Each dataset was then validated using a 10-fold cross-validation method. According to the literature, the correlation coefficient (R) and root mean square error (RMSE) were used in this study to evaluate the performance of the applied model. The two parameters can be described as follows:

$$R = \frac{\sum_{i=1}^{N}\left(y_i^* - \overline{y^*}\right)\left(y_i - \overline{y}\right)}{\sqrt{\sum_{i=1}^{N}\left(y_i^* - \overline{y^*}\right)^2}\sqrt{\sum_{i=1}^{N}\left(y^i - \overline{y}\right)^2}},$$
$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i^* - y_i\right)^2}, \tag{2}$$

where N represents the number of datum, which is 144 in the present study; $y*i$ and y_i represent the expected and measured compressive strengths of the concrete samples, respectively; $\overline{y}$ and $\overline{y^*}$ are the average values of the predicted and measured compressive strengths of the concrete samples, respectively.

3. Results and Discussion

3.1. Experimental Results of the Compressive Strength. Figure 2 gives the experimental results of the compressive strength of the concrete samples using the tailings as the aggregate and all the datasets, the solid mining waste to cement ratio, fine sand ratio, solid mining waste to water ratio, and curing time, are summarized in it.

As can be seen from Figure 2, the influence of fine sand ratio on the compressive strength of concrete specimens is generally negative, that is, with the decrease of fine tailings, the compressive strength of concrete specimens increases. However, it should be noted that, in some cases, the effect of fine tailings is not clear. In addition, with the increase of curing time, the compressive strength of concrete specimens increases rapidly. The solid mining waste to water ratio also has a similar influencing effect, and the compressive strength of concrete samples increases with the increase of the solid mining waste to water ratio. However, with the increase of solid mining waste to cement ratio, its strength decreases. The experimental results are consistent with the results of the previous literature, indicating that the dataset is reliable and accurate.

3.2. Hyperparameters Tuning. Figure 3 gives the relationship between iterations and RMSE values.

For different artificial intelligence algorithms, BAS is used to find the optimized hyperparameters. As can be seen from Figure 3, RMSE drops rapidly after several iterations and then remains stable, indicating that BAS can effectively tune these artificial intelligence models. The reduction rates of RMSE values of the three models are similar, indicating the lower difference in computational efficiency. Also, it can be clearly seen from Figure 3 that, due to the hyperparameters tuning of the BAS algorithm, the SVM model can obtain the minimum RMSE, while BAS algorithm is inefficient in DT and RF models. In summary, the BAS algorithm is efficient and accurate in the hyperparameter tuning

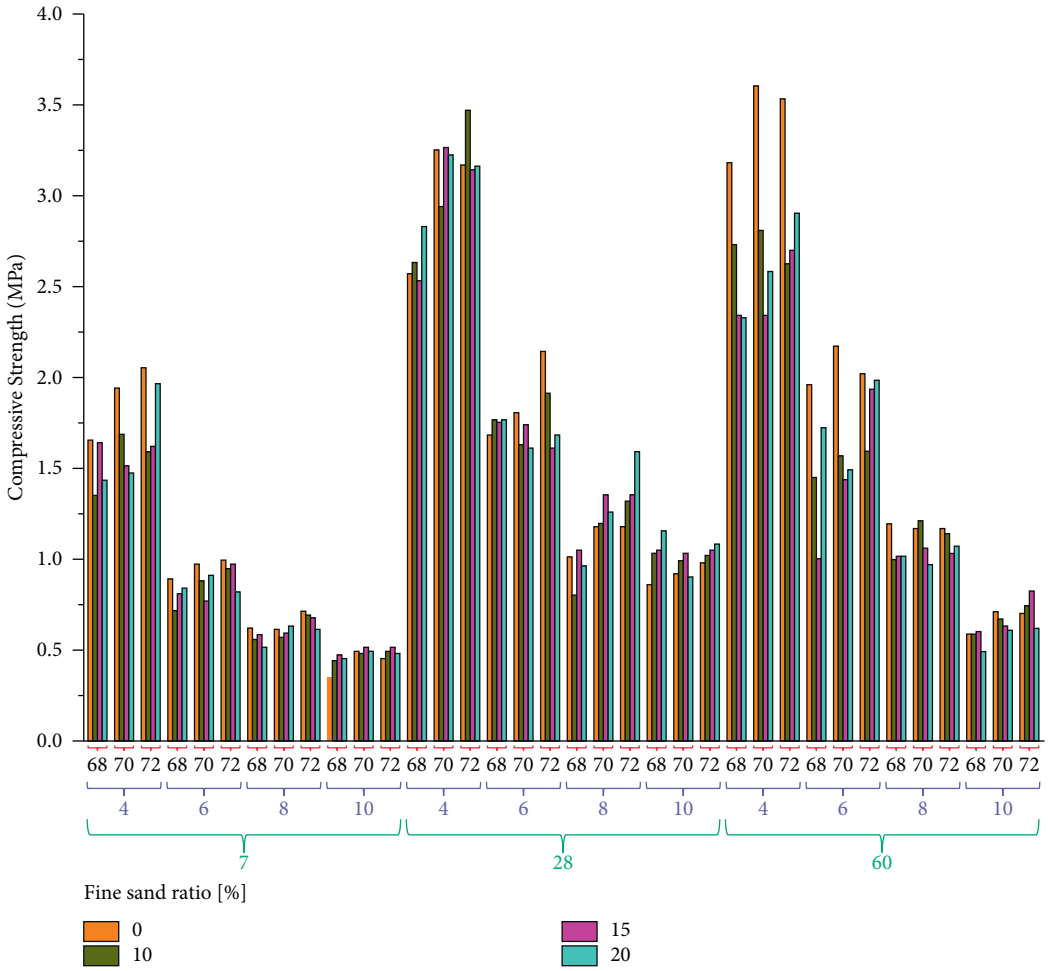


Figure 2: Experimental results.

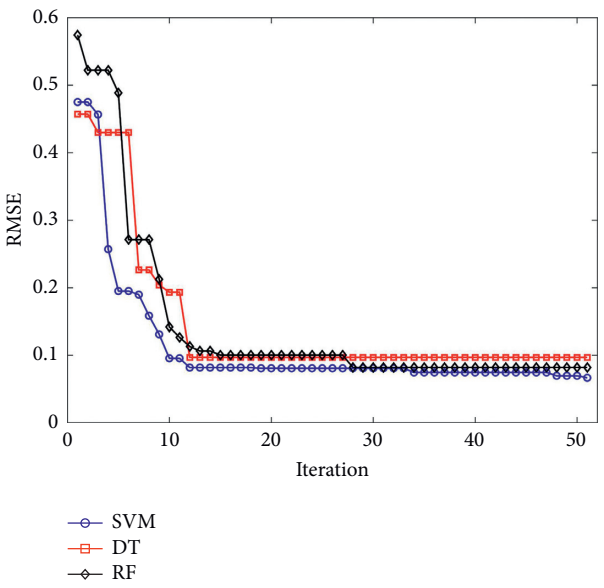


Figure 3: Relationship between the iterations and RMSE values for different models.

process of the three models and the following model evaluation process can be conducted.

The 10-fold cross-validation (CV) was employed for the hyperparameter tuning in the three machine learning models. Figures 4–6 give the RMSE values of different folds using the SVM, DT, and RF models, respectively.

It can be seen from the figures that the SVM model can obtain the minimum value of RMSE at the 4th fold with a value of 0.065 during the 10-fold CV process. The minimum value of RMSE in the DT model can be obtained at the 4th fold as well, with a value of 0.095. RF model can obtain the minimum value of RMSE at the 7th fold with a value of 0.08.

3.3. Model Evaluation. The predictive results of the three machine learning models employed in this study were systematically compared. Figures 7–9 give the results of the comparison between the actual compressive strength and predicted compressive strength using the SVM model, DT model, and RF model, respectively.

It can be observed that the RMSE values (test set) of the SVM model, DT model, and RF model were 0.2332, 0.24, and

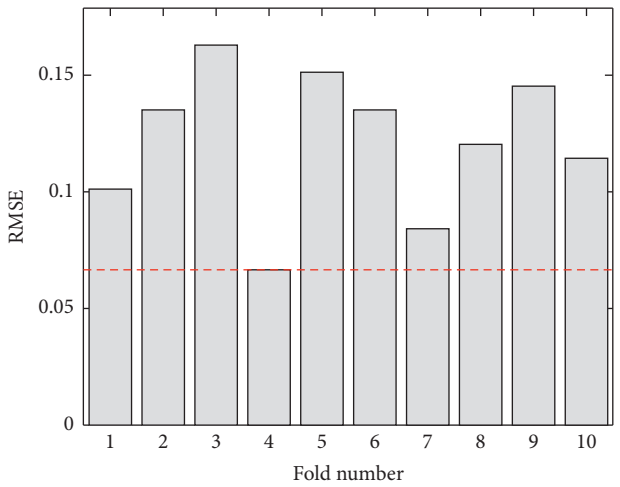


Figure 4: RMSE values of different folds (SVM model).

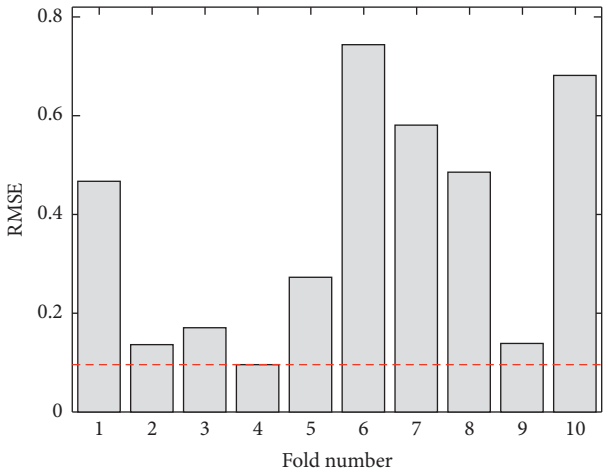


Figure 5: RMSE values of different folds (DT model).

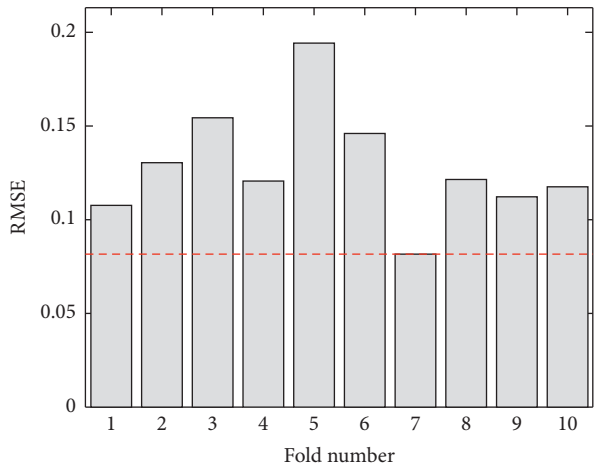


Figure 6: RMSE values of different folds (RF model).

0.2286, respectively. R values (test set) of the SVM model, DT model, and RF model were 0.9699, 0.9619, and 0.9731, respectively. It is indicated that the RF model is the

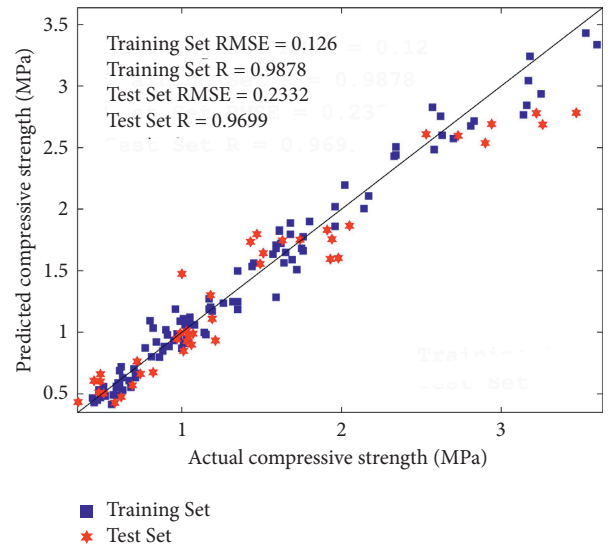


Figure 7: Comparison of the actual compressive strength and predicted compressive strength (SVM model).

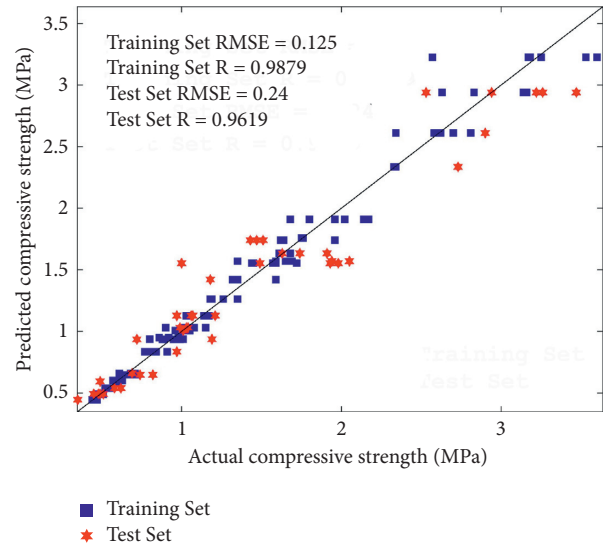


Figure 8: Comparison of the actual compressive strength and predicted compressive strength (DT model).

optimized machine learning algorithm to predict the compressive strength of the cement-based materials using the tailings as the aggregates. However, it should be noted that the predictive performances of the three models are quite similar (less than 5%).

3.4. Variable Importance Evaluation. The model can effectively predict the compressive strength of the cementing material of tailings as aggregate, and the relative importance of each variable can be determined by the machine learning model. To achieve optimal prediction performance, the RF model was used to calculate the importance of variables in this study, and the results are shown in Figure 10.

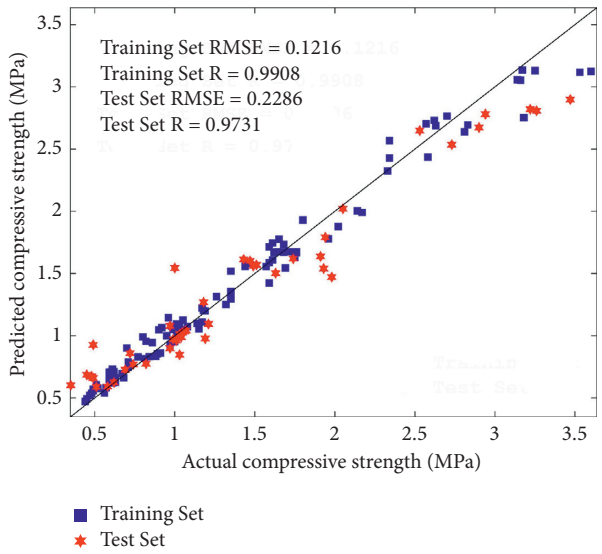


Figure 9: Comparison of the actual compressive strength and predicted compressive strength (RF model).

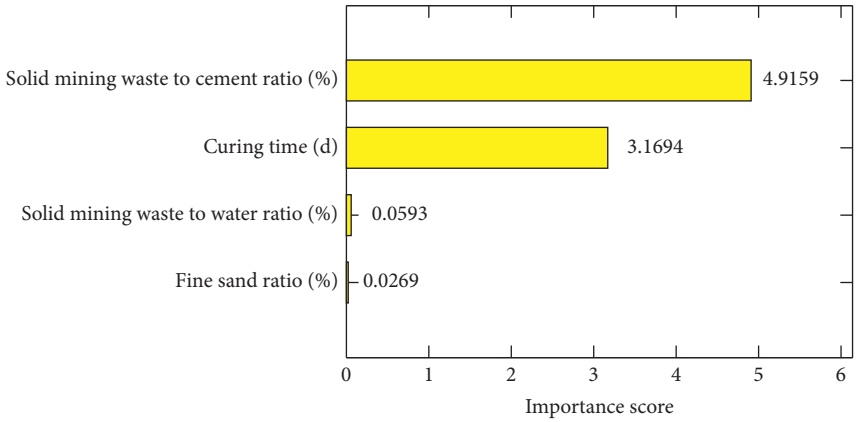


Figure 10: Importances of different variables.

It can be seen that the solid mining waste to cement ratio is the most important variable, and its influence score is 4.9159. Curing time (3.1694) can also be used as an important parameter in the compressive strength of cemented materials relative to the total amount of mining waste, followed by the water-solid ratio of mining waste (0.0593) and fine sand ratio (0.0269), indicating that engineers should pay more attention to the ratio of solid mining waste to cement. The above results have a certain guiding significance for the mix design of the cemented materials with mining waste as aggregate and field construction.

4. Conclusions

Based on a series of experimental studies on cement-based materials with mining waste as the aggregates, the dataset was constructed and the model was evaluated. Three machine learning models (SVM, DT, and RF) were used to predict the compressive strength of cement-based materials with mining waste as the aggregates, and the prediction results of different models were compared. The corresponding results are as follows:

(1) The results of compressive strength of cement-based materials show that, with the increase of water content and curing time of solid mine waste, the compressive strength of cement-based materials increases, while it decreases with the increase of fine sand ratio and solid waste rock cement ratio.

(2) The BAS algorithm can effectively tune these artificial intelligence models. The reduction rates of RMSE values of the three models are similar, indicating the lower difference in computational efficiency. Due to the hyperparameters tuning of the BAS algorithm, the SVM model can obtain the minimum RMSE, while the BAS algorithm is inefficient in DT and RF models.

(3) Based on the prediction results from the developed model, the SVM, DT, and RF models can be used to

predict the compressive strength of cement-based materials using solid mining waste as aggregate effectively and accurately, with high R values (0.9699, 0.9619, and 0.9731 for the SVM, DT, and RF models) and lower RMSE values (SVM, DT, and RF models were 0.2332, 0.24, 0.2286, respectively). Comparing the three developed models, the RF algorithm can obtain the highest value of R and the lowest value of RMSE, demonstrating the highest accuracy.

(4) The solid mining waste to cement ratio is the most important variable to affect the compressive strength of the cement-based materials using the mining waste as the aggregates. Curing time was also an important parameter in the compressive strength of cemented materials relative to the total amount of mining waste, followed by the water-solid ratio of mining waste and fine sand ratio. These results have a certain guiding significance for the mix design of the cemented materials with mining waste as aggregate and field construction in the future.

It should be pointed out that the accuracy and reliability of these developed machine learning models depend on the dataset, that is, the number and type of samples. Therefore, in the future, more samples will need to be obtained and more efficient models should be proposed.

Data Availability

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

Conflicts of Interest

There are no conflicts of interest.

References

[1] X. Zhang, Q. Jiang, N. Chen, W. Wei, and X. Feng, "Laboratory investigation on shear behavior of rock joints and a new peak shear strength criterion," *Rock Mechanics and Rock Engineering*, vol. 49, no. 9, pp. 3495–3512, 2016.

[2] G. You, M. A. Mandalawi, A. Soliman, K. Dowling, and P. Dahlhaus, "Finite element analysis of rock slope stability using shear strength reduction method," in *Proceedings of the International Congress and Exhibition "Sustainable Civil Infrastructures: Innovative Infrastructure Geotechnology*, pp. 227–235, Springer, Sharm El Sheikh, Egypt, July 2017.

[3] C. Xu, M. N. Amar, M. A. Ghriga, H. Ouaer, X. Zhang, and M. Hasanipanah, "Evolving support vector regression using grey wolf optimization; forecasting the geomechanical properties of rock," *Engineering with Computers*, pp. 1–15, 2020.

[4] H. Xie, J.-A. Wang, and W.-H. Xie, "Fractal effects of surface roughness on the mechanical behavior of rock joints," *Chaos, Solitons & Fractals*, vol. 8, no. 2, pp. 221–252, 1997.

[5] Z. Sun, J. Shi, K. Wu, T. Zhang, D. Feng, and X. Li, "Effect of pressure-propagation behavior on production performance: implication for advancing low-permeability coalbed-methane recovery," *SPE Journal*, vol. 24, no. 2, pp. 681–697, 2019.

[6] Z. Sun, X. Li, W. Liu, T. Zhang, M. He, and H. Nasrabadi, "Molecular dynamics of methane flow behavior through realistic organic nanopores under geologic shale condition: pore size and kerogen types," *Chemical Engineering Journal*, vol. 398, Article ID 124341, 2020.

[7] R. Wu, P. Kulatilake, H. Luo, and K. Zhao, "Design of the key bearing layer and secondary mining technology for previously mined areas of small coal mines," *Rock Mechanics and Rock Engineering*, vol. 53, pp. 1–15, 2019.

[8] G. Tiwari and G. M. Latha, "Shear velocity-based uncertainty quantification for rock joint shear strength," *Bulletin of Engineering Geology and the Environment*, vol. 78, no. 8, pp. 5937–5949, 2019.

[9] D. Tarchi, G. Antonello, N. Casagli et al., "On the use of ground-based sar interferometry for slope failure early warning: the cortenova rock slide (Italy)," in *Landslides*, pp. 337–342, Springer, Berlin, Germany, 2005.

[10] F. Xiao, T. Wang, X. Hou, J. Yuan, C. Jiang, and Y. Luo, "Waterproof and antiscour properties of asphalt-based composite seals for airfield base layer," *Journal of Materials in Civil Engineering*, vol. 32, no. 1, Article ID 04019328, 2020.

[11] T. Wang, J. Wang, X. Hou, and F. Xiao, "Effects of sara fractions on low temperature properties of asphalt binders," *Road Materials and Pavement Design*, vol. 22, no. 3, pp. 539–556, 2021.

[12] X. Hou, F. Xiao, R. Guo, Q. Xiang, T. Wang, and J. Wang, "Application of spectrophotometry on detecting asphalt content of emulsified asphalt," *Journal of Cleaner Production*, vol. 215, pp. 626–633, 2019.

[13] Z. C. Tang and L. N. Y. Wong, "New criterion for evaluating the peak shear strength of rock joints under different contact states," *Rock Mechanics and Rock Engineering*, vol. 49, no. 4, pp. 1191–1199, 2016.

[14] D. Sow, C. Carvajal, P. Breul et al., "Modeling the spatial variability of the shear strength of discontinuities of rock masses: application to a dam rock mass," *Engineering Geology*, vol. 220, pp. 133–143, 2017.

[15] M. Souley, F. Homand, and B. Amadei, "An extension to the saeb and amadei constitutive model for rock joints to include cyclic loading paths," *International Journal of Rock Mechanics and Mining Science & Geomechanics Abstracts*, pp. 101–109, Elsevier, 1995.

[16] K. Senthil, A. Arockiarajan, R. Palaninathan, B. Santhosh, and K. M. Usha, "Defects in composite structures: its effects and prediction methods - a comprehensive review," *Composite Structures*, vol. 106, pp. 139–149, 2013.

[17] R. Resende, J. Muralha, A. L. Ramos, and E. Fortunato, "Rock joint topography: three-dimensional scanning and numerical analysis," *Géotechnique Letters*, vol. 5, no. 4, pp. 318–323, 2015.

[18] H. Yu, Z. Zhu, Z. Leng et al., "Effect of mixing sequence on asphalt mixtures containing waste tire rubber and warm mix surfactants," *Journal of Cleaner Production*, vol. 246, Article ID 119008, 2020.

[19] H. Yu, Z. Leng, Z. Zhou, K. Shih, F. Xiao, and Z. Gao, "Optimization of preparation procedure of liquid warm mix additive modified asphalt rubber," *Journal of Cleaner Production*, vol. 141, pp. 336–345, 2017.

[20] H. Yu, G. Deng, Z. Zhang, M. Zhu, M. Gong, and M. Oeser, "Workability of rubberized asphalt from a perspective of particle effect," *Transportation Research Part D: Transport and Environment*, vol. 91, Article ID 102712, 2021.

[21] J. Zhao, D. Wang, P. Yan, D. Zhang, and H. Wang, "Self-cementitious property of steel slag powder blended with

gypsum," *Construction and Building Materials*, vol. 113, pp. 835–842, 2016.

[22] S. Zhang, Y. Fan, J. Huang, and S. P. Shah, "Effect of nano-metakaolinite clay on the performance of cement-based materials at early curing age," *Construction and Building Materials*, vol. 291, Article ID 123107, 2021.

[23] H. Zhang, H. Li, Y. Zhang, D. Wang, J. Harvey, and H. Wang, "Performance enhancement of porous asphalt pavement using red mud as alternative filler," *Construction and Building Materials*, vol. 160, pp. 707–713, 2018.

[24] S. Yin, A. Wu, K. Hu, Y. Wang, and Y. Zhang, "The effect of solid components on the rheological and mechanical properties of cemented paste backfill," *Minerals Engineering*, vol. 35, pp. 61–66, 2012.

[25] Y.-j. Xie, Q. Fu, K.-R. Zheng, Q. Yuan, and H. Song, "Dynamic mechanical properties of cement and asphalt mortar based on shpb test," *Construction and Building Materials*, vol. 70, pp. 217–225, 2014.

[26] K. Xie, Y. Du, and C. Sun, "Application of the mind-evolution-based machine learning in mixture-ratio calculation of raw materials cement," in *Proceedings of the 3rd World Congress on Intelligent Control and Automation (Cat. No. 00EX393)*, pp. 132–134, IEEE, Hefei, China, June 2000.

[27] M. Wu, R. Li, Y. Zhang, J. Wei, Y. Lv, and X. Ding, "Reinforcement effect of fiber and deoiled asphalt on high viscosity rubber/sbs modified asphalt mortar," *Petroleum Science*, vol. 11, no. 3, pp. 454–459, 2014.

[28] Z. Y. Wang, G. X. Mei, and X. B. Yu, "Dynamic shear modulus and damping ratio of waste granular rubber and cement soil mixtures," *Advanced Materials Research*, Trans Tech Publ, vol. 243-249, pp. 2091–2094, 2011.

[29] Q.-A. Wang, J. Zhang, and J. Huang, "Simulation of the compressive strength of cemented tailing backfill through the use of firefly algorithm and random forest model," *Shock and Vibration*, vol. 2021, Article ID 5536998, 9 pages, 2021.

[30] H. Wang, H. Li, X. Liang, H. Zhou, N. Xie, and Z. Dai, "Investigation on the mechanical properties and environmental impacts of pervious concrete containing fly ash based on the cement-aggregate ratio," *Construction and Building Materials*, vol. 202, pp. 387–395, 2019.

[31] A. Torres, J. Hu, and A. Ramos, "The effect of the cementitious paste thickness on the performance of pervious concrete," *Construction and Building Materials*, vol. 95, pp. 850–859, 2015.

[32] J. Huang, P. Leandri, G. Cuciniello, and M. Losa, "Mix design and laboratory characterisation of rubberised mixture used as damping layer in pavements," *International Journal of Pavement Engineering*, pp. 1–15, 2021.

[33] J. Huang, G. S. Kumar, and Y. Sun, "Evaluation of workability and mechanical properties of asphalt binder and mixture modified with waste toner," *Construction and Building Materials*, vol. 276, Article ID 122230, 2021.

[34] J. Huang, M. Koopialipoor, and D. J. Armaghani, "A combination of fuzzy delphi method and hybrid ann-based systems to forecast ground vibration resulting from blasting," *Scientific Reports*, vol. 10, pp. 19397–19421, 2020.

[35] J. Huang, T. Duan, Y. Zhang, J. Liu, J. Zhang, and Y. Lei, "Predicting the permeability of pervious concrete based on the beetle antennae search algorithm and random forest model," *Advances in Civil Engineering*, vol. 2020, Article ID 8863181, 11 pages, 2020.

[36] J. Huang, T. Duan, Y. Sun, L. Wang, and Y. Lei, "Finite element (fe) modeling of indirect tension to cylindrical (it-cy) specimen test for damping asphalt mixtures (dams)," *Advances in Civil Engineering*, vol. 2020, Article ID 6694180, 11 pages, 2020.

[37] J. Huang, T. Duan, Y. Lei, and M. Hasanipanah, "Finite element modeling for the antivibration pavement used to improve the slope stability of the open-pit mine," *Shock and Vibration*, vol. 2020, Article ID 6650780, 11 pages, 2020.

[38] J. Huang, G. Cuciniello, P. Leandri, and M. Losa, "Design of rubberized asphalt mixtures for noise and vibration damping layers," in *Proceedings of the Advances in Materials and Pavement Performance Prediction II: Contributions to the 2nd International Conference on Advances in Materials and Pavement Performance Prediction (AM3P 2020)*, p. 269, CRC Press, San Antonio, TX, USA, May 2020.

[39] J. Huang, P. G. Asteris, S. M. K. Pasha, A. S. Mohammed, and M. Hasanipanah, "A new auto-tuning model for predicting the rock fragmentation: a cat swarm optimization algorithm," *Engineering with Computers*, Springer, Berlin, Germany, 2020.

[40] J. Huang, R. Alyousef, M. Suhatril et al., "Influence of porosity and cement grade on concrete mechanical properties," *Advances in Concrete Construction*, vol. 10, pp. 393–402, 2020.

[41] Y. Sun, G. Li, J. Zhang, J. Sun, and J. Xu, "Development of an ensemble intelligent model for assessing the strength of cemented paste backfill," *Advances in Civil Engineering*, vol. 2020, Article ID 1643529, 6 pages, 2020.

[42] Y. Sun, G. Li, H. Basarir, A. Karrech, and M. R. Azadi, "Laboratory evaluation of shear strength properties for cement-based grouted coal mass," *Arabian Journal of Geosciences*, vol. 12, no. 22, p. 690, 2019.

[43] M. S. Sumanasooriya and N. Neithalath, "Pore structure features of pervious concretes proportioned for desired porosities and their performance prediction," *Cement and Concrete Composites*, vol. 33, no. 8, pp. 778–787, 2011.

[44] A. Subhy, D. Lo Presti, and G. Airey, "An investigation on using pre-treated tyre rubber as a replacement of synthetic polymers for bitumen modification," *Road Materials and Pavement Design*, vol. 16, no. 1, pp. 245–264, 2015.

[45] D. Shi, P. W. Brown, and W. Ma, "Lognormal simulation of pore size distributions in cementitious materials," *Journal of the American Ceramic Society*, vol. 74, no. 8, pp. 1861–1867, 1991.

[46] M. Sheshpari, "A review of underground mine backfilling methods with emphasis on cemented paste backfill," *Electronic Journal of Geotechnical Engineering*, vol. 20, pp. 5183–5208, 2015.

[47] A. Shadravan, M. Tarrahi, and M. Amani, "Intelligent tool to design fracturing, drilling, spacer and cement slurry fluids using machine learning algorithms," in *Proceedings of the SPE Kuwait Oil and Gas Show and Conference*, Mishref, Kuwait, October 2015.

[48] A. Shadravan, M. Tarrahi, and M. Amani, "Intelligent cement design: utilizing machine learning algorithms to assure effective long-term well integrity," in *Proceedings of the Carbon Management Technology Conference*, Sugar Land, TX, USA, November 2015.

[49] J. Ren, S. Wang, and G. Zang, "Effects of recycled aggregate composition on the mechanical characteristics and material design of cement stabilized cold recycling mixtures using road milling materials," *Construction and Building Materials*, vol. 244, Article ID 118329, 2020.

[50] J. Zhang, Y. Wang, Y. Sun, and G. Li, "Strength of ensemble learning in multiclass classification of rockburst intensity," *International Journal for Numerical and Analytical Methods in Geomechanics*, vol. 44, no. 13, pp. 1833–1853, 2020.

[51] J. Zhang, Y. Sun, G. Li, Y. Wang, J. Sun, and J. Li, "Machine-learning-assisted shear strength prediction of reinforced concrete beams with and without stirrups," *Engineering with Computers*, pp. 1–15, 2020.

[52] J. Zhang, D. Li, and Y. Wang, "Toward intelligent construction: prediction of mechanical properties of manufactured-sand concrete using tree-based models," *Journal of Cleaner Production*, vol. 258, Article ID 120665, 2020.

[53] J. Zhang, D. Li, and Y. Wang, "Predicting tunnel squeezing using a hybrid classifier ensemble with incomplete data," *Bulletin of Engineering Geology and the Environment*, vol. 79, pp. 1–12, 2020.

[54] J. Zhang, F. Jiang, S. Zhu, and L. Zhang, "Width design for gobs and isolated coal pillars based on overall burst-instability prevention in coal mines," *Journal of Rock Mechanics and Geotechnical Engineering*, vol. 8, no. 4, pp. 551–558, 2016.

[55] J. Zhang, Y. Huang, Y. Wang, and G. Ma, "Multi-objective optimization of concrete mixture proportions using machine learning and Metaheuristic algorithms," *Construction and Building Materials*, vol. 253, Article ID 119208, 2020.

[56] J. Zhang, Y. Huang, G. Ma, J. Sun, and B. Nener, "A metaheuristic-optimized multi-output model for predicting multiple properties of pervious concrete," *Construction and Building Materials*, vol. 249, Article ID 118803, 2020.

[57] J. Zhang, Y. Huang, F. Aslani, G. Ma, and B. Nener, "A hybrid intelligent system for designing optimal proportions of recycled aggregate concrete," *Journal of Cleaner Production*, vol. 273, Article ID 122922, 2020.

[58] J. Huang, Y. Zhang, Y. Sun, J. Ren, Z. Zhao, and J. Zhang, "Evaluation of pore size distribution and permeability reduction behavior in pervious concrete," *Construction and Building Materials*, vol. 290, Article ID 123228, 2021.

[59] J. Huang, J. Zhang, J. Ren, and H. Chen, "Anti-rutting performance of the damping asphalt mixtures (dams) made with a high content of asphalt rubber (ar)," *Construction and Building Materials*, vol. 271, Article ID 121878, 2021.

[60] J. Huang, Y. Sun, and J. Zhang, "Reduction of computational error by optimizing svr kernel coefficients to simulate concrete compressive strength through the use of a human learning optimization algorithm," *Engineering with Computers*, Springer, Berlin, Germany, 2021.

[61] J. Huang and Y. Sun, "Viscoelastic analysis of the damping asphalt mixtures (dams) made with a high content of asphalt rubber (ar)," *Advances in Civil Engineering*, vol. 2020, Article ID 8826926, 12 pages, 2020.

[62] J. Huang and Y. Sun, "Effect of modifiers on the rutting, moisture-induced damage, and workability properties of hot mix asphalt mixtures," *Applied Sciences*, vol. 10, no. 20, p. 7145, 2020.

[63] J. Huang, G. Shiva Kumar, J. Ren, J. Zhang, and Y. Sun, "Accurately predicting dynamic modulus of asphalt mixtures in low-temperature regions using hybrid artificial intelligence model," *Construction and Building Materials*, vol. 297, Article ID 123655, 2021.

[64] J. Huang, G. Shiva Kumar, J. Ren, Y. Sun, Y. Li, and C. Wang, "Towards the potential usage of eggshell powder as bio-modifier for asphalt binder and mixture: workability and mechanical properties," *International Journal of Pavement Engineering*, pp. 1–13, 2021.

[65] J. Huang, M. Losa, P. Leandri, S. G. Kumar, J. Zhang, and Y. Sun, "Potential anti-vibration pavements with damping layer: finite element (fe) modeling, validation, and parametrical studies," *Construction and Building Materials*, vol. 281, Article ID 122550, 2021.

[66] J. Huang, M. Losa, and P. Leandri, "Determining the effect of damping layers in flexible pavements on traffic induced vibrations," in *Proceedings of the Advances in Materials and Pavement Prediction: Papers from the International Conference on Advances in Materials and Pavement Performance Prediction (AM3P 2018)*, pp. 255–259, CRC Press, Doha, Qatar, April 2018.

[67] Y. Sun, G. Li, J. Zhang, and J. Xu, "Failure mechanisms of rheological coal roadway," *Sustainability*, vol. 12, no. 7, p. 2885, 2020.

[68] Y. Sun, G. Li, J. Zhang, and D. Qian, "Experimental and numerical investigation on a novel support system for controlling roadway deformation in underground coal mines," *Energy Science & Engineering*, vol. 8, no. 2, pp. 490–500, 2020.

[69] Y. Sun, G. Li, J. Zhang, and D. Qian, "Prediction of the strength of rubberized concrete by an evolved random forest model," *Advances in Civil Engineering*, vol. 2019, Article ID 5198583, 7 pages, 2019.

[70] Y. Sun, G. Li, J. Zhang, and D. Qian, "Stability control for the rheological roadway by a novel high-efficiency jet grouting technique in deep underground coal mines," *Sustainability*, vol. 11, no. 22, p. 6494, 2019.

[71] Y. Sun, G. Li, and J. Zhang, "Developing hybrid machine learning models for estimating the unconfined compressive strength of jet grouting composite: a comparative study," *Applied Sciences*, vol. 10, no. 5, p. 1612, 2020.

[72] Y. Sun, G. Li, and J. Zhang, "Investigation on jet grouting support strategy for controlling time-dependent deformation in the roadway," *Energy Science & Engineering*, vol. 8, 2020.

[73] F. Khademi, S. M. Jamal, N. Deshpande, and S. Londhe, "Predicting strength of recycled aggregate concrete using artificial neural network, adaptive neuro-fuzzy inference system and multiple linear regression," *International Journal of Sustainable Built Environment*, vol. 5, no. 2, pp. 355–369, 2016.

[74] F. Khademi and S. M. Jamal, "Estimating the compressive strength of concrete using multiple linear regression and adaptive neuro-fuzzy inference system," *International Journal of Structural Engineering*, vol. 8, no. 1, pp. 20–31, 2017.

[75] K. Cheng and Z. Lu, "Adaptive bayesian support vector regression model for structural reliability analysis," *Reliability Engineering & System Safety*, vol. 206, Article ID 107286, 2021.

[76] E. S. Chahnasir, Y. Zandi, M. Shariati et al., "Application of support vector machine with firefly algorithm for investigation of the factors affecting the shear strength of angle shear connectors," *Smart Structures and Systems*, vol. 22, pp. 413–424, 2018.

[77] J. Bi and K. P. Bennett, "A geometric approach to support vector regression," *Neurocomputing*, vol. 55, no. 1-2, pp. 79–108, 2003.

[78] N. Hasan, N. C. Nath, and R. I. Rasel, "A support vector regression model for forecasting rainfall," in *2015 2nd International Conference on Electrical Information and Communication Technologies (EICT)*, pp. 554–559, IEEE, Khulna, Bangladesh, December 2015.

[79] H. Drucker, C. J. Burges, L. Kaufman, A. Smola, and V. Vapnik, "Support vector regression machines," *Advances in Neural Information Processing Systems*, vol. 9, pp. 155–161, 1997.

[80] W.-C. Hong and G.-F. Fan, "Hybrid empirical mode decomposition with support vector regression model for short term load forecasting," *Energies*, vol. 12, no. 6, p. 1093, 2019.

[81] M. Hasanipanah, M. Monjezi, A. Shahnazar, D. Jahed Armaghani, and A. Farazmand, "Feasibility of indirect determination of blast induced ground vibration based on support vector machine," *Measurement*, vol. 75, pp. 289–297, 2015.

Hindawi
Advances in Civil Engineering
Volume 2021, Article ID 6671448, 12 pages
https://doi.org/10.1155/2021/6671448

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

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^3, 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, kg/m^3 (X_1); water content, kg/m^3 (X_2); coarse aggregate, kg/m^3 (X_3); fine aggregate or sand, kg/m^3 (X_4); GGFBS content, kg/m^3 (X_5); hyperplasticizing, kg/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 (kg/m^3), but it is mainly in the range of 180 to 270 (kg/m^3). The highest sample number is about 79, which corresponds to 180 kg/m^3 of cement content. Similarly, the water content (X_2) ranges from 70 to 295 kg/m^3. As shown in Figure 1, the coarse aggregate content (X_3) is varied from about 400 to 1200 (kg/m^3), but no sample has coarse aggregate content in the range of 500 to 700 (kg/m^3). The fine aggregate or sand content (X_4) is mainly in two ranges from 500 to 950 (kg/m^3) and 1150 to 1550 (kg/m^3). The highest sample number (X_4) corresponds to 680 kg/m^3 of fine aggregate (or sand content). The GGBFS content (X_5) varies from 40 to 460 kg/m^3, but the values are mostly in the range of 70 to 270 (kg/m^3). The carboxylic-type hyperplasticizing content (X_6) ranges from 2 to 14 kg/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%

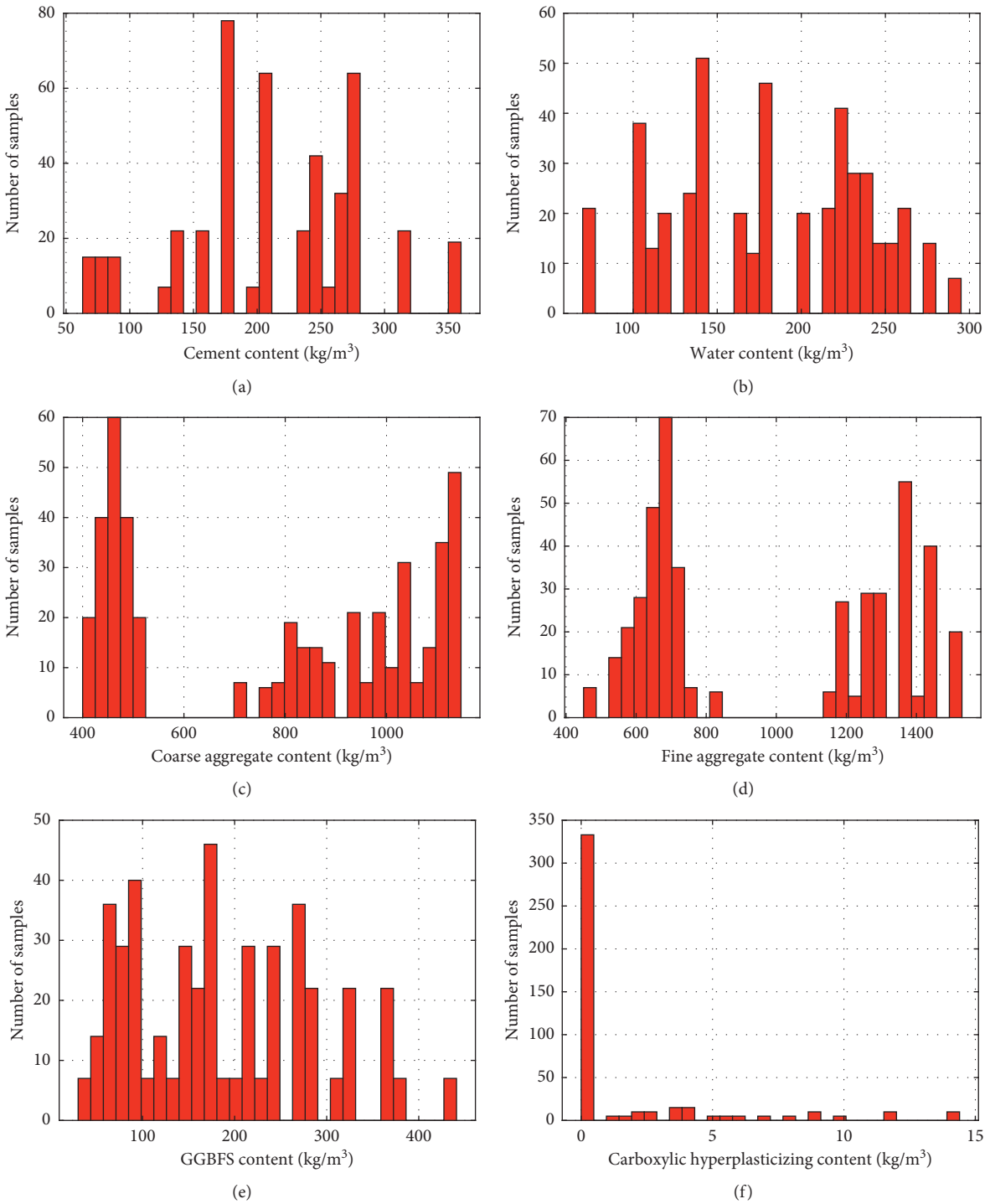


Figure 1: Continued.

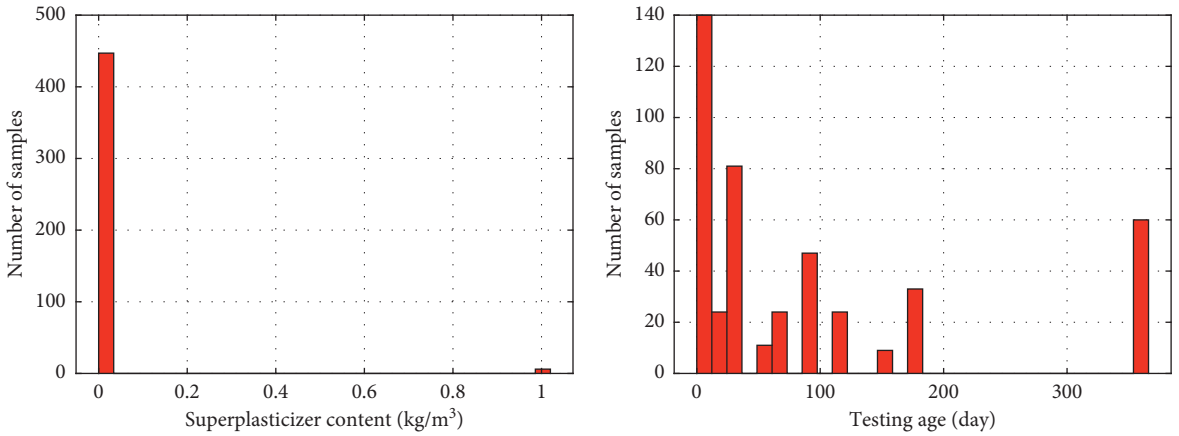


Figure 1: Histograms of the input variables used in this study: (a) cement content; (b) water content; (c) coarse aggregate content; (d) fine aggregate (or sand) content; (e) ground granulated blast furnace slag content; (f) carboxylic-type hyperplasticizing content; (g) superplasticizer content; and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

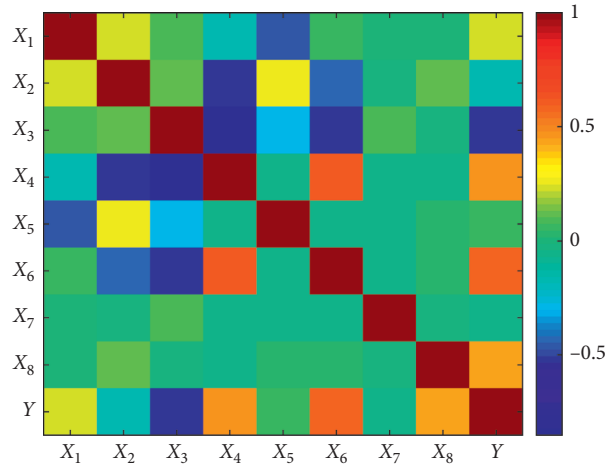


Figure 2: Multicorrelation graph of input and output variables used in this study.

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 $D_1, D_2, \ldots, D_k$ to build trees $T_1, T_2, \ldots, T_k$

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:

$$R = \frac{\sum_{j=1}^{N}\left(Q_{0,j} - \overline{Q}_0\right)\left(Q_{t,j} - \overline{Q}_t\right)}{\sum_{j=1}^{N}\left(Q_{0,j} - \overline{Q}_0\right)^2 \sum_{j=1}^{N}\left(Q_{t,j} - \overline{Q}_t\right)^2},$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(Q_{0,j} - Q_{t,j}\right)^2}, \quad (1)$$

$$\text{MAE} = \frac{1}{N}\sum_{j=1}^{N}\left|Q_{0,j} - Q_{t,j}\right|,$$

where N is the number of database, Q_0 and $\overline{Q}_0$ are the actual experimental value and the average real experimental value, and Q_t and $\overline{Q}_t$ 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

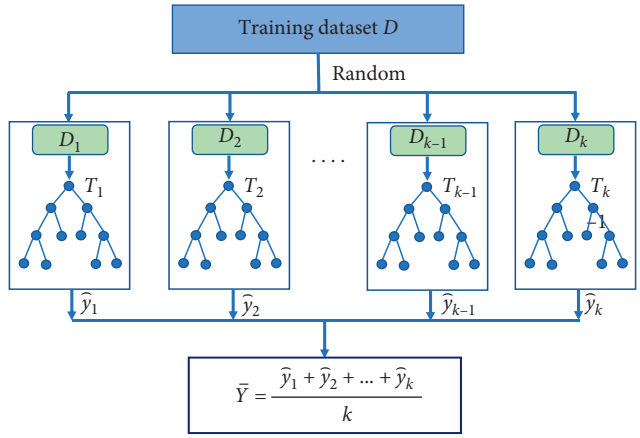


Figure 3: The architecture of the random forest algorithm.

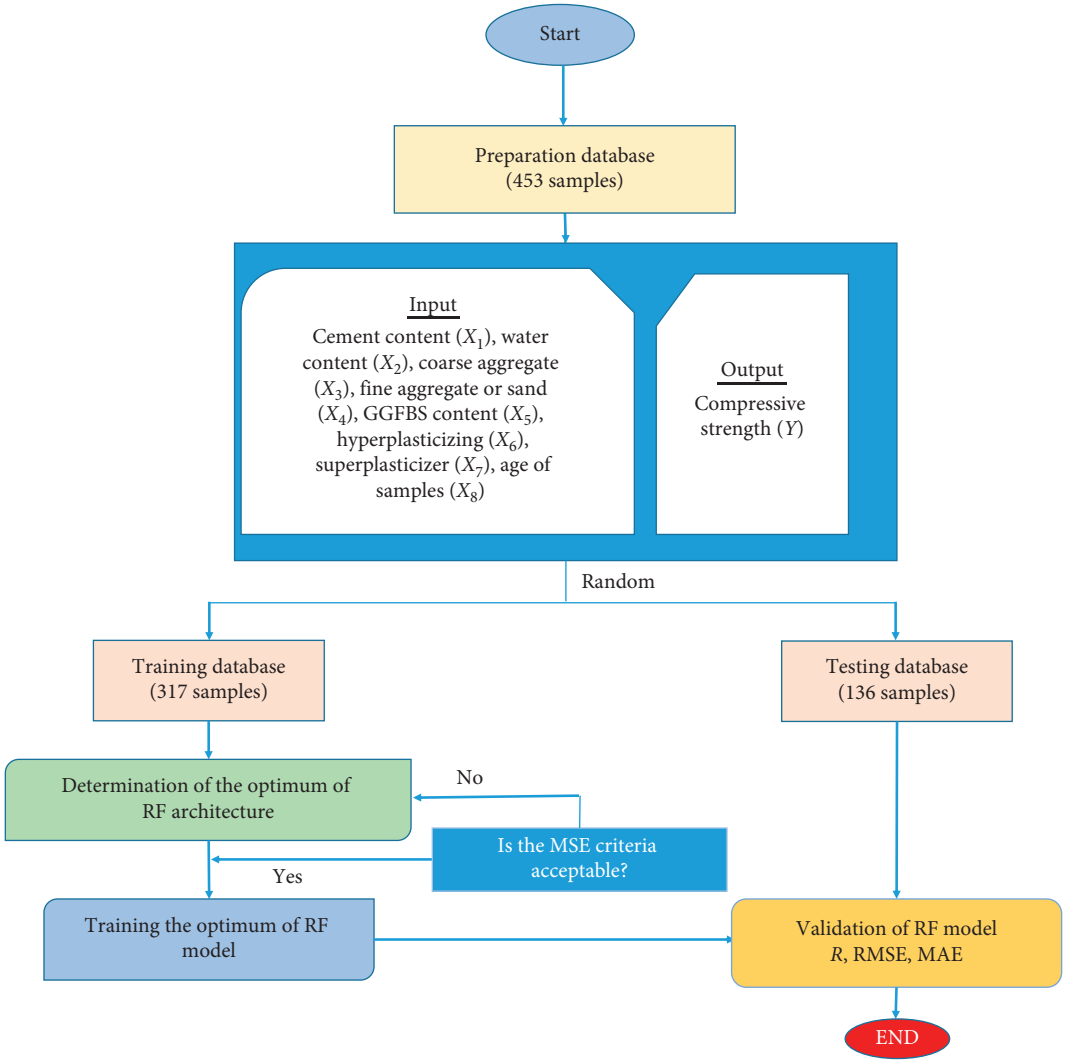


Figure 4: Methodology flow chart.

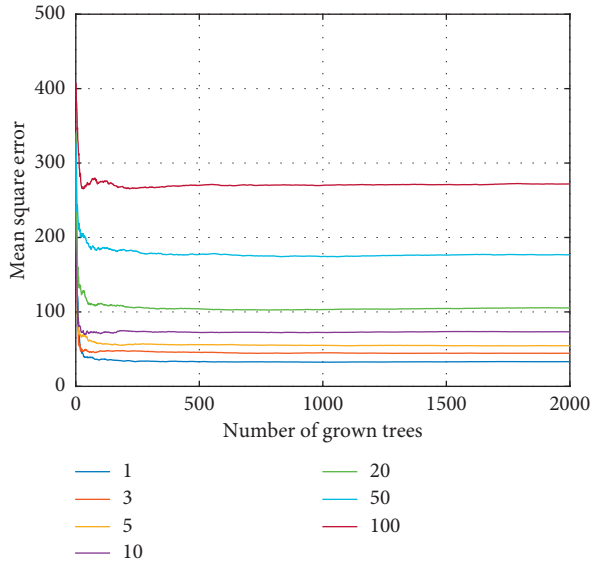


Figure 5: Values of MSE in function of grown tree number and leaf size number.

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-

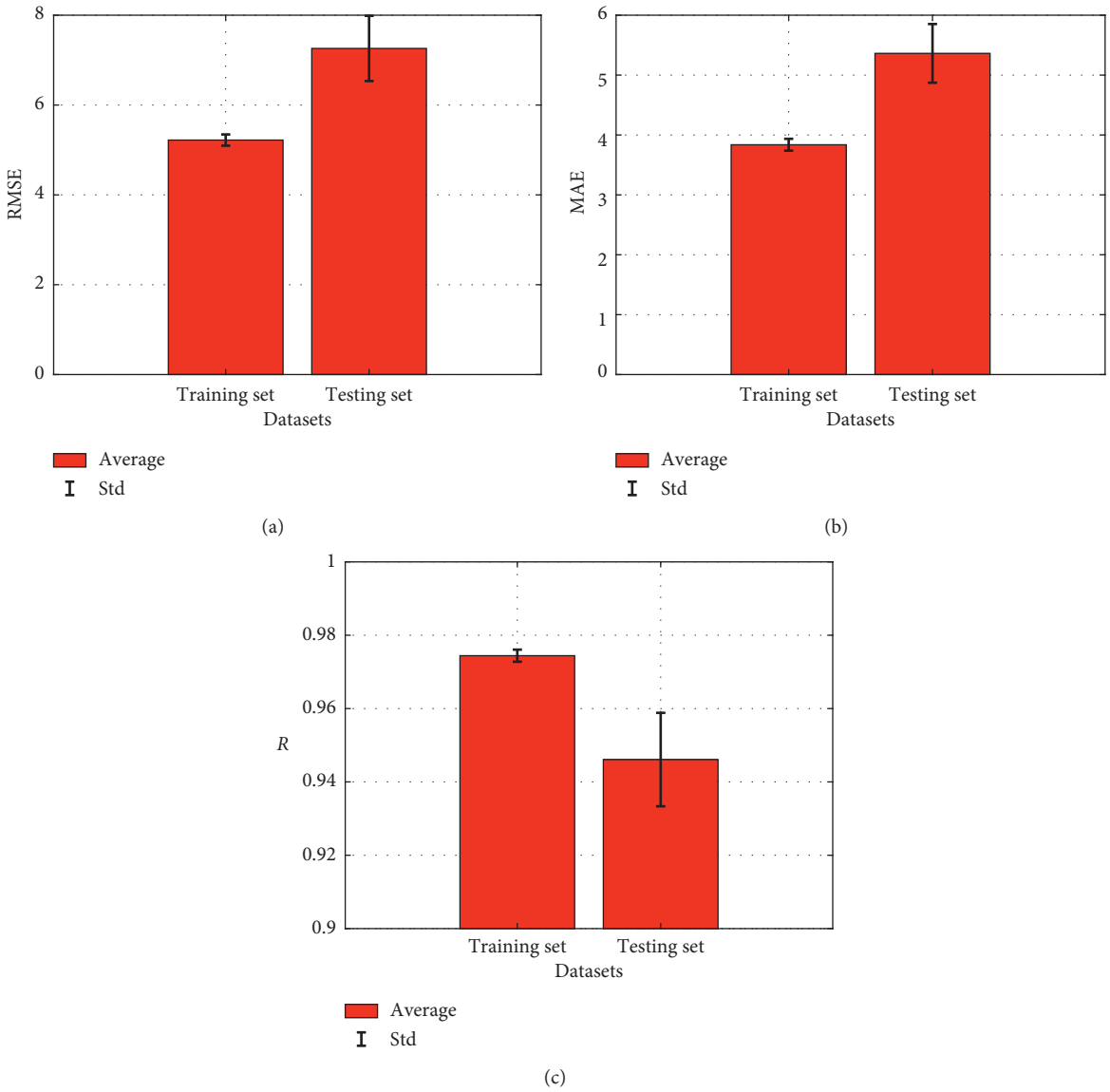


Figure 6: Analysis of the results over 500 simulations (presented in average values with standard deviation) using different RF architectures: (a) R; (b) RMSE; (c) MAE.

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

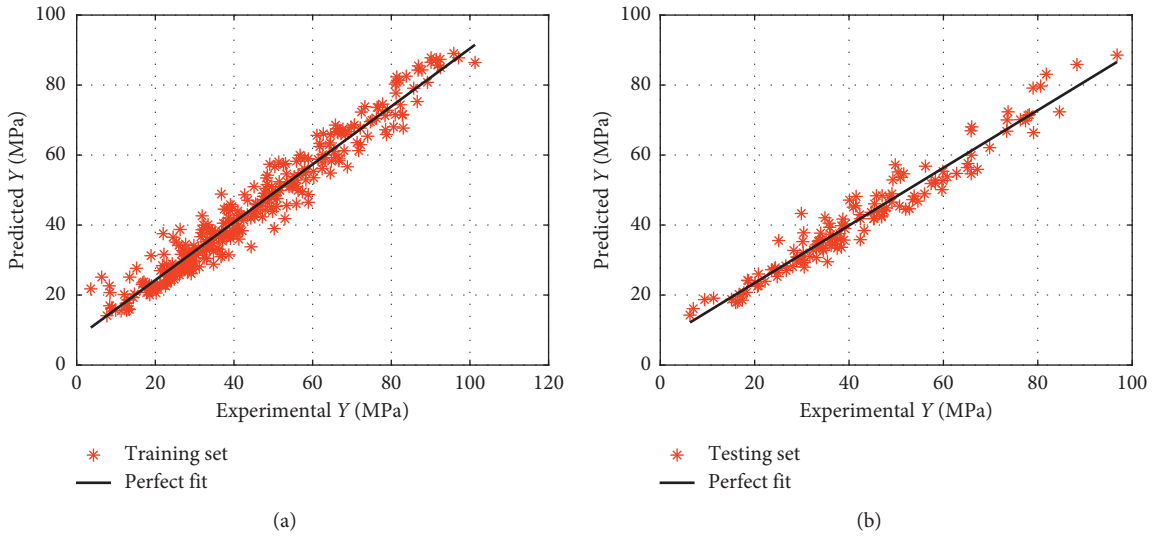


Figure 7: Regression graphs of the best predictor RF between experimental and predicted compressive strength: (a) training dataset and (b) testing dataset.

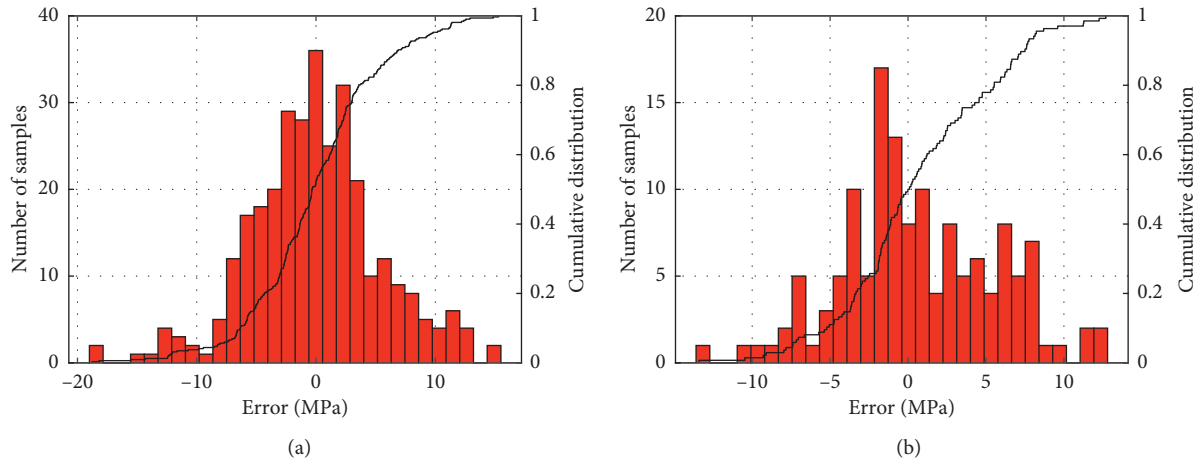


Figure 8: Error between target and output value plots for the case of the best RF architecture: (a) training dataset and (b) testing dataset.

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.

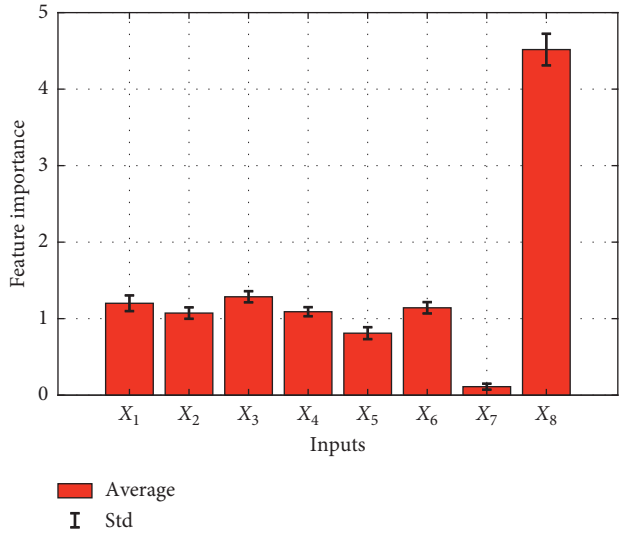


Figure 9: Feature importance over 500 simulations.

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] Move 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. Atiş, H. Tanyildizi, 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. Severcan, "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.

Hindawi
Advances in Civil Engineering
Volume 2021, Article ID 8878396, 14 pages
https://doi.org/10.1155/2021/8878396

Research Article

Application of Extreme Gradient Boosting Based on Grey Relation Analysis for Prediction of Compressive Strength of Concrete

Liyun Cui,[1] **Peiyuan Chen**,[1] **Liang Wang**,[1] **Jin Li**,[1] **and Hao Ling**[2]

[1]*School of Civil Engineering and Architecture, Anhui University of Science and Technology, Huainan 232001, China*
[2]*School of Materials Science and Engineering, Southeast University, Nanjing 211189, China*

Correspondence should be addressed to Peiyuan Chen; peiyuan29@126.com

Received 9 August 2020; Revised 3 October 2020; Accepted 5 February 2021; Published 18 March 2021

Academic Editor: Yuantian Sun

The prediction of concrete strength is an interesting point of investigation and could be realized well, especially for the concrete with the complex system, with the development of machine learning and artificial intelligence. Therefore, an excellent algorithm should put emphasis to receiving increased attention from researchers. This study presents a novel predictive system as follows: extreme gradient boosting (XGBoost) based on grey relation analysis (GRA) for predicting the compressive strength of concrete containing slag and metakaolin. One of its highlights is a feature selection methodology, i.e., GRA, which was used to determine the main input variables. Another highlight is that its performance was compared with the frequently used artificial neural network (ANN) and genetic algorithm-artificial neural network (GA-ANN) by using random dataset and the same testing datasets. For three same testing datasets, the average R^2 values of ANN, GA-ANN, and XGBoost are 0.674, 0.829, and 0.880, respectively, indicating that XGBoost has the highest absolute fraction of variance (R^2). XGBoost can provide best result by testing the root mean squared error (RMSE) and mean absolute percentage error (MAPE). The average RMSE values of ANN, GA-ANN, and XGBoost are 15.569 MPa, 10.530 MPa, and 9.532 MPa, respectively, and those of MAPE of ANN, GA-ANN, and XGBoost are 11.224%, 9.140%, and 8.718%, respectively. Thus, the XGBoost definitely performed better than the ANN and GA-ANN. Finally, a type of application software based on XGBoost was developed for practical applications. This vivid software interfaces could help users in prediction and easy and efficient analysis.

1. Introduction

A series of dilemmas including waste emission and overconsumption of energy and natural resources have been currently pressing worldwide concerns because of global population explosion and rapid urbanization. According to the International Energy Agency report, 4.8 billion tons of cement will be produced worldwide by 2050, resulting in the emission of 3.8 billion tons of CO_2 [1]. Therefore, it is beneficial to find and utilize active admixtures with high-quality and low-energy consumption as alternatives of cement, partly or totally [1, 2]. In addition, those active admixtures can even enhance the properties of concrete such as compressive strength, antipermeability, and corrosion resistance [3]. Therefore, mix proportion of concrete, especially the high-performance concrete, has been in increasing demand. Mix design of concrete is difficult to achieve for researchers, as it must reach a specific strength level, which is the most important property of concrete.

Currently, combining with machine learning, some researchers utilize various basic prediction models, such as logistic regression (LR), random forest model (BRF), support vector machine (SVM), and artificial neural network (ANN) [4–6]. These models are commonly used for predicting compressive strength of concrete, irrespective of costly and time-consuming nature. The artificial neural network, inspired by biological systems of the human brain, can learn and generalize from experience without prior knowledge. It has powerful pattern classification and pattern recognition capabilities [7]. Bilim et al. [8] conducted various ANNs and indicated their strong potential for predicting the compressive strength of granulated blast furnace

slag concrete. Saridemir et al. [9] investigated ANNs for predicting the compressive strength of concrete including metakaolin and silica fume, exhibiting good performance. However, the determination of weights and biases of ANN is a complicated process, and the standard ANN algorithm has a slow convergence rate and is easy to fall into the local minimum. As a matter of course, there are many kinds of intelligent algorithms used to optimize the superparameter parameters of the basic model such as conjugate simulated annealing [10], gradient algorithm [11], and genetic algorithm (GA) [12]. Among those algorithms, the performance of GA is excellent, attributing to its advantages such as scalability, adaptation, speed, fault tolerance, modularity, autonomy, and parallelism. Firouzi et al. [13] proposed an integrated GA-ANN approach, appropriate for finding optimum reliability-based inspection plans based on minimum life cycle costs.

The ANN model needs a lot of parameters, and the learning process is like a black box. Hence, it is difficult to carry out the actual derivation step by step, affecting the credibility. In addition, in the learning process, it is easy to appear overfitting state; therefore, some researchers used GA to optimize the internal structure parameters of the neural network. Although GA is a solution for optimizing the neural network, its programming process is complex and the search speed is slow. Therefore, another new model, XGBoost, will be introduced later.

The extreme gradient boosting (XGBoost) is a novel machine learning algorithm mainly used for supporting classification, regression, and ranking. In recent years, it has gradually attracted attention owing to its excellent performance of strong learning ability and fast convergence during parallel learning computations shown in many AI competitions. In KDDCup competition including commercial sales forecast, the team of Top 10 used the XGBoost algorithm for web page text classification, customer behavior prediction, ad click rate prediction, and hazard risk prediction and other fields [14]. In consideration of those outstanding achievements, in this study, XGBoost was applied for predicting the compressive strength of concrete containing slag and metakaolin. In addition, the performance of the model was compared with the commonly used ANN and superior GA-ANN in order to reflect the advantages or disadvantages of XGBoost.

Before predicting the compressive strength of concrete based on ANN, GA-ANN, and XGBoost, the determination of main influencing factors of concrete strength (i.e., the input variables of prediction models) is a primary thing because it is beneficial to reduce the dimensions of the model and improve prediction accuracy [15]. In this study, there is a highlight that a feature selection methodology, grey relation analysis (GRA), was used to determine the main input variables. Figure 1 shows the overall process of predicting compressive strength of concrete containing slag and metakaolin. To train and test those prediction models, 600 groups of data selected from 18 research papers (Table 1) were utilized, after a necessary pretreatment process, i.e., unifying the compressive strength of concrete under different dimensions to avoid the influence of dimension effect. Those data were divided into training dataset and testing dataset. The quantitative analysis of prediction performance of ANN, GA-ANN, and XGBoost was obtained by testing the value of absolute fraction of variance (R^2), mean absolute percentage error (MAPE), and root mean squared error (RMSE).

2. Preprocessing and Preanalysis of Selected Data

2.1. Preprocessing of Data

2.1.1. Unifying the Compressive Strength of Concrete. Because of the dimension effect, the values of compressive strength will be valid just when they are compared under a uniform dimension. A cube with a length of 100 mm was frequently used in 600 groups of collected data. Therefore, it acts as the uniform dimension in this study. For other values of compressive strength of concrete under different dimensions, they would be first transformed by the formula of Neville [32], as shown in the following equation:

$$\frac{P}{P_o} = 0.56 + \frac{0.697}{(V/6h\ d) + (h/d)}, \tag{1}$$

where P is the compressive strength of concrete under other dimensions; P_0 is the compressive strength of cube (length = 150 mm, approximately 5.9 inches); V is the specimen volume; h is the height; and d is the maximum transverse dimension of specimen. Notably, all the numerical values must be calculated with a unit of inch. By the formula of Neville and another simple conversion (cube with 150 mm length to cube with 100 mm length), the scaling factor (the value of P/P_0) based on the uniform dimension of cube with a length of 100 mm was calculated, as listed in Table 2.

2.1.2. Normalization of Data. In the system of data-driven modelling, in order to eliminate the influences of outliers, missed values, and bad data, all raw data should be normalized to fit the range of [0, 1]. It ensures that the raw data collected by major journals are more suitable for modelling, and it also helps to achieve better results and significantly speeds up the calculation [32]. The linear mapping function is as follows:

$$X_m = \frac{X - X_{\min}}{X_{\max} - X_{\min}}, \tag{2}$$

where X represents the original value; $X_{\min}$ and $X_{\max}$ are the minimum value and maximum value of dataset, respectively; and X_{m} is the value after mapping, noting that the outputs will be remapped to their corresponding real values by the inverse mapping function before calculating any performance criterion [31].

3. Preanalysis of Data: Feature Selection Methodology

Generally speaking, the compressive strength of traditional concrete was determined by various factors, including cement type, water-binder ratio (W/B), sand ratio, dosage of

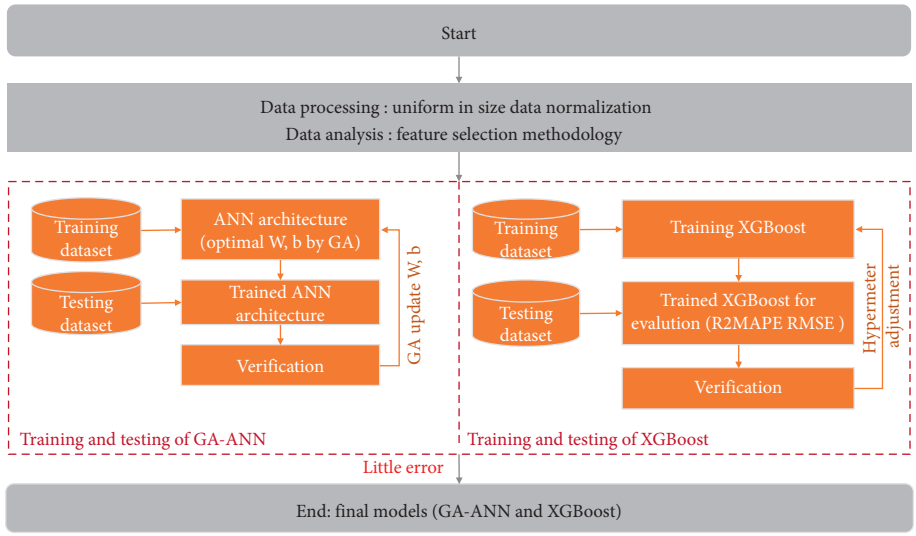


Figure 1: The overall process of GA-ANN and XGBoost for predicting the compressive strength of concrete.

Table 1: Data source and number of data.

Author of data source	Year	Number of data
Liu et al. [16]	2018	12
Duan et al. [17]	2013	16
Sujjavanich et al. [18]	2017	55
Mermerdaş et al. [19]	2012	84
Güneyisi et al. [20]	2008	36
Yunfeng et al. [21]	2017	90
Ramezanianpour et al. [22]	2012	48
Shekarchi et al. [23]	2010	20
Shi et al. [24]	2015	18
Shannag et al. [25]	1995	32
Shafiq et al. [26]	2015	21
Khatib et al. [27]	2008	30
El-diadamony et al. [28]	2016	20
Güneyisi et al. [29]	2012	30
Perez-cortes et al. [30]	2019	14
Dinakar et al. [31]	2013	16
Poon et al. [1]	2001	28
Bilim et al. [8]	2009	30
Total		600

water-reducing agent, and curing age. For concrete containing slag and metakaolin, the content of slag, MK, and the Si/Al of metakaolin (Si/Al of MK) should also be included in consideration of some samples; therefore, the Fly ash content should be considered as an important factor. The selection of main influencing factors among various factors is significant in order to improve the efficiency and accuracy of the prediction models. In this study, GRA, a feature selection methodology, was used to address the abovementioned problem for obtaining the most influential and significant factors.

3.1. Determination of the Analysis Sequence. First, we construct the reference matrix (compressive strength of concrete) $X_0\ (k)$ (where $k = 1, 2, \ldots, 600$, i.e., the number of sample datasets) and comparative matrix (cement type, water-binder ratio (W/B), sand ratio, dosage of water-reducing agent, curing age, slag content, fly ash content, MK content, and Si/Al of MK) $X_i\ (k)$ (where $i = 1, 2, \ldots, 9$, i.e., the number of factors and $k = 1, 2, \ldots, 600$). The equations in Table 3 are the mathematical forms for the construction of the reference matrix and the comparative matrix [33].

3.2. Dimensionless Treatment. For the purpose of reducing the numerical fluctuation, the normalization process can be determined as follows:

$$x_i(k) = \frac{X_i(k)}{(1/n)\sum_{k=1}^{n} X_i(k)}, n = 600, \quad i = 1, 2, 3, \ldots, 9. \tag{3}$$

3.3. Calculation of Grey Relational Coefficient. Subsequently, the grey relational coefficient ζ_i can be calculated, according to the following equation [33]:

$$\zeta_i(k) = \frac{\min_i \min_k \left|x_0(k) - x_i(k)\right| + \rho \cdot \max_i \max_k \left|x_0(k) - x_i(k)\right|}{\left|x_0(k) - x_i(k)\right| + \rho \cdot \max_i \max_k \left|x_0(k) - x_i(k)\right|}, \tag{4}$$

where $|x_0(k) - x_i(k)|$ represents the absolute difference between the two sequences; ρ, the distinguishing coefficient, is usually 0.5, and it represents the significance of $\max_i \max_k |x_0(k) - x_i(k)|$. The smaller the ρ is, the higher its distinguishability is. Therefore, $\rho = 0.5$ was considered, as it can offer more moderate distinguishing effects and better stability [34].

3.4. Calculation of Grey Relational Grade (γ_i) and Rank of Results. As shown in equation (5), it is worth noting that the grey relational grade (γ_i) can be used for measuring the degree of correlation between the comparative matrix and the reference matrix. The γ_i being close to 1 indicates the

Table 2: Scaling factor of compressive strength of concrete under different dimensions.

Shape	Size (mm)	Scaling factor (P/P_0)
Cylinder	$d = 100$, $h = 200$	0.72
Cube	1 = 40	1.16
Cube	1 = 50	1.13
Cube	1 = 70	1.07
Cube	1 = 75	1.06
Cube	1 = 100	1.00
Cube	1 = 150	0.91

strong correlation, whereas γ_i being close to 0 depicts the weak correlation.

$$\gamma_i = \frac{1}{n}\sum_{k=1}^{n}\zeta_i(k), \quad k = 1, 2, \ldots, 600. \tag{5}$$

The size of the grey correlation grade γ_i objectively reflects the influencing degree of each factor on the compressive strength of concrete; therefore, choosing the influential and significant factors is important. In this study, MATLAB 7.0 was used to calculate γ_i of nine factors. According to the results of grey relational grade, the rank (cement type > sand ratio > W/B > MK content > slag content > curing age > dosage of water-reducing agent > Si/Al of MK > fly ash content) was obtained, as shown in Figure 2. However, too many input factors will decrease the calculation efficiency of the models and complicate the model architecture of the input layer and output layer pattern and thus will be unfavorable for the building models. Thus, the factors such as Si/Al of MK and fly ash content, considered as the weak factors, were not taken into account to optimize the model structure because of its low grey correlation grade ($\gamma_i < 0.9$).

After the abovementioned three steps of preprocessing and preanalysis of data source, i.e., unifying the compressive strength of concrete, normalization of data, and selection of main influencing factors, the datasets are ready for training and testing the prediction models. The characteristics of the prepared dataset are listed in Table 4.

4. Main Theories and Parameters of ANN, GA-ANN, and XGBoost

4.1. Artificial Neural Network (ANN)

4.1.1. Artificial Neuron. ANN is inspired by understanding of the biological nervous system [35] and is a mathematical model that imitates the behavioral feature of the human neural network for distributed and parallel processing of information. These are massively parallel complex systems made up of many processing neurons connected by connection weights (w_i), as shown in Figure 3.

A single neuron's structure is simple. However, a complete network system containing a mass of basic neurons can be skilled in processing nonlinear problems. Figure 3 shows the notational convention where the orange graphics represents a computational unit and X_i is an input variable and multiplied by the respective weight W_i. After that, the results combined with bias term b will be transferred to every neuron in the next layer of the network. The summed values (one per neuron in the layer) are then forwarded to the next layer of the network usually *via* an activation function such as sigmoid function, as shown in equation (6) [36]. The output from the activation function Y can be determined by equation (7), and it will act as the final result or the input variable of the next layer.

$$f(x) = \frac{1}{1 + e^{-x}}, \tag{6}$$

$$Y = f\left(\sum_{i=1}^{n} W_i X_i + b\right). \tag{7}$$

4.1.2. Feed-Forward Neural Network and Backpropagation Algorithm. The feed-forward neural network is one of the simplest neural networks. Each neuron is only connected to the one in the previous layer. Take the output of the previous layer and output it to the next layer. There is no feedback between the layers. It is one of the most widely used and rapidly developed ANNs [37]. Commonly, the feed-forward neural network consists of one input layer, several hidden layers, and one output layer of neurons [38]. The original data are considered as the input information accepted by the first layer (input layer) and then is transmitted to the corresponding neurons of the second layer (hidden layer); at last, it will pass forward to the output layer. Besides, the number of input nodes can be regarded as the input parameters (main influencing factors mentioned previously), and the output node is the target prediction result. Figure 4 shows the framework of the ANN with 7 input nodes and 1 output node used in this study. Notably, in terms of empirical and experimental data types, the number of neurons of the hidden layer needs to be debugged multiple times for creating an optimal predictive neural network model.

There is no feedback process between the layers in the abovementioned ANN model. In order to get a better prediction model, some algorithms such as the frequently used backpropagation (BP) algorithm were added to the ANN for adjusting parameters. The essence of the BP algorithm is decreasing the difference between the actual result and the desirable result in each output layer by the fine adjustment of the weights in the back-propagated process. When the BP algorithm is injected into the forward-feeding neural network, the feed-forward neural network shows excellent performance such as strong generalization ability, fast convergence, and simple structure [39].

During the process of back-propagation, all of the weights should be updated by the BP algorithm in the neural network according to the following formula [40]:

$$W_{ji}(n+1) = W_{ji}(n) + \eta \cdot \delta_j \cdot X_i + \cdot \Delta W_{ji}(n). \tag{8}$$

In equation (8), n is the number of times, η is the learning rate, δ_j means the error signal, and α is the expression of

Table 3: Mathematical forms for the construction of the reference matrix and the comparative matrix.

Compressive strength: $X_0\ (k) = X_0\ (1), X_0\ (2), X_0\ (3), \ldots, X_0\ (600)$	Reference matrix
Cement type: $X_1\ (k) = X_1\ (1), X_1\ (2), X_1\ (3), \ldots, X_1\ (600)$	
W/B: $X_2\ (k) = X_2\ (1), X_2\ (2), X_2\ (3), \ldots, X_2\ (600)$	
Sand ratio: $X_3\ (k) = X_3\ (1), X_3\ (2), X_3\ (3), \ldots, X_3\ (600)$	
Dosage of water-reducing agent: $X_4\ (k) = X_4\ (1), X_4\ (2), X_4\ (3), \ldots, X_4\ (600)$	
Curing age: $X_5\ (k) = X_5\ (1), X_5\ (2), X_5\ (3), \ldots, X_5\ (600)$	Comparative matrix
Slag content: $X_6\ (k) = X_6\ (1), X_6\ (2), X_6\ (3), \ldots, X_6\ (600)$	
Fly ash content: $X_7\ (k) = X_7\ (1), X_7\ (2), X_7\ (3), \ldots, X_7\ (600)$	
MK content: $X_8\ (k) = X_8\ (1), X_8\ (2), X_8\ (3), \ldots, X_8\ (600)$	
Si/Al of MK: $X_9\ (k) = X_9\ (1), X_9\ (2), X_9\ (3), \ldots, X_9\ (600)$	

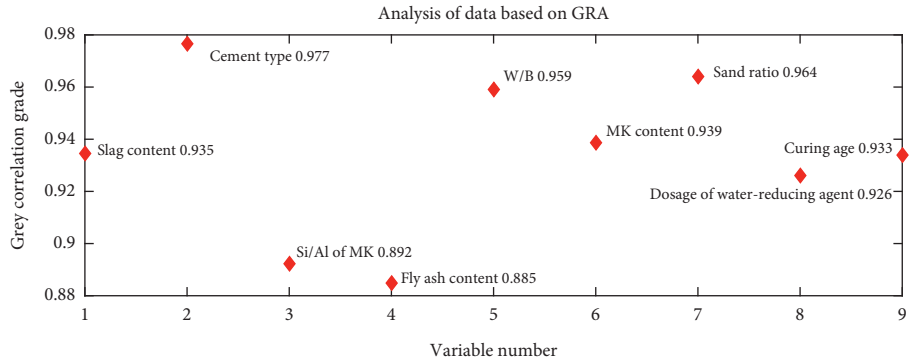


Figure 2: The grey correlation grade values of nine factors.

Table 4: The characteristics of dataset (600 groups of data selected from 18 research articles).

Factors	Minimum values	Maximum values	Average	Standard deviation
W/B	0.18	0.60	0.35	0.10
Sand ratio (%)	0.33	1	0.45	0.09
Cement type (four types)	1	4	—	—
Slag content (%)	0	20	0.83	0.83
MK content (%)	0	20	9.01	6.90
Dosage of water-reducing agent (%)	0	3.5	0.83	0.84
Curing age (days)	1	180	37.37	43.48
Compressive strength (MPa)	8.93	130.85	68.55	22.68

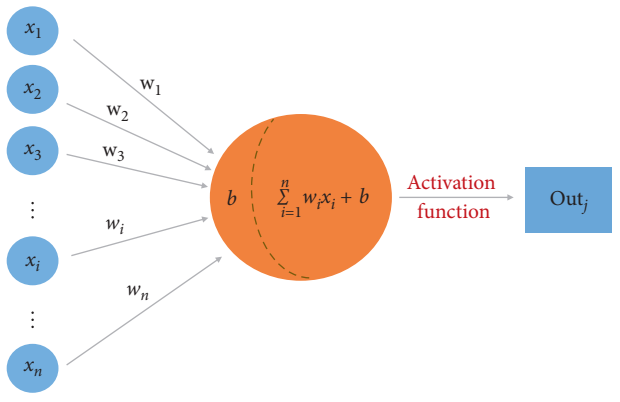


Figure 3: A simple artificial neuron model.

momentum factor. The difference between the target value and the true value can be evaluated by the error signal δ_j for an output unit j [41]:

$$\delta_j = \left(t_j - X_j\right) \cdot X_j \cdot \left(1 - X_j\right). \tag{9}$$

The error signal δ_j is a function of the error signals of those units in the next higher layer, connected to unit j and the weights of those connections [9]:

$$\delta_j = X_j \cdot \left(1 - X_j\right) \cdot \sum_k \delta_k W_{kj}. \tag{10}$$

As shown in equation (10), it is worth noting that the k_{th} layer means the layer in front of the j_{th} layer. The training procedure should be working until the iterative process converges. The updated connection weights are captured from the final trained network [42].

Although the BP algorithm has been widely used, it still has some shortcomings, such as long training time and easy to fall into local minimum. Therefore, some optimized algorithms were developed by researchers for combining with the ANN model to get a better prediction performance.

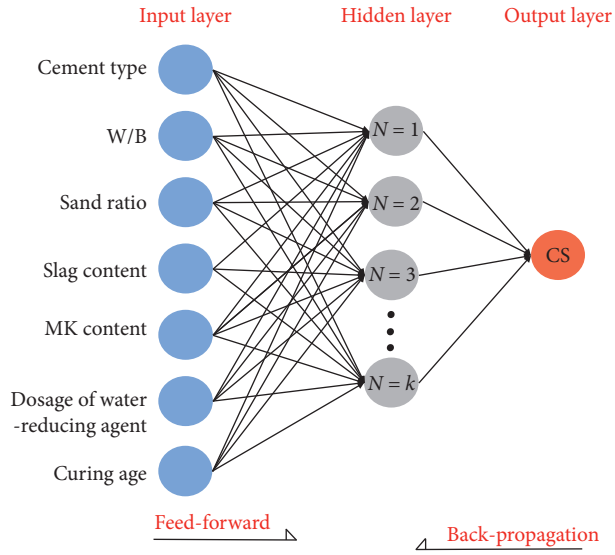


Figure 4: Feed-forward and backpropagation of ANN.

5. Parameters Optimized by Genetic Algorithm

5.1. Genetic Algorithm. The training process of ANN is a procedure of optimization of model parameters, and it starts from a random initial solution and iteratively finds out the optimal parameter values. During each iteration, the error function is first calculated at the current gradient, and then the search direction is determined based on the gradient. However, on this occasion, the optimization of parameters is likely to sink into local minimum (e.g., using the BP algorithm). At this point, the genetic algorithm (GA) was chosen for training the networks to better approximate the global minimum [43].

As one of the most commonly used intelligent optimization algorithms, GA is a mathematical computation model simulating the natural selection of Darwin's biological evolution theory and the process of biological evolution of genetic variation. Moreover, it is a characteristic methodology inspired by the physical phenomenon of biological evolution (selection, crossover, and mutation) to seek the global optimal solutions in space. In the field of artificial intelligence, it produces significant influence on optimization main parameters of the artificial neural network (ANN) [44]. Thus, for creating a powerful model with a higher prediction accuracy, it is necessary to utilize the genetic algorithm to optimize initial weights and thresholds of the ANN model.

5.2. The Operating Procedure of GA. Figure 5 shows the main operation procedure of GA used for the optimization of ANN. GA's key elements such as population size, numbers of generation, and fitness function and key operation procedure including selection, evolution, and mutation directly affect the final optimization result, making them definitely important in the operating procedure.

In this study, GA was combined with ANN to get its desired weights and thresholds. During the optimization process, the initial weights and thresholds of the network could be construed as individuals with prediction errors of the initial ANN as its genetic information. Then, the genetic operation of selection, crossover, and mutation is repeated, to search the optimal individual, namely, the optimal weights and thresholds of ANN (GA-ANN).

6. The Extreme Gradient Boosting Methodology (XGBoost)

XGBoost is a machine learning algorithm based on decision tree with a process of gradient promotion [45]. It is one of the most successful machine learning algorithms at present and has been extensively used by researchers in some competitions of machine learning attributing to its excellent learning performance and fast calculation speed, even more used than the popular deep neural network method. In addition, the XGBoost algorithm is generally considered to be superior to ANNs in processing small- and medium-sized structured data although ANNs show excellent performance in analyzing unstructured model data such as pictures, text, audio, and video.

In a word, the main calculation process of XGBoost is an accumulation of iterative results after T times, as shown in equation (11) [46], where i is the number of samples; T is the number of decision trees, and $\hat{\mathbf{y}}_{\mathbf{i}}^{(\mathbf{T})}$ is the final predicted value of the i_{th} sample in the Decision tree with number T. Function $\mathbf{f_t}(\mathbf{x_i})$ represents the calculation formula of the i_{th} sample in the T decision tree equation (12)), ω is the weight vector corresponding to the leaf node, and $q(x_i)$ is a function of the feature vector $\mathbf{x_i}$ mapped to the leaf node of the decision tree.

$$\hat{\mathbf{y}}_{\mathbf{i}}^{(\mathbf{T})} = \hat{\mathbf{y}}_{\mathbf{i}}^{(0)} + \sum_{\mathbf{t}=1}^{\mathbf{T}} \mathbf{f_t}(\mathbf{x_i}), \ \textbf{where}\ \mathbf{f}_0(\mathbf{x_i}) = \hat{\mathbf{y}}_{\mathbf{i}}^{(0)} = 0, \tag{11}$$

$$\mathbf{f_t}(\mathbf{x_i}) = \boldsymbol{\omega}_{\mathbf{q}(\mathbf{x_i})}. \tag{12}$$

To better understand the running process of the XGBoost model, the training process of a traditional decision tree was first introduced, as shown in Figure 6, indicating W/B, curing age, and sand ratio (not all the main factors for simple example), as the decision items for processing the data of four samples. The training process steps are as follows: (i) sample 1 is screened out when the limit is W/B > 0.4; (ii) sample 3 is separated from other samples for curing age ≤3; (iii) samples 2 and 4 are divided by the sand ratio of 0.3; and (iv) a traditional decision tree completes an accurate prediction model with some potential problems.

Figures 7–9 show the training process of the XGBoost model. Figure 7 represents the Tree 1 with W/B as the discriminant condition, similar to the first branch of Figure 6. The difference is that the output results of the leaf nodes are averaged to represent the training output values of the four samples (i.e., $\omega_{1(\mathbf{x_i})}$, where $i = 1, 2, 3, 4$), and then the training error of each sample is obtained by calculating the deviation of the predicted value from the true value. The obtained errors in Tree 1 will be used as the input values of Tree 2 (Figure 8) and the curing age (≤3 or >3) will act as the

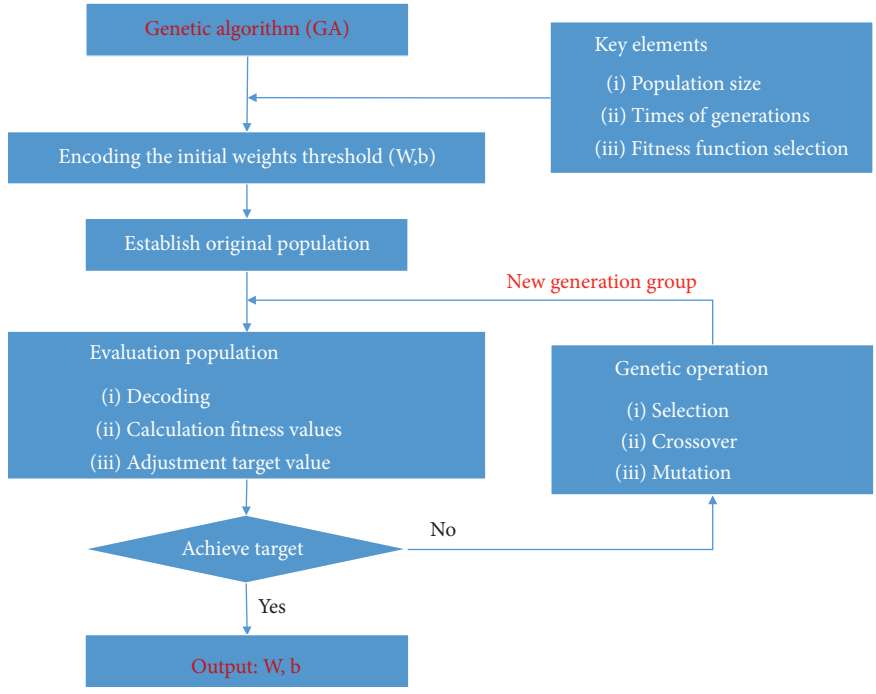
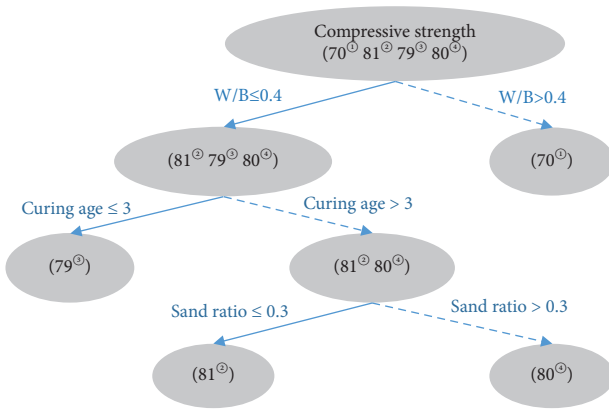


Figure 5: The working principle of GA in the optimal procedure.



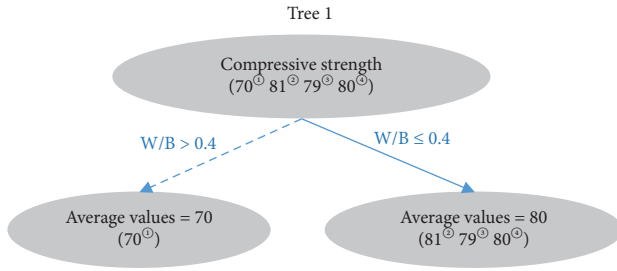
Factors	Training set				Testing set
	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
W/B	0.45	0.4	0.4	0.4	0.4
Sand ratio	0.4	0.21	0.21	0.35	0.21
Cement type (four types)	3	3	3	3	3
Slag content (%)	10	20	10	0	10
MK content (%)	0	0	0	0	0
Dosage of water-reducing agent (%)	2	2	2	1	2
Curing age (days)	3	7	3	28	7
Compressive strength (MPa)	70①	81②	79③	80④	Unknown

Note: $70^1\,81^2\,79^3\,80^4$ in the decision tree on the left represents the compressive strength of the first, second, third and fourth samples respectively

Figure 6: The training process of a traditional decision tree.

discrimination standard. Then, the value of $\omega_{2(\mathbf{x}_i)}$ will be obtained based on the average value of decided errors. The predictive values of samples were obtained by the summation of ω_1 and ω_2. The calculated training errors in Tree 2 will be used as input into Tree 3. Similarly, a decision tree based on sand ratio was established, as shown in Figure 9, and the average values are all 0, i.e., $\omega_{3(\mathbf{x}_i)} = 0$, where $i = 1, 2, 3, 4$. Then, the final predicted values of samples 1, 2, 3, and 4 were calculated by equation $\omega_{1(\mathbf{x}_i)} + \omega_{2(\mathbf{x}_i)} + \omega_{3(\mathbf{x}_i)}$ and found as 69.5, 80.5, 79.5, and 80.5 MPa, respectively. At this point, the model training is finished. You can find that the output values in XGBoost are different than those in the traditional decision tree. To be more specific, the predictive values of four samples are not equal to their true values in XGBoost and thus seem to be worse than the traditional decision tree; however, it can solve a big problem of traditional decision tree, i.e., overfitting.

For instance, the traditional decision tree and the XGBoost were used, respectively, to predict the strength value of sample 5, indicating that the traditional decision tree will output the strength value of 81 MPa and that by the XGBoost model is 80.5 MPa. The following analysis can prove the competitive prediction ability of XGBoost. Compared with sample 2, except that the slag content of sample 5 is less than sample 2, other variables have exactly the same values. In combination with the general conclusion of the influence of slag content on the concrete strength, the compressive strength of sample 5 was

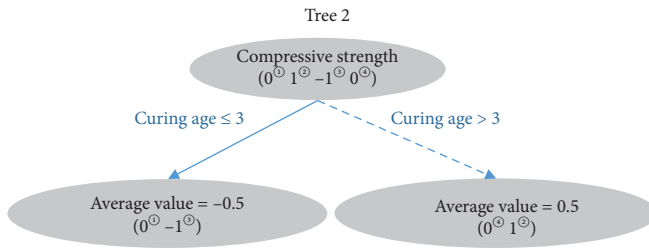


Serialnumber	1	2	3	4
W/B	0.45	0.4	0.4	0.4
$f_1(x_i) = \omega_1$	$70^{①}$	$80^{②}$	$80^{③}$	$80^{④}$
$\hat{y}^{(1)}_{\text{predict}}(\omega_1)$	$70^{①}$	$80^{②}$	$80^{③}$	$80^{④}$
y_{true}	$70^{①}$	$81^{②}$	$79^{③}$	$80^{④}$
Training error (Tree 1)	$0^{①}$	$1^{②}$	$-1^{③}$	$0^{④}$

$f_t(x) = \omega_{q(x)}$, $t = 3$ x–sample

Training error = $y_{\text{true}} - \hat{y}^{(t)}_{\text{predict}}$

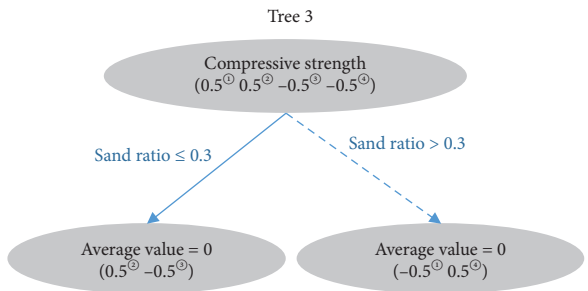
Figure 7: The training process of an XGBoost model (Tree 1).



Serial number	1	2	3	4
Curing age	3	7	3	28
$f_1(x_i) = \omega_1$	$70^{①}$	$80^{②}$	$80^{③}$	$80^{④}$
$f_2(x_i) = \omega_2$	$-0.5^{①}$	$0.5^{②}$	$-0.5^{③}$	$0.5^{④}$
$\hat{y}^{(2)}_{\text{predict}}(\omega_1 + \omega_2)$	$69.5^{①}$	$80.5^{②}$	$79.5^{③}$	$80.5^{④}$
y_{true}	$70^{①}$	$81^{②}$	$79^{③}$	$80^{④}$
Training error (Tree 2)	0.5	0.5	−0.5	−0.5

$f_t(x) = \omega_{q(x)}$, $t = 3$ x–sample

Figure 8: The training process of an XGBoost model (Tree 2).



Serial number	1	2	3	4
Sand ratio	0.4	0.21	0.21	0.35
$f_1(x_i) = \omega_1$	$70^{①}$	$80^{①}$	$80^{①}$	$80^{①}$
$f_2(x_i) = \omega_2$	$-0.5^{①}$	$0.5^{①}$	$-0.5^{①}$	$0.5^{①}$
$f_3(x_i) = \omega_3$	$0^{①}$	$0^{①}$	$0^{①}$	$0^{①}$
$\hat{y}^{(3)}_{\text{predict}}(\omega_1 + \omega_2 + \omega_3)$	$69.5^{①}$	$80.5^{①}$	$79.5^{①}$	$80.5^{①}$
y_{true}	$70^{①}$	$81^{①}$	$79^{①}$	$80^{①}$

$f_t(x) = \omega_{q(x)}$, $t = 3$ x–sample

Figure 9: The training process of an XGBoost model (Tree 3).

obtained as predicted to be less than 81 MPa. In the same way, it was easy to know that the strength value of sample 5 will be more than that of sample 3 (i.e., 79 MPa) on account of the increasing curing age of samples under the same mix proportion. Hence, the approximate range of the strength of sample 5 will be in the range 79–81 MPa, indicating that the predictive result of XGBoost (80.5 MPa) is more reasonable and the predictive performance of XGBoost is superior than that of the traditional decision tree. This phenomenon is caused by the overfitting training process of the traditional decision tree. It is sometimes inclined to focus on the characteristics of the training data and summarizes the internal rules of the training set (four samples). However, for the new test dataset (sample 5), it cannot reveal its internal change rule. Therefore, a single decision tree model is not generalizable.

7. Setting Structural Parameters and Learning Rate of ANN, GA-ANN, and XGBoost

In this study, the prediction performances of ANN, GA-ANN, and XGBoost were tested and compared. In order to avoid the influence of network structures on the ability of prediction models, the structural parameters of ANN and GA-ANNs were fixed, namely, with the same hidden layers and nodes, as listed in Table 5. Because XGBoost is a definitely different model with a tree structure compared with ANN and GA-ANN, it has different parameters, but some parameters similar to those in ANN and GA-ANN were set.

The programming of ANN, GA-ANN, and XGBoost was performed on the MATLAB 7.0 platform, combined with the dataset (600 groups of data selected from 18 research papers) prepared for the training and testing prediction models.

Table 5: Setting parameters of ANN, GA-ANN, and XGBoost.

Structural parameters	ANNs	GA-ANNs	Hyperparameters	XGBoost
Input layer nodes	6	6	Maximum depth	6
Hidden layers	2	2	Minimum child weight	1
Hidden layer nodes	80 (first) 60 (second)	80 (first) 60 (second)	Gamma	0
Output layer nodes	1	1	Subsample	1
Training parameters			Colsample by tree	1
Epoch times	10	10	Regularization alpha	0
Initial learning rate	0.01	0.01	Initial learning rate	0.01
Maximum iterations	100	100	Number of iterations	30

The prediction ability of those models was tested by comparing the value of absolute fraction of variance (R^2), root mean squared error (RMSE), and mean absolute percentage error (MAPE). The detailed results are as follows.

8. Results and Discussion

8.1. Comparison of R^2, RMSE, and MAPE of ANN, GA-ANN, and XGBoost with Random Testing Dataset. To train and test those prediction models, 600 groups of data after preprocessing were randomly divided into a training dataset (containing 480 groups of data, i.e., 80% of the total dataset) and a testing dataset (containing 120 groups of data, i.e., 20% of total dataset). The values of the absolute fraction of variance (R^2), root mean squared error (RMSE), and mean absolute percentage error (MAPE) were adopted to comprehensively evaluate the prediction performances of these models. The detailed equations are as follows:

$$\mathbf{R^2} = \frac{\left(\mathbf{n}\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{t_i O_i} - \sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{t_i} \sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{O_i}\right)^2}{\left(\mathbf{n}\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{t_i^2} - \left(\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{t_i}\right)^2\right)\left(\mathbf{n}\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{O_i^2} - \left(\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{O_i}\right)^2\right)},$$

$$\mathbf{RMSE} = \sqrt{\frac{1}{\mathbf{n}}\sum_{\mathbf{i=1}}^{\mathbf{n}} \left(\mathbf{t_i} - \mathbf{O_i}\right)^2},$$

$$\mathbf{MAPE} = \frac{1}{\mathbf{n}}\left[\frac{\sum_{\mathbf{i=1}}^{\mathbf{n}} \left|\mathbf{t_i} - \mathbf{O_i}\right|}{\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{t_i}} \times 100\right], \tag{13}$$

where t_i is the true value, O_i is the predictive value, and n is the total number [41, 47].

Figure 10 shows the values of R^2, RMSE, and MAPE of the three prediction models by using the test dataset. It is easy to find that the values of R^2, RMSE, and MAPE of ANN are 0.708, 13.38 MPa, and 9.72%, respectively. With the parameters optimized by the genetic algorithm (GA), the prediction performance is better attributing to the better values of R^2 (0.837), RMSE (11.74 MPa), and MAPE (8.26%) of GA-ANN. The third image shows the prediction results of XGBoost. As we know, XGBoost is often known to deal with the classifying problems owing to its excellent performance, but surprisingly, it has also achieved good performance for dealing with regression prediction problems. As shown in Figure 10 (the third), the data points are more concentrated on the line $y = x$ and nearly without the deviation beyond ± 25 MPa (the red dotted lines) of predictive values compared with target values. The results of R^2 (0.872), RMSE (8.62 MPa), and MAPE (8.25%) are also good, better than those obtained by GA-ANN.

The above analysis and discussion indicate superior prediction performance of XGBoost than those by ANN and GA-ANN; however, the results were obtained based on the random testing dataset, indicating that the dataset applied to XGBoost is different with those of ANN and GA-ANN. A special case where the dataset applied to XGBoost is easy to predict was tested, exhibiting good result. In order to eliminate the possibility of that special case and get a more persuasive conclusion, the R^2, RMSE, and MAPE were tested by using the same testing dataset as described in the following section.

8.2. Comparison of R^2, RMSE, and MAPE of ANN, GA-ANN, and XGBoost with the Same Testing Dataset. The R^2, RMSE, and MAPE values were calculated thrice by using the three datasets, i.e., dataset 1, 2, and 3 with 120 data in each dataset generated by random selection. In order to easily compare the deviation degree of predictive values (red hollow points) and true values (black solid points) of ANN, GA-ANN, and XGBoost, the diagrams of the results are presented side-by-side (as shown in Figure 11). It is obvious that the predictive values of XGBoost have the maximal degree of coincidence with true values shown in Figure 11, and the degree of coincidence of those points in the ANN is the lowest. It is a subjective conclusion; therefore, the objective results of ANN, GA-ANN, and XGBoost were compared. The specific R^2 values obtained by the ANN, GA-ANN, and XGBoost are 0.708, 0.833, and 0.874, respectively. Second, for the same 120 groups of testing data, the RMSE values obtained by the ANN, GA-ANN, and XGBoost are 13.377, 9.413, and 10.982 MPa, respectively. Lastly, the MAPE values obtained by the ANN, GA-ANN, and XGBoost are 9.733%, 9.289%, and 9.544%, respectively.

The values in Figure 11 indicate that only the R^2 value obtained by XGBoost is the highest, whereas the RMSE and MAPE obtained by the GA-ANN are the lowest than those obtained by the ANN and XGBoost, indicating better prediction performance of GA-ANN than that of XGBoost. However, the results obtained from dataset 2 and dataset 3, as shown in Figures 12 and 13, respectively, show different conclusions. For getting the comprehensive results of comparison, the average R^2, RMSE, and MAPE values were calculated by the ANN, GA-ANN, and XGBoost. Table 6

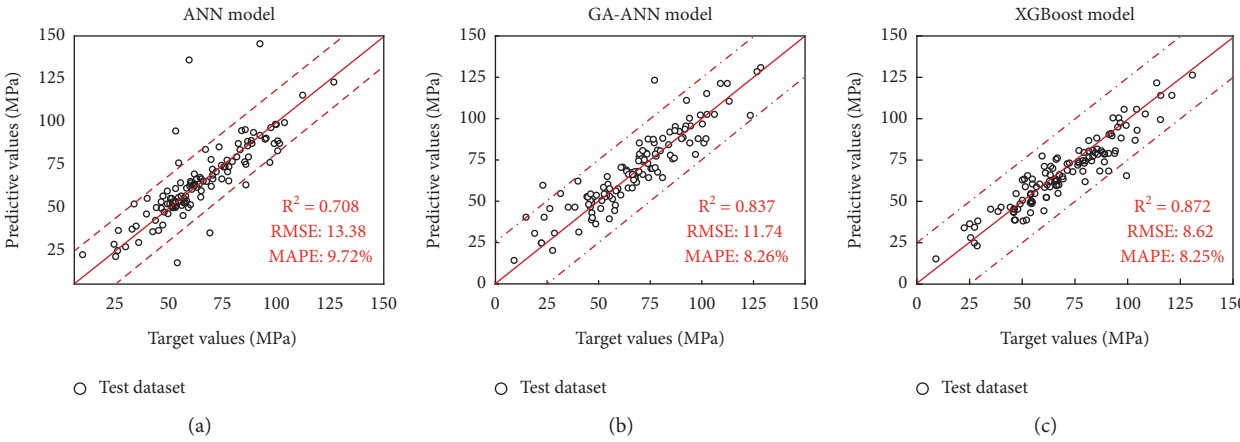


Figure 10: Prediction precision of three models for test datasets.

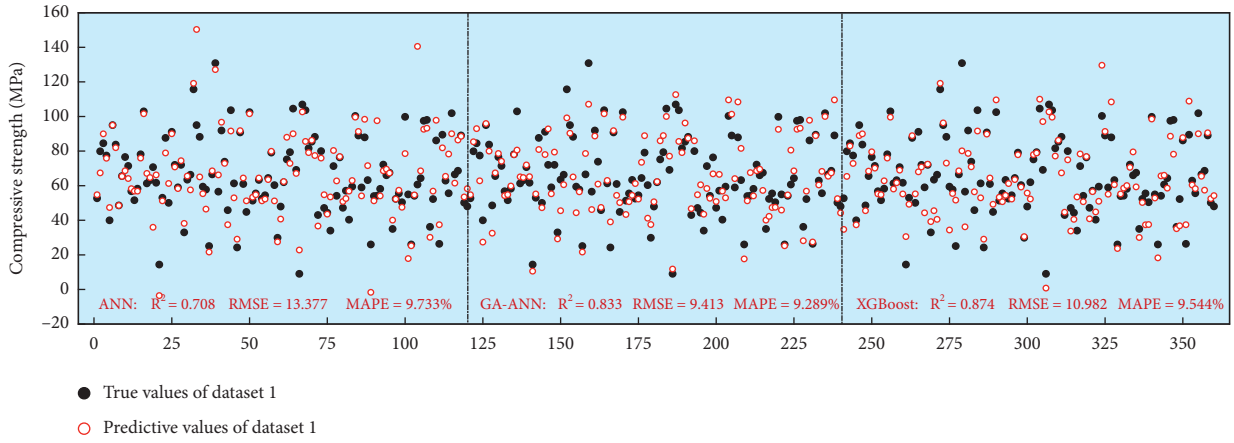


Figure 11: Predictive and true values of ANN, GA-ANN, and XGBoost for dataset 1.

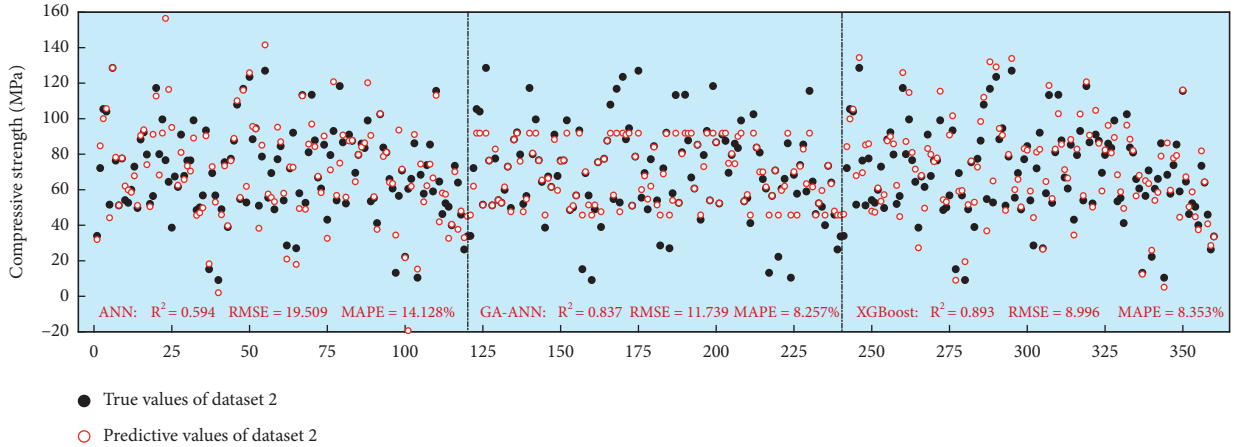


Figure 12: Predictive and true values of ANN, GA-ANN, and XGBoost for dataset 2.

shows the average R^2 values (i.e., $\overline{\mathbf{R}}$) of the ANN, GA-ANN, and XGBoost as 0.674, 0.829, and 0.880, respectively, indicating that the highest absolute fraction of variance (R^2) by the XGBoost. The RMSE and MAPE values obtained by XGBoost were the best. The average RMSE value of the ANN, GA-ANN, and XGBoost was 15.569, 10.530, and 9.532 MPa, respectively, and the average MAPE value of the ANN, GA-ANN, and XGBoost was 11.224%, 9.140%, and 8.718%, respectively, certainly verifying the superior prediction performance of XGBoost.

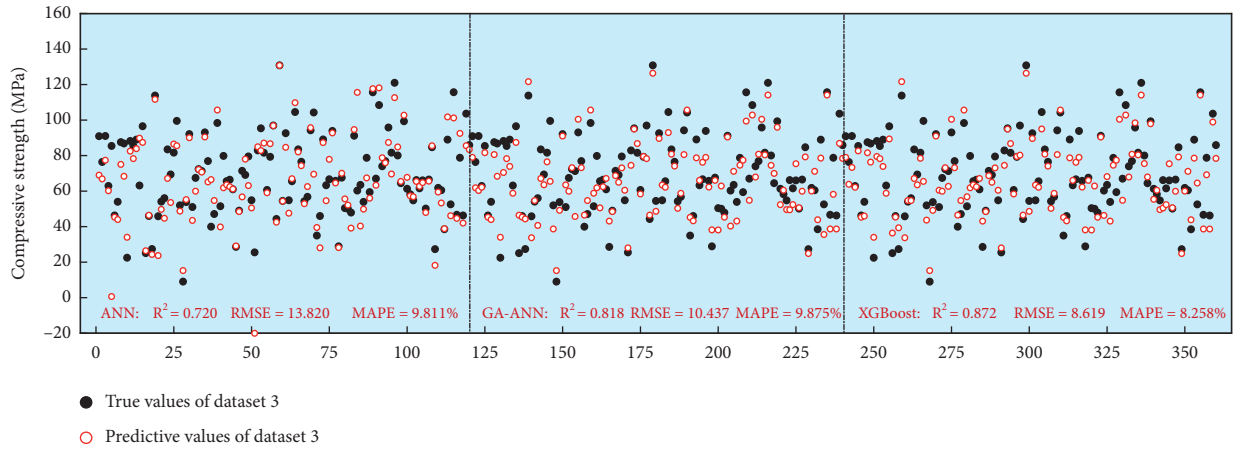


Figure 13: Predictive and true values of ANN, GA-ANN, and XGBoost for dataset 3.

Table 6: Individual and average values of the R^2, RMSE, and MAPE of ANN, GA-ANN, and XGBoost.

	R^2		RMSE (MPa)		MAPE (%)	
	R_i $(i = 1, 2, 3)$	$\overline{R}$	M_i $(i = 1, 2, 3)$	$\overline{M}$	A_i $(i = 1, 2, 3)$	$\overline{A}$
ANN	0.708	0.674	13.377	15.569	9.733	11.224
	0.594		19.509		14.128	
	0.720		13.820		9.811	
GA-ANN	0.833	0.829	9.413	10.530	9.289	9.140
	0.837		11.739		8.257	
	0.818		10.437		9.875	
XGBoost	0.874	0.880	10.982	9.532	9.544	8.718
	0.893		8.996		8.353	
	0.872		8.619		8.258	

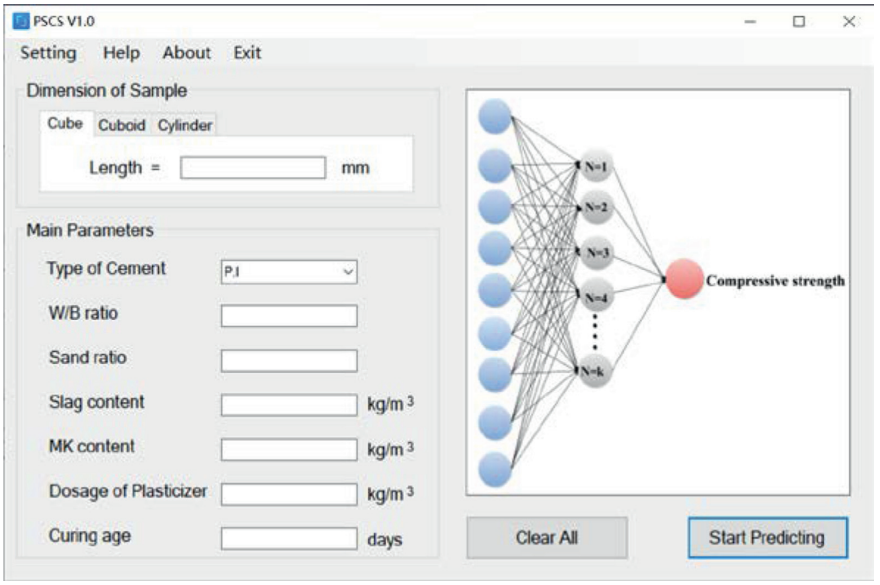


Figure 14: Interface for inputting parameters.

The above analyses and discussions of the results obtained by using the random testing dataset and three same testing datasets clearly indicate more stable prediction performance of XGBoost with the highest prediction accuracy. Therefore, XGBoost is a recommended prediction model and could be used for predicting the compressive strength of concrete containing slag and metakaolin.

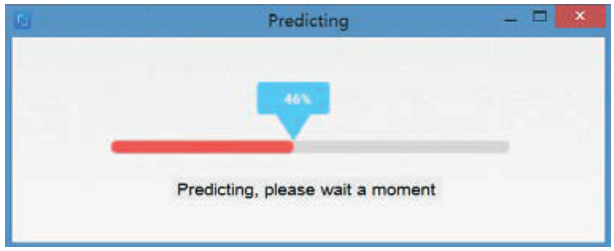


Figure 15: The procedure of prediction.

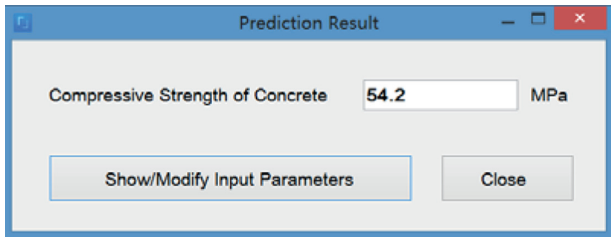


Figure 16: The results of compressive strength prediction.

9. Application Software for Easy Prediction Based on XGBoost and GRA

At present, the design and development of application software based on various prediction models is receiving increasing attention. This behavior can encourage to develop software from theory to practical application. Therefore, in this study, C# programming was utilized for integrating the involved research results and methods into an application software (named PSCS V1.0) with straightforward interactive interfaces to help users without the knowledge of MATLAB to predict and analyze easily and efficiently. With the help of PSCS, users can just input the dimension of sample (unifying the compressive strength of concrete under different dimensions automatically) and main parameters (the parameters obtained by GRA, i.e., type of cement, W/B ratio, sand ratio, slag content, MK content, dosage of water-reducing agent, and curing age), as shown in Figure 14. After a short wait time for computing of XGBoost (Figure 15), the final prediction results obtained are shown in Figure 16.

PSCS is a short but strong system and could be used for predicting the compressive strength of concrete containing slag and metakaolin. As the dataset used for the training XGBoost model was collected from various journal articles, it can ensure the applicability for different cases. In the following study, relevant literature data were collected continually for building up a complete and more comprehensive database.

10. Conclusions

In conclusion, the proposed novel predictive system XGBoost was successfully applied to predict the compressive strength of concrete containing slag and metakaolin, and its performance was compared with the commonly used ANN and optimized GA-ANN models. The datasets used for the training and testing three models were selected from 18 research articles (containing 600 groups of data), with the treatments of unifying the compressive strength of concrete, normalization of data, and selection of main influencing factors. With the help of GRA, the main input parameters such as the type of cement, W/B ratio, sand ratio, slag content, MK content, dosage of water-reducing agent, and curing age were determined scientifically and reasonably. After that, the prediction models were trained and tested by calculating the values of R^2, RMSE, and MAPE. The main results and conclusions can be drawn as follows:

(1) For the random testing dataset, the R2, RMSE, and MAPE obtained by the ANN are 0.708, 13.38 MPa, and 9.72%, respectively, and those obtained by the GA-ANN are 0.837, 11.74 MPa, and 8.26%, respectively, and the results obtained by XGBoost are 0.872, 8.62 MPa, and 8.25%, respectively, proving a superior prediction performance of XGBoost than those of ANN and GA-ANN.

(2) XGBoost also exhibited superior prediction performance for prediction of the models based on the same testing dataset. The average R^2 values of ANN, GA-ANN, and XGBoost are 0.674, 0.829, and 0.880, respectively. For the RMSE and MAPE, the average RMSE values of the ANN, GA-ANN, and XGBoost are 15.569, 10.530, and 9.532 MPa, respectively, and the average MAPE values of the ANN, GA-ANN, and XGBoost are 11.224%, 9.140%, and 8.718%, respectively. Therefore, XGBoost was the best prediction model for testing the same datasets.

(3) Based on the analyses and discussions obtained by using the random testing dataset and the same testing datasets, XGBoost exhibited better performance than the ANN and GA-ANN for prediction capability. Therefore, XGBoost is the best method for predicting the compressive strength of concrete containing slag and metakaolin in this study. Lastly, in order to make XGBoost user friendly and easy to operate, the application software (PSCS V1.0) was encoded by C# in this study, making it a short but strong system that could predict the compressive strength of concrete containing slag and metakaolin efficiently.

Data Availability

The data used in this paper are from references, and the corresponding literature sources have been marked, so the data have reliability and authenticity.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors would like to express gratitude to all those who helped them during the data analysis and writing of this thesis. This work was supported by the National Natural

Science Foundation of China (no. 51828201), China Postdoctoral Science Foundation (no. 2018M632518), Anhui Postdoctoral Science Foundation (nos. 2017B150 and 2018B248), Natural Science Foundation of Anhui University (no. KJ2018A0074), and Key Research and Development Program Project of Anhui Province (no. 201904a07020081).

References

[1] C.-S. Poon, L. Lam, S. C. Kou, Y.-L. Wong, and R. Wong, "Rate of pozzolanic reaction of metakaolin in high-performance cement pastes," *Cement and Concrete Research*, vol. 31, no. 9, pp. 1301–1306, 2001.

[2] D. M. Roy, P. Arjunan, and M. R. Silsbee, "Effect of silica fume, metakaolin, and low-calcium fly ash on chemical resistance of concrete," *Cement and Concrete Research*, vol. 31, no. 12, pp. 1809–1813, 2001.

[3] L. Kanamarlapudi, K. B. Jonalagadda, D. C. K. Jagarapu, and A. Eluru, "Different mineral admixtures in concrete: a review," *SN Applied Sciences*, vol. 2, no. 4, 2020.

[4] F. Özcan, C. D. Atiş, O. Karahan, E. Uncuoğlu, and H. Tanyildizi, "Comparison of artificial neural network and fuzzy logic models for prediction of long-term compressive strength of silica fume concrete," *Advances in Engineering Software*, vol. 40, no. 9, pp. 856–863, 2009.

[5] S. Safarzadegan Gilan, H. Bahrami Jovein, and A. A. Ramezanianpour, "Hybrid support vector regression-particle swarm optimization for prediction of compressive strength and RCPT of concretes containing metakaolin," *Construction and Building Materials*, vol. 34, pp. 321–329, 2012.

[6] H. Huang, X. Wei, and Y. Zhou, "Twin support vector machines: a survey," *Neurocomputing*, vol. 300, pp. 34–43, 2018.

[7] G. Zhang, B. Eddy Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks:," *International Journal of Forecasting*, vol. 14, no. 1, pp. 35–62, 1998.

[8] C. Bilim, C. D. Atiş, H. Tanyildizi, 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.

[9] M. Sarıdemir, "Prediction of compressive strength of concretes containing metakaolin and silica fume by artificial neural networks," *Advances in Engineering Software*, vol. 40, no. 5, pp. 350–355, 2009.

[10] E. Aarts and J. Korst, *Simulated Annealing and Boltzmann Machines*, John Wiley & Sons, New York, NY, USA, 1988.

[11] C. Charalambous, "Conjugate gradient algorithm for efficient training of artificial neural networks," *Institute of Electrical and Electronics Engineers Proceedings G Circuits, Devices and Systems*, vol. 139, no. 3, pp. 301–310, 1992.

[12] J. H. Holland, *Genetic Algorithms and Adaptation*, Springer US, New York, NY, USA, 1984.

[13] A. Firouzi and A. Rahai, "An integrated ANN-GA for reliability based inspection of concrete bridge decks considering extent of corrosion-induced cracks and life cycle costs," *Scientia Iranica*, vol. 19, no. 4, pp. 974–981, 2012.

[14] G. Dror, M. Boulle, and I. Guyon, *The 2009 Knowledge Discovery and Data Mining Competition (KDD Cup 2009): Challenges in Machine Learning*, Microtome Publishing, Brookline, MA 02446, USA, 2011.

[15] R. L. Wilson and R. Sharda, "Bankruptcy prediction using neural networks," *Decision Support Systems*, vol. 11, no. 5, pp. 545–557, 1994.

[16] F. Liu, Z. You, X. Yang, and H. Wang, "Macro-micro degradation process of fly ash concrete under alternation of freeze-thaw cycles subjected to sulfate and carbonation," *Construction and Building Materials*, vol. 181, pp. 369–380, 2018.

[17] P. Duan, Z. Shui, W. Chen, and C. Shen, "Effects of metakaolin, silica fume and slag on pore structure, interfacial transition zone and compressive strength of concrete," *Construction and Building Materials*, vol. 44, pp. 1–6, 2013.

[18] S. Sujjavanich, P. Suwanvitaya, D. Chaysuwan, and G. Heness, "Synergistic effect of metakaolin and fly ash on properties of concrete," *Construction and Building Materials*, vol. 155, pp. 830–837, 2017.

[19] E. Güneyisi, M. Gesoğlu, S. Karaoğlu, and K. Mermerdaş, "Strength, permeability and shrinkage cracking of silica fume and metakaolin concretes," *Construction and Building Materials*, vol. 34, no. 34, pp. 120–130, 2012.

[20] E. Güneyisi, M. Gesoğlu, and K. Mermerdaş, "Improving strength, drying shrinkage, and pore structure of concrete using metakaolin," *Materials and Structures*, vol. 41, pp. 937–949, 2008.

[21] L. I. Yunfeng, M. A. Jiandong, S. Zhang, L. I. Qianglong, and D. Wang, *Effects of Metakaolin Admixture on the Compressive Strength of High-Strength Concrete*, China Sciencepaper, Beijing, China, 2017.

[22] A. A. Ramezanianpour and H. Bahrami Jovein, "Influence of metakaolin as supplementary cementing material on strength and durability of concretes," *Construction and Building Materials*, vol. 30, pp. 470–479, 2012.

[23] M. Shekarchi, A. Bonakdar, M. Bakhshi, A. Mirdamadi, and B. Mobasher, "Transport properties in metakaolin blended concrete," *Construction and Building Materials*, vol. 24, no. 11, pp. 2217–2223, 2010.

[24] Z. Shi, Z. Shui, Q. Li, and H. Geng, "Combined effect of metakaolin and sea water on performance and microstructures of concrete," *Construction and Building Materials*, vol. 74, pp. 57–64, 2015.

[25] M. J. Shannag and A. Yeginobali, "Properties of pastes, mortars and concretes containing natural pozzolan," *Cement and Concrete Research*, vol. 25, no. 3, pp. 647–657, 1995.

[26] N. Shafiq, M. F. Nuruddin, S. U. Khan, and T. Ayub, "Calcined kaolin as cement replacing material and its use in high strength concrete," *Construction and Building Materials*, vol. 81, pp. 313–323, 2015.

[27] J. M. Khatib, "Metakaolin concrete at a low water to binder ratio," *Construction and Building Materials*, vol. 22, no. 8, pp. 1691–1700, 2008.

[28] H. El-Diadamony, A. A. Amer, T. M. Sokkary, and S. El-Hoseny, "Hydration and characteristics of metakaolin pozzolanic cement pastes," *HBRC Journal*, vol. 14, no. 2, pp. 150–158, 2018.

[29] K. Mermerdaş, M. Gesoğlu, E. Güneyisi, and T. Özturan, "Strength development of concretes incorporated with metakaolin and different types of calcined kaolins," *Construction and Building Materials*, vol. 37, pp. 766–774, 2012.

[30] P. Perez-Cortes and J. I. Escalante-Garcia, "Metakaolin based geopolymers with high limestone contents–statistical modeling of strength and environmental and cost analyses," *Cement and Concrete Composites*, vol. 106, Article ID 103450, 2019.

[31] P. Dinakar, P. K. Sahoo, and G. Sriram, "Effect of metakaolin content on the properties of high strength concrete," *International Journal of Concrete Structures and Materials*, vol. 7, no. 3, p. 215, 2013.

[32] J. Sola and J. Sevilla, "Importance of input data normalization for the application of neural networks to complex industrial problems," *Institute of Electrical and Electronics Engineers Transactions on Nuclear Science*, vol. 44, no. 3, pp. 1464–1468, 1997.

[33] J. Xu, X. Zhao, Y. Yu, T. Xie, G. Yang, and J. Xue, "Parametric sensitivity analysis and modelling of mechanical properties of normal- and high-strength recycled aggregate concrete using grey theory, multiple nonlinear regression and artificial neural networks," *Construction and Building Materials*, vol. 211, pp. 479–491, 2019.

[34] T. C. Chang and S. J. Lin, "Grey relation analysis of carbon dioxide emissions from industrial production and energy uses in Taiwan," *Journal of Environmental Management*, vol. 56, no. 4, pp. 247–257, 1999.

[35] F. Altun, Ö. Kişi, and K. Aydin, "Predicting the compressive strength of steel fiber added lightweight concrete using neural network," *Computational Materials Science*, vol. 42, no. 2, pp. 259–265, 2008.

[36] J.-R. Zhang, J. Zhang, T.-M. Lok, and M. R. Lyu, "A hybrid particle swarm optimization-back-propagation algorithm for feedforward neural network training," *Applied Mathematics and Computation*, vol. 185, no. 2, pp. 1026–1037, 2007.

[37] B. B. Adhikary and H. Mutsuyoshi, "Prediction of shear strength of steel fiber RC beams using neural networks," *Construction and Building Materials*, vol. 20, no. 9, pp. 801–811, 2006.

[38] A. Mukherjee and S. Nag Biswas, "Artificial neural networks in prediction of mechanical behavior of concrete at high temperature," *Nuclear Engineering and Design*, vol. 178, no. 1, pp. 1–11, 1997.

[39] J. Bai, S. Wild, J. A. Ware, and B. B. Sabir, "Using neural networks to predict workability of concrete incorporating metakaolin and fly ash," *Advances in Engineering Software*, vol. 34, no. 11-12, pp. 663–669, 2003.

[40] T. P. Vogl, J. K. Mangis, A. K. Rigler, W. T. Zink, and D. L. Alkon, "Accelerating the convergence of the back-propagation method," *Biological Cybernetics*, vol. 59, no. 4-5, pp. 257–263, 1988.

[41] M. Sarıdemir, "Predicting the compressive strength of mortars containing metakaolin by artificial neural networks and fuzzy logic," *Advances in Engineering Software*, vol. 40, no. 9, pp. 920–927, 2009.

[42] R. Ince, "Prediction of fracture parameters of concrete by artificial neural networks," *Engineering Fracture Mechanics*, vol. 71, no. 15, pp. 2143–2159, 2004.

[43] D. E. Goldberg and J. H. Holland, "Genetic algorithms and machine learning," *Machine Learning*, vol. 3, no. 2/3, pp. 95–99, 1988.

[44] R. Rebouh, B. Boukhatem, M. Ghrici, and A. Tagnit-Hamou, "A practical hybrid NNGA system for predicting the compressive strength of concrete containing natural pozzolan using an evolutionary structure," *Construction and Building Materials*, vol. 149, pp. 778–789, 2017.

[45] T. Chen and C. Guestrin, "Xgboost: a scalable tree boosting system," in *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, ACM, New York, NY, USA, August 2016.

[46] S. Lim and S. Chi, "Xgboost application on bridge management systems for proactive damage estimation," *Advanced Engineering Informatics*, vol. 41, Article ID 100922, 2019.

[47] R. Siddique, P. Aggarwal, and Y. Aggarwal, "Influence of water/powder ratio on strength properties of self-compacting concrete containing coal fly ash and bottom ash," *Construction and Building Materials*, vol. 29, pp. 73–81, 2012.

Hindawi
Advances in Civil Engineering
Volume 2021, Article ID 6682283, 15 pages
https://doi.org/10.1155/2021/6682283

Research Article

Machine Learning-Based Modeling with Optimization Algorithm for Predicting Mechanical Properties of Sustainable Concrete

Muhammad Izhar Shah,[1] **Shazim Ali Memon**,[2] **Muhammad Sohaib Khan Niazi**,[3] **Muhammad Nasir Amin**,[4] **Fahid Aslam**,[5] **and Muhammad Faisal Javed**[1]

[1]*Department of Civil Engineering, COMSATS University Islamabad, Abbottabad Campus, Abbottabad 22060, Pakistan*
[2]*Department of Civil and Environmental Engineering, School of Engineering and Digital Sciences, Nazarbayev University, Nur-Sultan 010000, Kazakhstan*
[3]*Civil Engineering Department, Qurtuba University of Science and Information Technology, Dera Ismail Khan, Pakistan*
[4]*Department of Civil and Environmental Engineering, College of Engineering, King Faisal University (KFU), P. O. 380, Al-Hofuf, Al Ahsa 31982, Saudi Arabia*
[5]*Department of Civil Engineering, College of Engineering, Prince Sattam Bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia*

Correspondence should be addressed to Muhammad Izhar Shah; mizhar@cuiatd.edu.pk and Shazim Ali Memon; shazim.memon@nu.edu.kz

Received 1 December 2020; Revised 4 February 2021; Accepted 18 February 2021; Published 4 March 2021

Academic Editor: Junfei Zhang

In this research, multiexpression programming (MEP) has been employed to model the compressive strength, splitting tensile strength, and flexural strength of waste sugarcane bagasse ash (SCBA) concrete. Particle swarm optimization (PSO) algorithm was used to fine-tune the hyperparameter of the proposed MEP. The formulation of SCBA concrete was correlated with five input parameters. To train and test the proposed model, a large number of data were collected from the published literature. Afterward, waste SCBA was collected, processed, and characterized for partial replacement of cement in concrete. Concrete specimens with varying proportion of SCBA were prepared in the laboratory, and results were used for model validation. The performance of the developed models was then evaluated by statistical criteria and error assessment tests. The result shows that the performance of MEP with PSO algorithm significantly enhanced its accuracy. The essential input variables affecting the output were revealed, and the parametric analysis confirms that the models are accurate and have captured the essential properties of SCBA. Finally, the cross validation ensured the generalized capacity and robustness of the models. Hence, the adopted approach, i.e., MEP-based modeling with PSO, could be an effective tool for accurate modeling of the concrete properties, thus directly contributing to the construction sector by consuming waste and protecting the environment.

1. Introduction

The construction industry consumes one-third of the world's energy and is a significant contributor of greenhouse gas emission to the environment [1]. Concrete is the most commonly used construction material. A single ton of concrete releases about 0.13 ton of carbon dioxide [2, 3]. In order to move towards sustainability, the concept of green concrete is getting popular to reduce the adverse effects of concrete. Green concrete is produced by replacing the conventional cementitious material with some waste as replacement of cement. Commonly used materials are fly ash, waste foundry sand, blast furnace slag, glass, metakaolin, rice husk ash, recycled aggregate, and bagasse ash [4]. Utilization of such materials is considered as low-carbon substitute to conventional construction materials. Sugarcane bagasse, which is an agricultural waste obtained after crushing and extraction, is used as fuel in the sugarcane industry [5]. Each ton of sugarcane generates approximately 26 percent of bagasse and 0.62 percent of residual ash [6].

The obtained ash is disposed by dumping in landfills and poses serious environmental issues [7]. Therefore, alternative and eco-friendly utilization methods of sugarcane bagasse ash (SCBA) are being discovered in the construction sector. Various research studies have concluded the viable use of SCBA in concrete as a cement replacement with a significant increase in mechanical properties of concrete. Chusilp et al. [8] reported higher compressive strength and lower permeability when concrete contained 20% SCBA by weight of cement. Sobuz [9] reported that maximum strength of SCBA concrete was obtained when cement was replaced with 10% of SCBA. Jagadesh et al. [10] reported that the strength of concrete made with 30% raw SCBA as a cement replacement reduced by almost 50%. The same authors reported about 28% increase in the strength of concrete when cement was replaced with 10% processed SCBA. The increase in strength was attributed to finer silica which reacted with calcium hydroxide to form additional CSH. Also, the finer SCBA particles filled voids and increased the packing density, which in turn increased the compressive strength of concrete. Bahurudeen et al. [11] reported higher strength of SCBA concrete as compared to normal concrete. Maximum compressive strength was achieved when cement was replaced with 10% SCBA. Strength reduction was linked with the dilution effect of the matrix caused by higher percentage replacement. Several researchers have also concluded that the utilization of SCBA also results in improved durability properties such as chloride penetration, chloride conductivity, water sorptivity, and water permeability [8, 12–14]. The above discussion highlights the point that SCBA concrete behaves differently at low and high replacement levels. This behavior could be attributed to several aspects, i.e., composition and dosage of SCBA, mix proportions, and the properties and type of the concrete constituents. Therefore, it is essential to correlate and figure out the factors influencing the mechanical properties of SCBA concrete, which, in turn, will have profound effect on the construction industry.

In order to address this issue, the unique features of artificial intelligence (AI) techniques such as random forest (RF), support vector machine (SVM), artificial neural network (ANN), gene expression programming (GEP), M5P, support vector regression (SVR), and convolution neural network (CNN) have been used to develop, correlate, and find the factors influencing the mechanical properties [15–19]. Zhang et al. [20] developed the RF model for the investigation of the hardened properties of synthetic-sand concrete. From results, it was found that RF showed reduced performance in comparison to other models. In a study conducted by Sun et al. [21], the authors utilized RF combined with an optimization algorithm for predicting the uniaxial compressive strength of rubberized concrete. The output of the study reported good accuracy of the model with a high correlation. Huang et al. [22] used the RF model along with beetle antenna search algorithm to predict the permeability of pervious concrete. The result of the study suggested improved performance of RF optimized model. ANN algorithm was used to model the compressive strength of lightweight concrete, foamed concrete, silica fume concrete, and high-performance concrete and elastic modulus of recycled aggregate concrete [17, 23–28]. A good correlation was observed in these studies for estimating the underlying concrete properties. However, the ANN is considered as black box algorithm, since it does not consider information or physical phenomena of the related problem [29]. Moreover, due to lack of parametric studies, the ANN models may not perform well on unseen datasets [30]. Recently, the advanced GEP technique was applied to predict the mechanical properties of SCBA and waste foundry sand concrete. Sensitivity and parametric analyses were performed to assess the performance of the models developed for mechanical properties [4, 31]. In these studies, the results of the comparative study revealed superior performance of GEP over regression methods. However, the GEP was identified with certain limitations since it fails to consider a few deviating datasets for model development, thus reducing its range of applicability [4]. Such deviating datasets should be removed from both training and testing phases to improve the model performance. Furthermore, the GEP encodes only a single chromosome and is suitable for simple relationship of input and output variables [32].

Considering the above limitations of certain AI techniques, an advanced algorithm, i.e., multiexpression programming with particle swarm optimization (PSO-MEP), has been adopted to model the mechanical properties of SCBA concrete. MEP with optimization technique (PSO) has been rarely used in civil engineering field despite its distinguished features. The mechanical properties of SCBA concrete in terms of compressive strength (CS), splitting tensile strength (ST), and flexural strength (FS) were modeled using PSO-MEP to solve complex relationship. A large dataset was collected from the literature for model training and testing. The validity of the developed models was verified utilizing the results obtained from laboratory testing. Furthermore, the variable importance, parametric study, and cross validation were used to assess the robustness and accuracy of the developed models.

2. Methods and Datasets

2.1. Multiexpression Programming. A linear variant of machine learning, i.e., multiexpression programming (MEP), has been proposed recently. The individual entities can be represented as a variable length in MEP [32, 33]. The linear variants permit MEP to extricate the genotype and phenotype [34]. MEP is considered expedient over other techniques due to linear chromosomes and encoding multiple solutions in an individual chromosome. This unique feature allows to search in a wider space to gather the best possible solution. In comparison to genetic programming, the MEP applies simple decoding procedures and is given particular importance in case of unknown complexity of targeted expression [29]. MEP can handle exceptions such as invalid expressions, divide by zero, etc. MEP can handle exceptions such as invalid expressions and divide by zero and it transforms into arbitrary terminal symbol so that the process continues. This produces a margin in structure of chromosome

during evaluation process [32]. The various steps involved in MEP process are shown in Figure 1. The MEP algorithm is similar to C language and pascal compiler. The outcome of the MEP process is a combination of mathematical operators or variables in a linear string of instruction form [35].

2.2. Particle Swarm Optimization. Particle swarm optimization (PSO) is a computational technique mainly used for problem optimization to enhance a solution by iterative process given a set of quality measures. This technique was developed by Kennedy and Eberhart based on motion of bird flocking and schooling fish [36]. In PSO, a process is initiated with a population of arbitrary solutions and the generations are updated to search for optimal solution. This technique is widely used for problem optimization such as forecasting the compressive strength, image contrast enhancement, and evaluating energy performance of building [37, 38]. In the present study, PSO was applied to fine-tune the hyperparameter of MEP and improve the modeling accuracy.

2.3. Modeling Database. A detailed dataset of different properties of SCBA concrete was collected from the published literature [8–11, 13, 14, 39–60]. The collected datasets contained information about SCBA concrete at 28 days. Some of the aforementioned research studies used concrete cubes to determine the compressive strength of SCBA concrete. In order to get homogenous data, the cube strength was converted to cylinder strength according to the method suggested by Elwell and Fu [61]. The collected literature data were statistically analyzed to get the most influential parameters affecting properties of SCBA concrete. The statistical parameters of different variables used in modeling are given in Table 1 [62]. An extensive study of design codes revealed that different models are available correlating the mechanical properties of normal concrete with compressive strength [63–67]. Therefore, the water to cementitious ratio (W/C), proportions of bagasse ash (SCBA%), fine aggregate content (FA), cement content (CC), and quantity of coarse aggregate (CA) are chosen as input for prediction of SCBA concrete mechanical properties. The formulation of compressive strength (CS), splitting tensile strength (ST), and flexural strength (FS) of SCBA concrete is considered to be a function of the following variables:

$$\mathrm{CS, ST, FS} = f\left(\frac{W}{C}, \mathrm{SCBA\%, CA, CC, FA}\right). \quad (1)$$

2.4. Modeling Parameters for MEP. For developing a generalized relationship, several fitting parameters are required for MEP. The chosen parameters for MEP modeling are presented in Table 2. Basic mathematical operators were considered to get simple expressions. The trial and error method was adopted to get the fitting parameters [68]. The population size was used to specify the number of programs required in the population. High population can lead to complex and long convergence time and often causes overfitting problem beyond specified limit. Moreover, an algorithm run with a large number of generations could lead to a model with minimum error. Several combinations of

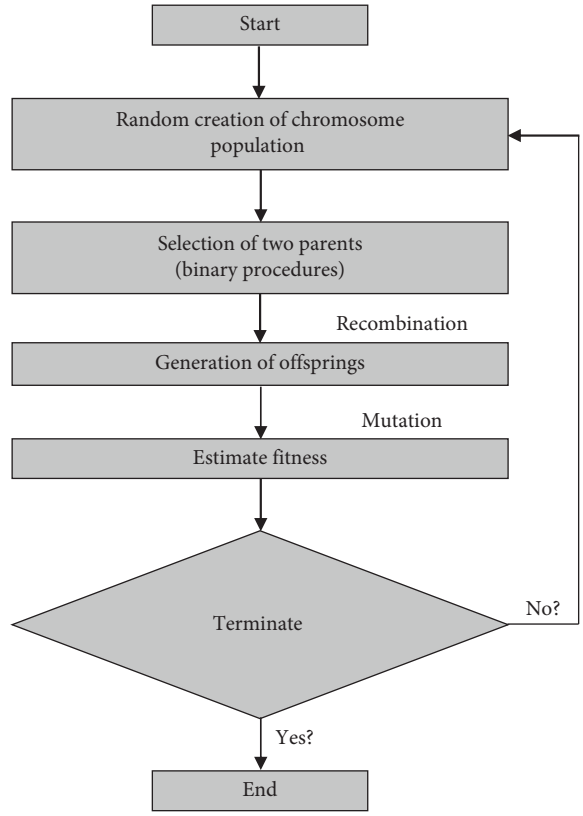


Figure 1: Schematic diagram of MEP technique.

Table 1: Statistics of the input parameters.

Parameter	W/C	CC	SCBA%	FA	CA
Unit	—	kg/m^3	%	kg/m^3	kg/m^3
Range	0.3	444	50	614	772
Min	0.3	112	0	239	477
Max	0.6	555	50	853	1249
Mean	0.47	336.5	13.98	603.5	884.6
SD	0.074	98.5	10.46	232.1	392.3

Table 2: Optimum parameter setting for MEP.

Parameters	Setting
Number of subpopulation	50
Size of subpopulation	250
Code length	40
Crossover probability	0.9
Mathematical operators	+, −, ×, ÷
Mutation probability	0.01
Tournament size	4
Operators	0.5
Variables	0.5
Number of generations	1000

parameters were initiated and the best possible grouping was selected based on the model performance.

2.5. Performance Evaluation. The performance of the developed PSO-MEP model was assessed by measuring various statistical indicators including correlation coefficient (R), root mean squared error (RMSE), Nash–Sutcliffe efficiency (NSE), mean absolute error (MAE), relative root mean squared error (RRMSE), relative squared error (RSE), and performance index (ρ). Moreover, another measure to reduce the model overfitting is to select the best model by minimizing the objective function (OF) as suggested by Gandomi et al. and Azim [34, 69]. The same approach has been applied in this study, and OF is termed as fitness function. The mathematical expressions for the statistical indicators are shown below from equations (2) to (9). The high values of R and NSE and low values of RMSE and MAE indicate better performance. The indicator R quantifies the linear relationship between input and output [27], and R value more than 0.8 signifies excellent correlation among predicted and actual data [70]. However, it alone cannot be considered to judge the efficiency of a model. Despotovic et al. [71] highlighted that a model can be considered excellent if the value of RRMSE is between 0 and 0.10 and good if the values are between 0.11 and 0.20. The minimum and maximum value of NSE is negative infinity and 1, respectively, with 1 showing the best output. The values of ρ and OF range from 0 to positive infinity with a value near to zero representing a good model. It can be noted that the OF taken into account the effect of RRMSE, R, and relative percentage of data in different sets. Hence, low OF value shows superior performance of a model.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(P_i - M_i\right)^2}{N}}, \tag{2}$$

$$\text{NSE} = 1 - \frac{\sum_{i=1}^{n}\left(M_i - P_i\right)^2}{\sum_{i=1}^{n}\left(M_i - \overline{M}_i\right)^2}, \tag{3}$$

$$R = \frac{\sum_{i=1}^{n}\left(M_i - \overline{M}_i\right)\left(P_i - \overline{P}_i\right)}{\sqrt{\sum_{i=1}^{n}\left(M_i - \overline{M}_i\right)^2 \sum_{i=1}^{n}\left(P_i - \overline{P}_i\right)^2}}, \tag{4}$$

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

$$\text{RSE} = \frac{\sum_{i=1}^{n}\left(P_i - M_i\right)^2}{\sum_{i=1}^{n}\left(\overline{M}_i - \overline{M}_i\right)^2}, \tag{6}$$

$$\text{RRMSE} = \frac{1}{\left|\overline{M}\right|}\sqrt{\frac{\sum_{i=1}^{n}\left(P_i - M_i\right)^2}{N}}, \tag{7}$$

$$\rho = \frac{\text{RRMSE}}{1 + R}, \tag{8}$$

$$\text{OF} = \left(\frac{n_T - n_{\text{TE}}}{n}\right)\rho_T + 2\left(\frac{n_{\text{TE}}}{n}\right)\rho_{\text{TE}}, \tag{9}$$

where n, P_i, M_i, $\overline{P}_i$, and $\overline{M}_i$ show the number of data points, predicted data, measured data, mean of predicted data, and mean of measured data, respectively. T and TE represent the training and testing datasets, respectively.

2.6. Hyperparameter Tuning and K-Fold Cross Validation. Fine-tuning of hyperparameters is a major concern in machine learning-based modeling. Various researchers used different optimization techniques for hyperparameter tuning such as beetle antennae search [22, 72, 73] and grid search method [74]. In our study, particle swarm optimization (PSO) was employed for hyperparameter tuning to improve the accuracy of the model. Firstly, the dataset was divided into 70% and 30% for model training and testing, respectively. Then k-fold cross validation was applied to assess the hyperparameter tuning process by PSO. The k-fold method divides the actual data to k subclasses. Moreover, the efficiency of the 10-fold cross validation method is reported in the literature [75, 76]. Among all the ten subsets, each subset was used for validation and the same method was repeated for all the remaining subsets. Consequently, the optimized MEP model and the associated optimized hyperparameters were obtained after 10 rounds. After getting the optimum structure of MEP by PSO, the result of training and testing dataset was evaluated employing statistical indicators. Finally, 10-fold cross validation was used again to ensure the generalized capability of PSO-MEP and output was expressed in terms of mean accuracy. Figure 2 shows the hyperparameter tuning process for MEP with the help of PSO for both training and testing.

3. Experimental Investigation

3.1. SCBA Processing. The sugarcane bagasse ash (SCBA), resulting from burning of bagasse (as a fuel), was collected from a sugar industry located in Malakand, Pakistan. It has been reported in the literature that the presence of fibrous and unburnt content in raw bagasse ash decreases the pozzolanic activity and also elevates the loss on ignition. Removal of such particles, proper characterization, and grinding up to cement fineness significantly increase its pozzolanic activity [11, 77]. In our study, the collected SCBA was sieved from #200 sieve to remove undesirable particles and subsequently grinded to reduce the particle size. Grinding was carried out in ball mill machine with ceramic balls as grinding media. The grinding media to SCBA ratio of 5 by weight was kept constant, and the machine was operated at 100 rpm [78]. Different SCBA samples were obtained at a grinding duration of 15, 30, 45, and 60 minutes. Blaine fineness values were determined according to ASTM C204 in order to obtain the effect of grinding on particle size.

The chemical composition of grounded SCBA was evaluated through X-ray fluorescence (XRF). The composition is given in Table 3. It can be inferred from XRF results

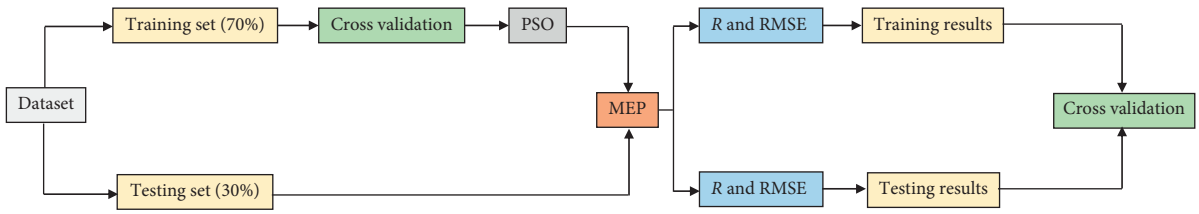


Figure 2: Hyperparameter tuning by PSO.

Table 3: XRF results of SCBA.

Composition	Percentage
SiO_2	66.28
Al_2O_3	8.36
Fe_2O_3	1.39
CaO	9.06
MgO	5.56
P_2O_5	2.46
K_2O	3.52
Na_2O	1.30
TiO_2	0.19
MnO	0.02
LOI	1.67
Moisture content	1.15

that the quantity of silica, alumina, and iron oxide is above 70%, meeting the chemical requirement of a pozzolan according to ASTM C618-05 standard.

Scanning electron micrographs of SCBA are shown in Figure 3. Images were taken at different magnification ranging from X500 to X5000. In the micrographs, the heterogeneous nature of SCBA is prominent. The SCBA exhibited various shapes such as elongated, needle, flat, oval, and irregular. According to available literature [79], the irregular shape particles are mainly rich in silica. The size of needle and oval shape particles is about 50 μm. Furthermore, voids over the surface ranging from 5 μm to 10 μm are visible in the form of black spots. Overall, the observed size of different particles ranges from 5 μm to 50 μm.

3.2. Mix Proportions and Properties of SCBA Concrete. Experimental testing of SCBA concrete was conducted to check the performance of the model required for validation purpose. Concrete specimens were casted at room temperature of 25°C, and comparison of fresh and hardened properties was made between concrete made with bagasse ash (BC) and control specimens (CM). Different dosages of SCBA, i.e., 0–40%, were incorporated in concrete as cement replacement. The desired target strength was formulated based on published data as adamant variation was observed. The detailed mix design with the formulation is depicted in Table 4.

For the validation of the PSO-MEP model, concrete specimens, i.e., cylinders and $4" \times 4" \times 20"$ beams, were prepared with different proportions of SCBA and tested for compressive strength (CS), splitting tensile strength (ST), and flexural strength (FS), respectively, at the curing age of 28 days. ASTM standards C39, C496, and C293 were, respectively, followed for compressive, splitting tensile, and flexural strength of the CM and BC.

4. Results and Discussion

4.1. Mechanical Properties of SCBA Concrete. The results of mechanical properties, i.e., compressive strength (CS), splitting tensile strength (ST), and flexural strength (FS), of SCBA concrete were determined in laboratory by casting concrete cylinders and beams with varying proportions of SCBA (0% to 40%), as presented in Table 5. It can be observed that strength increased up to 10% SCBA and then consistently decreased at higher SCBA level. The maximum strength gained is at 10% replacement. The strength gain at 10% SCBA may be related to the pozzolanic reaction resulting in additional calcium silicate hydrate (CSH). For higher replacement level, the decrease in strength was found to be 6.5%, 17.3%, and 30.3% for 20BC, 30BC, and 40BC, respectively. This decrease in strength may be attributed to unavailability of the adequate amount of calcium hydroxide.

Similar results were observed for splitting tensile strength and flexural strength as illustrated in Table 5. For 10% and 20% replacement of SCBA, the increase in tensile strength as compared to control concrete was 25.3% and 15.8%, respectively. This shows that the maximum tensile strength was attained at 10% addition of SCBA as cement replacement. For 30% and 40% SCBA replacement level, the tensile strength decreased by 7.9% and 23.8%, respectively, as compared to CM. For the case of flexural strength, the maximum strength was also achieved at 10% SCBA. The flexural strength at higher replacement level (20%, 30%, and 40% SCBA) reduced by 15.5%, 28.8%, and 42.5%, respectively. According to available literature, the increased tensile and flexural strength at 10% SCBA may be due to the microfibrous nature of SCBA, which is related with CSH formation, and also due to the formation of aluminates, resulting in needle-like structure [80, 81]. The interlocking of these needles takes place between hydrated paste and may directly enhance the tensile and flexural strength. In short, the enhanced properties may be due to formation of more hydrated products, less porous structure of concrete made with SCBA, and the enhanced interfacial transition zone (ITZ) [80, 82, 83].

4.2. Formulation of Mechanical Properties with PSO-MEP. The results obtained from PSO-MEP for compressive strength (CS), splitting tensile strength (ST), and flexural strength (FS) are interpreted to get the empirical mathematical expressions for the prediction of aforementioned

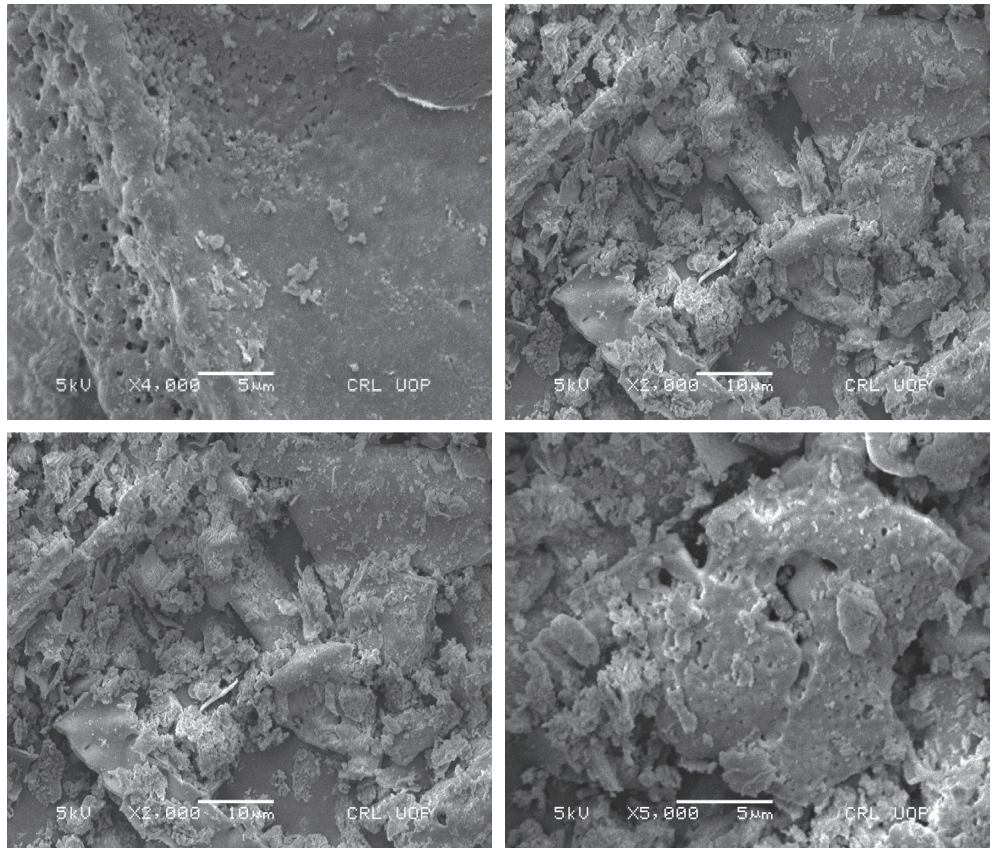


Figure 3: SEM images of SCBA at different magnifications.

Table 4: Mix Proportions of the concrete mix.

Mix design	Cement (kg/m^3)	Coarse aggregate (kg/m^3)	SCBA (kg/m^3)	W/C	Fine aggregate (kg/m^3)	Water (kg/m^3)
0BC (CM)	366	1013.5	0	0.5	742.3	183
10BC	329.4	1013.5	36.6	0.5	742.3	183
20BC	292.8	1013.5	73.2	0.5	742.3	183
30BC	256.2	1013.5	109.8	0.5	742.3	183
40BC	219.6	1013.5	146.4	0.5	742.3	183

properties based on five input variables. The derived equations are shown as equations (10), (11), and (12) for CS, ST, and FS, respectively. A comparative graph of experimental and predicted CS is presented in Figure 4 for training, testing, and validation. The expressions for regression lines are also shown. For an ideal situation, it is known that the slope of the line should be nearly equal to 1. It can be deduced from Figure 4 that the developed PSO-MEP model considered the effect of input variables and retains a strong correlation between experimental and predicted data as evident from the slope, i.e., 0.8951, 0.9315, and 0.9014 for training, testing, and validation, respectively.

$$\mathrm{CS\ (MPa)} = \left(1.1x_1 + 1.1x_2\right) + \left(\frac{8x_0^2 \times x_4}{x_3 - x_4}\right)\left(16x_0^3\left(1.1x_1 + 1.1x_2\right) + \frac{4\left(5x_1 - x_3\right)}{1.1x_1 + 1.1x_2}\right)^2, \tag{10}$$

$$\mathrm{ST\,(MPa)} = \left(x_0 + \frac{x_0^2}{x_0 - 0.375}\right) - \left(\frac{x_0 - 0.375}{\left(x_0 - x_1\right) + \left(x_0/x_0 - 0.375\right)}\right) + \frac{\left(x_0 - 0.375\right)^2}{\left(x_0^2 - 0.375\right)^2} - \left(\frac{x_0 \times x_1^2}{x_2}\right) + \frac{x_0^2 \times x_3}{x_4 - 0.375}, \tag{11}$$

$$\mathrm{FS\,(MPa)} = \left(\frac{2x_3x_0}{x_4 + \left(3x_2 + 0.97\right)^2\left(2x_1 - 89x_0\right)}\right) + \left(\frac{2x_0}{\left(\left(x_1 - 89x_0\right)/\left(100x_0 - 48.5\right)\right)}\right), \tag{12}$$

Table 5: Results of laboratory-derived mechanical properties of SCBA concrete.

Mix	Compressive strength (MPa)				
	0BC	10BC	20BC	30BC	40BC
Sample 1	23.5	23.9	21.5	18.5	16.7
Sample 2	22.7	23.6	21.6	19.6	15.6
Sample 3	22.9	23.7	21.2	19.1	16.4
Sample 4	23.4	24.2	22.3	19.5	15.7
Average	23.1	23.8	21.6	19.1	16.1
Splitting tensile strength (MPa)					
Sample 1	6.3	7.9	7.2	6.7	5.3
Sample 2	6.2	7.8	7.3	5.6	4.7
Sample 3	6.2	8.1	7.5	5.3	4.4
Sample 4	6.7	8.1	7.5	5.8	4.9
Average	6.3	7.9	7.3	5.8	4.8
Flexural strength (MPa)					
Sample 1	4.7	5.1	3.9	3.1	2.8
Sample 2	4.3	5.1	3.8	3.3	2.6
Sample 3	4.6	5.2	3.8	3.3	2.6
Sample 4	4.6	5.3	3.7	3.2	2.5
Average	4.5	5.1	3.8	3.2	2.6

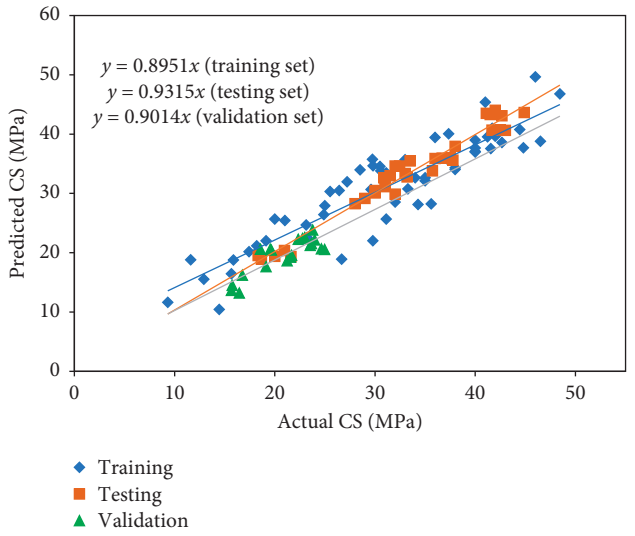


Figure 4: Comparison of actual and predicted CS.

where

$$x_0 = \frac{w}{c},$$
$$x_1 = \text{SCBA\%}$$
$$x_2 = \text{CC}, \quad (13)$$
$$x_3 = \text{FA},$$
$$x_4 = \text{CA}.$$

A similar comparison has been drawn for the tensile strength (ST) output as shown in Figure 5. It can be seen that the model exhibited an excellent correlation among experimental and predicted data to estimate ST. The respective slopes of the regression line are close to ideal fit, i.e., 0.9351, 0.8903, and 0.9273 for training, testing, and validation. Similar to CS, the model for ST performs exceptionally well on training dataset revealing that the issue of overfitting by the model has been reduced to a larger extent. Moreover, the performance and accuracy of a model depend on the number of data points incorporated in the model [84]. In this study, 110 data points have been selected to model ST; therefore, an accurate model with minimum error has been achieved.

The PSO-MEP model results for flexural strength (FS) are graphically presented in Figure 6. During model training, testing, and validation, the slope of the regression line was observed to be 0.9494, 0.9026, and 0.9332. Compared with the models for CS and ST, an excellent estimate is observed for experimental and predicted data point as illustrated in Figure 6. The model performs exceptionally well for training and testing data.

4.3. Model Evaluation by Statistical Measures. The reliability of a model depends on the amount of data used for model development. The literature survey recommended that the ratio of number of data points to the number of input variables for both training and testing should be greater than 5 [85]. For model training, the aforementioned ratio is 11.8, 13.8, and 8.2 for CS, ST, and FS, respectively. For model testing, the values are 6.2, 6.6, and 5.7 for CS, ST, and FS, respectively. The performance of the developed models was evaluated by various statistical criteria such as RMSE, NSE, RSE, RRMSE, OF, ρ, MAE, and R. The values of these indicators for CS, ST, and FS are given in Table 6 for training, testing, and validation. It can be observed from Table 6 that models exhibit a strong correlation as evident by the R value which is 0.91, 0.90, and 0.91 for training and 0.94, 0.92, and 0.91 for testing of CS, ST, and FS, respectively. The maximum and minimum values of NSE are 0.89 and 0.87 for CS, 0.91 and 0.85 for ST, and 0.86 and 0.87 for FS models, respectively. The values of RMSE and MAE are considerably

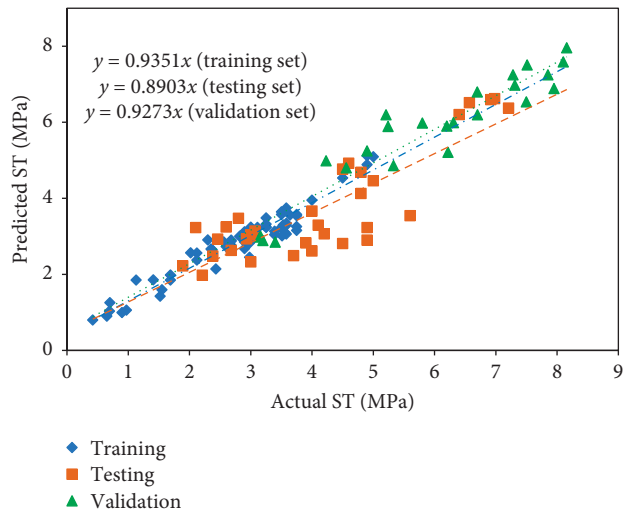


Figure 5: Comparison of actual and predicted ST.

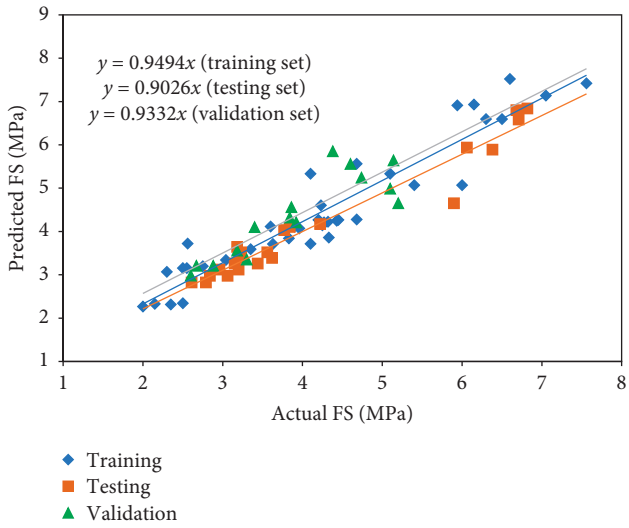


Figure 6: Comparison of actual and predicted FS.

low for the three datasets which highlighted the generalized capability and high accuracy of the models. Based on RMSE, the model for ST can be categorized as excellent, as the values for all the three datasets are 2.43, 2.65, and 3.25, respectively. It can be deduced from results that for all models, MAE lies in the good range from 1.45 to 3.98. Furthermore, the OF values for CS, ST and FS are 0.036, 0.031, and 0.052, respectively. All these values are close to zero, indicating the accurate performance and further validating that the overfitting problem has been suitably addressed by the models. For all the developed models, the RRMSE ranged from 0.04 to 0.16, thus highlighting the accuracy of the model in predicting the mechanical properties of SCBA concrete. In order to infer the absolute error, the data points are plotted in Figure 7 showing the error among actual and model simulated data. The mean absolute error for CS, ST, and FS is 2.87, 0.405, and 0.675, respectively. The minimum and maximum absolute errors are 0.1 and 7.76 for CS, 0.08 and 2.15 for ST, and 0.075 and 1.95 for FS, respectively. It is worth mentioning that almost 80% of results for CS, ST, and FS have error less than 3, 0.5, and 0.6, respectively.

The criteria for external validation of the models are given in Table 7. It has been suggested that regression line slope, i.e., k and k', passing through the origin should be nearly equal to 1 [86]. Roy and Roy [87] reported that criteria for external predictability of a model are satisfied when the indicator R_m is greater than 0.5. It can be observed from Table 7 that all the three models (CS, ST, and FS) satisfy the conditions for external predictability.

4.4. Model Cross Validation Results. The 10-fold cross validation was applied to evaluate the CS, ST, and FS models, and the results are graphically shown in Figures 8 and 9 for R and RMSE, respectively. A variation in the results can be observed at individual level as depicted in the figure. However, it demonstrated a good mean accuracy. The average R value obtained for CS, ST, and FS is 0.85, 0.89, and 0.85, respectively. In all the 10-fold cross validations, the maximum and minimum R values of 0.72 and 0.91, respectively, were achieved by ST. Similarly, the mean RMSE values of 4.54, 3.89, and 4.78 were accomplished by CS, ST, and FS, respectively. For individual subset, the lowest RMSE, i.e., 1.86, was also attained by ST. Overall, the results from 10-fold cross validation shows the accurate performance, generalized capability, and robustness of the PSO-MEP models.

4.5. Variable Importance and Parametric Analysis. Variable importance is a process to find out the most influencing input variables affecting the targeted output. The most sensitive input parameters and their relative contribution to output were determined, and the results are illustrated in Figure 10. The result shows that cement content is the most important variable contributing 55% to mechanical properties of SCBA concrete. Similarly, water-cement ratio and amount of coarse aggregate turned out to be important variables with 17.15% and 16.97% contribution to the output. The result further reveals that quantity of fine aggregate is the least important parameter affecting the mechanical properties of SCBA concrete.

Parametric analysis was performed to assess the variation of the model output with every single input variable. This process is recommended in research studies to determine the effect of all the physical phenomena and inputs. In this method, all the variables were kept constant at their mean values and the variation of the model output is plotted with a single input variable. The same procedure was applied for the individual input parameter. Figure 11 shows the parametric analysis results for the developed CS model only as similar patterns were obtained for other mechanical properties (ST and FS). Therefore, parametric study results are discussed in detail for CS in the following.

It is a known fact that an increase in water-cement ratio decreases the strength of concrete [84, 88]. It can be seen form Figure 11(a) that an increase in W/C resulted in a consistent decrease in compressive strength. The effect of SCBA% on the mechanical properties of SCBA concrete depends on physical and chemical composition and

Table 6: Statistics for the training, testing, and validation dataset of the models.

Models	Dataset	NSE	R	RMSE	MAE	RSE	RRMSE	ρ	OF
CS	Training	0.87	0.91	3.47	2.96	0.16	0.04	0.020	0.036
	Testing	0.89	0.94	2.98	2.98	0.12	0.09	0.046	
	Validation	0.89	0.93	2.87	1.67	0.15	0.04	0.020	
ST	Training	0.85	0.90	2.43	3.67	0.23	0.09	0.047	0.031
	Testing	0.91	0.92	2.65	3.69	0.26	0.12	0.062	
	Validation	0.90	0.92	3.25	3.98	0.31	0.10	0.052	
FS	Training	0.86	0.91	3.92	1.87	0.29	0.13	0.068	0.052
	Testing	0.87	0.91	3.34	1.45	0.28	0.15	0.078	
	Validation	0.86	0.93	3.67	2.87	0.19	0.16	0.079	

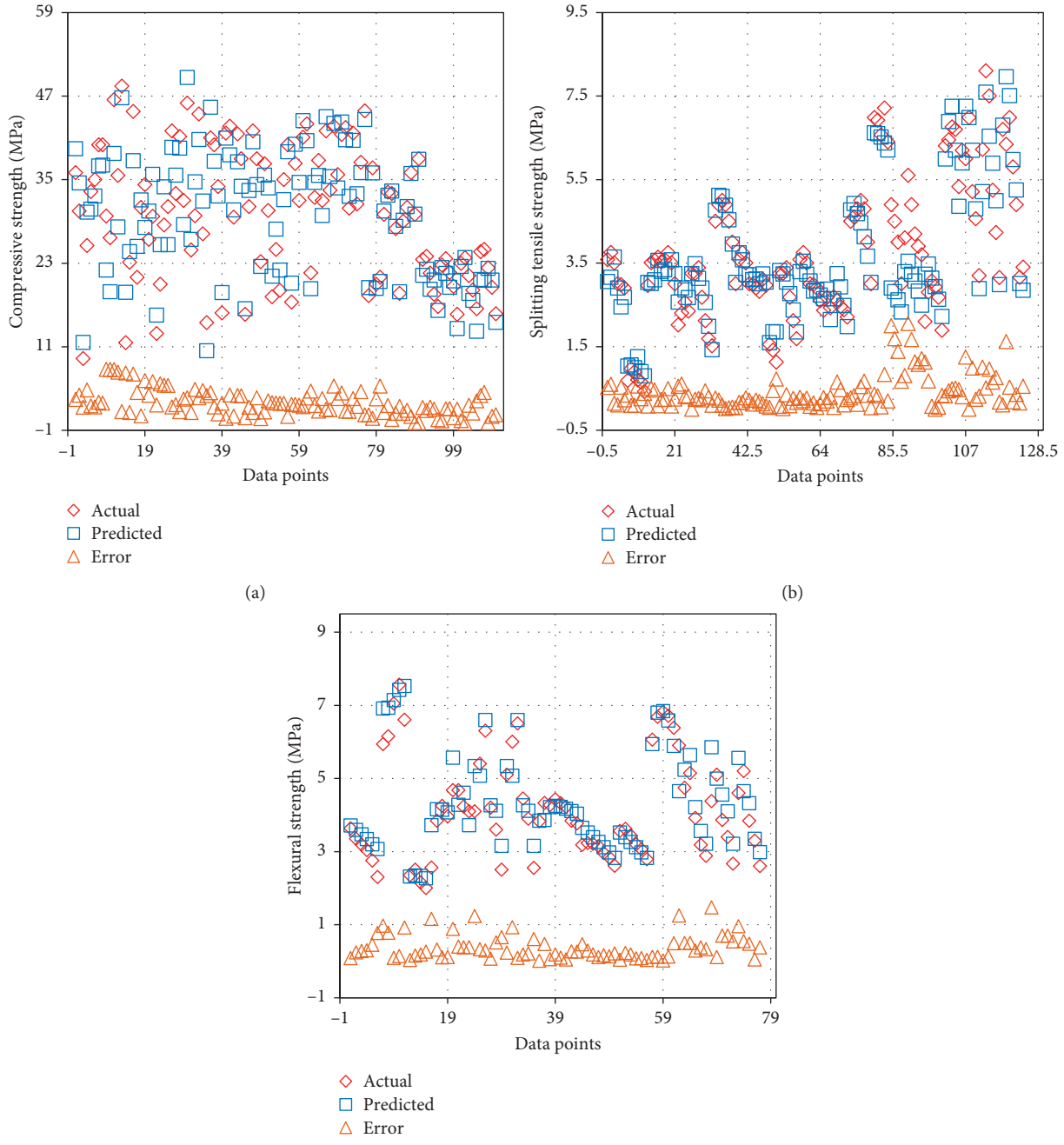


Figure 7: Description of the error among the actual and predicted data: (a) CS; (b) ST; (c) FS.

Table 7: Statistical indicators of verified models for external predictability.

S.No.	Equation	Condition	CS	ST	FS	Suggested by
1.	$R = \frac{\sum_{i=1}^{n}(M_i-\overline{M_i})(P_i-\overline{P_i})}{\sqrt{\sum_{i=1}^{n}(M_i-\overline{M_i})^2\sum_{i=1}^{n}(P_i-\overline{P_i})^2}}$	$R > 0.8$	0.92	0.92	0.91	Frank and Todeschini [85]
2.	$k = \frac{\sum_{i=1}^{n}(M_i-P_i)}{M_i^2}$	$0.85 < k < 1.15$	1.00	0.99	1.01	Golbraikh and Tropsha [86]
3.	$k' = \frac{\sum_{i=1}^{n}(M_i-P_i)}{P_i^2}$	$0.85 < k'' < 1.15$	0.98	0.98	1.05	Golbraikh and Tropsha [86]
4.	$R_m = R^2 \times (1 - \sqrt{\lvert R^2 - R_0^2\rvert}$	$R_m > 0.5$	0.67	0.71	0.64	Roy and Roy [87]
	$R_0^2 = \frac{\sum_{i=1}^{n}(P_i-M_i^0)^2}{\sum_{i=1}^{n}(P_i-\overline{P_i^0})^2}, \quad M_i^0 = k \times P_i$	$R_0^2 \cong 1$	0.98	0.98	0.97	
	$R_0^{2'} = \frac{\sum_{i=1}^{n}(M_i-P_i^0)^2}{\sum_{i=1}^{n}(M_i-\overline{M_i^0})^2}, \quad P_i^0 = k' \times P_i$	$R_0^{2'} \cong 1$	0.98	0.99	0.98	

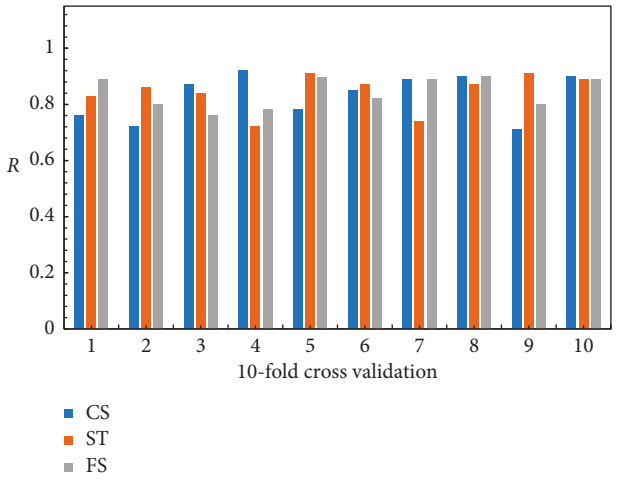


Figure 8: Cross validation results of CS, ST, and FS models based on R.

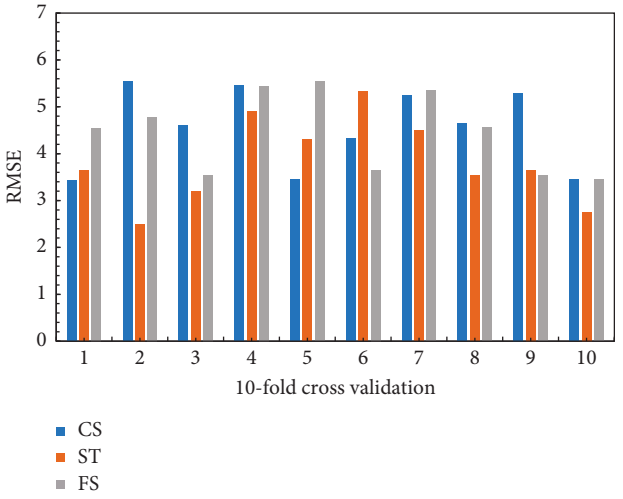


Figure 9: Cross validation results of CS, ST, and FS models based on RMSE.

replacement percentage of SCBA [11]. It can observed in Figure 11(b) that the compressive strength of SCBA concrete increased up to a certain level and then decreased with

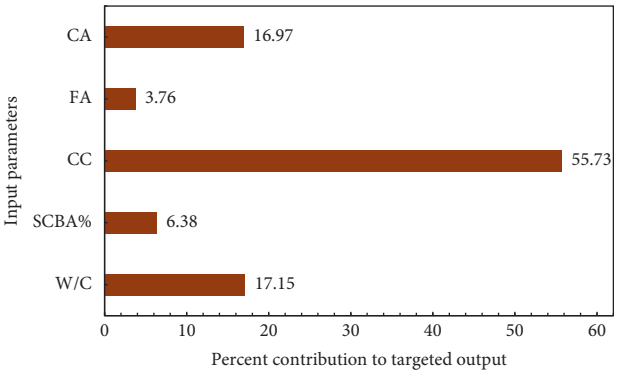


Figure 10: Importance of input variables on the targeted output.

increase in the level SCBA%. It can be deduced from the figure that maximum strength has been attained at 10% SCBA replacement. A similar trend was also observed for experimental testing of SCBA concrete as described in Section 4.1. Hence, the results of parametric analysis are in close agreement with laboratory testing where maximum strength was achieved at 10% SCBA. Similar findings were reported in [9–11] which showed that 10% SCBA replacement attained higher strength.

Cement is the principal cementitious material in concrete, and the increase in the cement content enhances the mechanical properties of concrete. A similar trend can also be seen in Figure 11(c) where the compressive strength linearly increased with increase in cement content. The higher cement content produces more calcium silicate hydrate (CSH) which forms a dense structure, thereby increasing the compressive strength. Figures 11(d) and 11(e) illustrate the variation of compressive strength with fine and coarse aggregate content, respectively. Both figures demonstrated that CS decreased with the increase in FA and CA content. The aggregates are inert materials and are used to provide volume stability to concrete. The quantity of aggregates affects the mechanical properties of concrete. However, keeping all the quantities constant in a mix, the strength of concrete decreases with increase in the quantity of aggregate. Similar results were also observed in the current study.

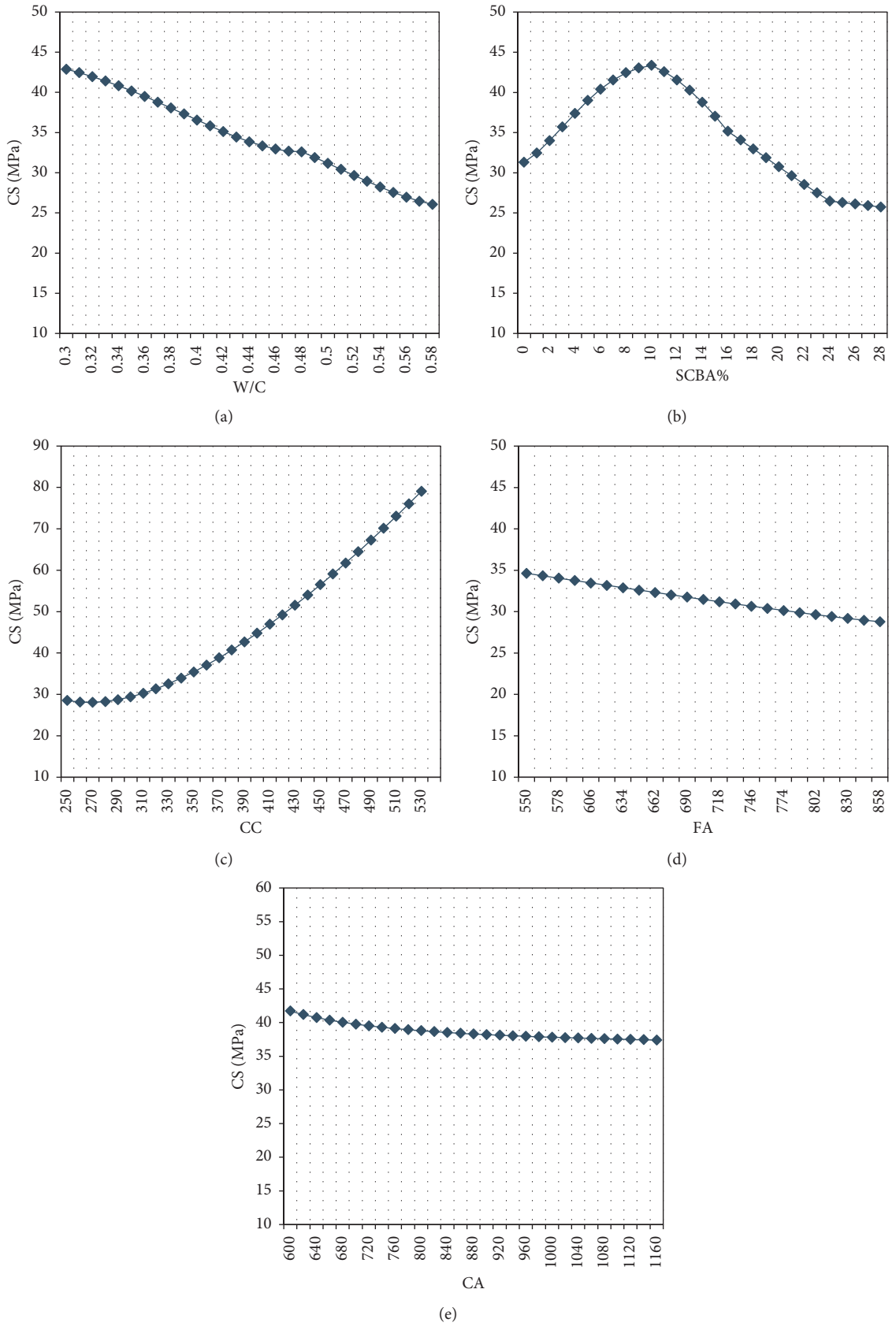


Figure 11: Parametric analysis results for (a) W/C, (b) SCBA%, (c) CC, (d) FA, and (e) CA.

From the aforementioned discussion, it can be concluded that the developed PSO-MEP model has successfully incorporated the effect of all input variables, specifically the complex physical behavior of SCBA, thus making it more suitable for estimation of complex problems.

5. Conclusion

This study adopted the two-fold objective. Firstly, the compressive strength, splitting tensile strength, and flexural strength of SCBA concrete were modeled by utilizing multiexpression programming (MEP). The particle swarm optimization (PSO) was employed to tune the hyperparameter of the model. The proposed model was developed and formulated based on extensive literature data. Secondly, sugarcane bagasse ash (SCBA) was characterized, optimized, and used in different proportions (10%, 20%, 30%, and 40%) as a partial replacement with cement. SCBA concrete specimens were prepared and tested for mechanical properties, and the results were used for model validation. The performance and accuracy of the final models were evaluated with the help of statistical indicators, i.e., RMSE, NSE, RSE, RRMSE, OF, ρ, MAE, and R. The results obtained from the developed models exhibited an excellent correlation with the experimental data with R value above 0.9, MAE and RMSE below 5, and OF values close to zero for CS, ST, and FS models. The developed PSO-MEP models also satisfied the criteria for external validation available in the literature. The variable importance and parametric analysis revealed that the developed model has taken into account the effect of all the inputs. The final outcome of the model was also cross verified with the 10-fold validation; the results ensured that the models produced generalized outcome, and the overfitting issue has also been addressed. It is obvious from the current study that utilization of different wastes is indispensable for sustainability viewpoint and machine learning models play a crucial role in its success.

Data Availability

The data used in this study were collected from the published literature and are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] H. Du and S. D. Pang, "Value-added utilization of marine clay as cement replacement for sustainable concrete production," *Journal of Cleaner Production*, vol. 198, pp. 867–873, 2018.

[2] Z. He, X. Zhu, J. Wang, M. Mu, and Y. Wang, "Comparison of CO2 emissions from OPC and recycled cement production," *Construction and Building Materials*, vol. 211, pp. 965–973, 2019.

[3] L.-x. Mao, Z. Hu, J. Xia et al., "Multi-phase modelling of electrochemical rehabilitation for ASR and chloride affected concrete composites," *Composite Structures*, vol. 207, pp. 176–189, 2019.

[4] M. F. Iqbal, Q.-f. Liu, I. Azim et al., "Prediction of mechanical properties of green concrete incorporating waste foundry sand based on gene expression programming," *Journal of Hazardous Materials*, vol. 384, Article ID 121322, 2020.

[5] W. A. Pippo and C. A. Luengo, "Sugarcane energy use: accounting of feedstock energy considering current agro-industrial trends and their feasibility," *International Journal of Energy and Environmental Engineering*, vol. 4, no. 1, p. 10, 2013.

[6] G. C. Cordeiro, D. C. T. F. Romildo, M. F. Eduardo et al., "Influence of mechanical grinding on the pozzolanic activity of residual sugarcane bagasse ash," in *Proceedings of International RILEM Conference on the Use of Recycled Materials in Building and Structures*, Barcelona, Spain, January 2004.

[7] K. Pedersen, A. Jensen, M. Skjothrasmussen, and K. Damjohansen, "A review of the interference of carbon containing fly ash with air entrainment in concrete," *Progress in Energy and Combustion Science*, vol. 34, no. 2, pp. 135–154, 2008.

[8] N. Chusilp, C. Jaturapitakkul, and K. Kiattikomol, "Utilization of bagasse ash as a pozzolanic material in concrete," *Construction and Building Materials*, vol. 23, no. 11, pp. 3352–3358, 2009.

[9] M. H. R. Sobuz, "Properties of concrete by using bagasse ash and recycle aggregate," *Concrete Research Letters*, vol. 5, 2014.

[10] P. Jagadesh, A. Ramachandramurthy, and R. Murugesan, "Evaluation of mechanical properties of sugar cane bagasse ash concrete," *Construction and Building Materials*, vol. 176, pp. 608–617, 2018.

[11] A. Bahurudeen, D. Kanraj, V. Gokul Dev, and M. Santhanam, "Performance evaluation of sugarcane bagasse ash blended cement in concrete," *Cement and Concrete Composites*, vol. 59, pp. 77–88, 2015.

[12] A. Bahurudeen and M. Santhanam, "Performance evaluation of sugarcane bagasse ash-based cement for durable concrete," in *Proceedings of the International Conference on the Durability of Concrete Structures*, Dunbeath, Caithness, UK, July 2014.

[13] A. Bahurudeen, S. Manu, W. Kaiser et al., "Assesment of pozzolanic performance of sugarcane bagasse ash," *Journal of Materials in Civil Engineering*, vol. 28, no. 2, Article ID 04015095, 2016.

[14] A. Rerkpiboon, W. Tangchirapat, and C. Jaturapitakkul, "Strength, chloride resistance, and expansion of concretes containing ground bagasse ash," *Construction and Building Materials*, vol. 101, pp. 983–989, 2015.

[15] M. I. Khan, "Predicting properties of high performance concrete containing composite cementitious materials using artificial neural networks," *Automation in Construction*, vol. 22, pp. 516–524, 2012.

[16] A. T. A. Dantas, M. Batista Leite, and K. de Jesus Nagahama, "Prediction of compressive strength of concrete containing construction and demolition waste using artificial neural networks," *Construction and Building Materials*, vol. 38, pp. 717–722, 2013.

[17] E. M. Golafshani, A. Behnood, and M. Arashpour, "Predicting the compressive strength of normal and high-performance concretes using ANN and ANFIS hybridized with grey wolf optimizer," *Construction and Building Materials*, vol. 232, Article ID 117266, 2020.

[18] R. Parichatprecha and P. Nimityongskul, "Analysis of durability of high performance concrete using artificial neural networks," *Construction and Building Materials*, vol. 23, no. 2, pp. 910–917, 2009.

[19] J. Sun, J. Zhang, Y. Gu, Y. Huang, Y. Sun, and G. Ma, "Prediction of permeability and unconfined compressive strength of pervious concrete using evolved support vector regression," *Construction and Building Materials*, vol. 207, pp. 440–449, 2019.

[20] J. Zhang, D. Li, and Y. Wang, "Toward intelligent construction: prediction of mechanical properties of manufactured-sand concrete using tree-based models," *Journal of Cleaner Production*, vol. 258, Article ID 120665, 2020.

[21] Y. Sun, G. Li, J. Zhang et al., "Prediction of the strength of rubberized concrete by an evolved random forest model," *Advances in Civil Engineering*, Article ID 5198583, 7 pages, 2019.

[22] J. Huang, T. Duan, Z. Yi et al., "Predicting the permeability of pervious concrete based on the beetle antennae search algorithm and random forest model," *Advances in Civil Engineering*, vol. 2020, Article ID 8863181, 11 pages, 2020.

[23] A. Ashteyat, Y. T. Obaidat, Y. Z. Murad, and R. Haddad, "Compressive strength prediction of lightweight short columns at elevated temperature using gene expression programing and artificial neural network," *Journal of Civil Engineering and Management*, vol. 26, no. 2, pp. 189–199, 2020.

[24] 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.

[25] A. Sadrmomtazi, J. Sobhani, and M. A. Mirgozar, "Modeling compressive strength of EPS lightweight concrete using regression, neural network and ANFIS," *Construction and Building Materials*, vol. 42, pp. 205–216, 2013.

[26] A. Öztaş, M. Pala, M. Bhatti et al., "Predicting the compressive strength and slump of high strength concrete using neural network," *Construction and Building Materials*, vol. 20, no. 9, pp. 769–775, 2006.

[27] T. Nguyen, A. Kashani, T. Ngo, and S. Bordas, "Deep neural network with high-order neuron for the prediction of foamed concrete strength," *Computer-Aided Civil and Infrastructure Engineering*, vol. 34, no. 4, pp. 316–332, 2019.

[28] M. A. Getahun, S. M. Shitote, and Z. C. Abiero Gariy, "Artificial neural network based modelling approach for strength prediction of concrete incorporating agricultural and construction wastes," *Construction and Building Materials*, vol. 190, pp. 517–525, 2018.

[29] M. Oltean and C. Groşan, "Evolving evolutionary algorithms using multi expression programming," in *European Conference on Artificial Life*Springer, Berlin, Germany, 2003.

[30] A. H. Gandomi and D. A. Roke, "Assessment of artificial neural network and genetic programming as predictive tools," *Advances in Engineering Software*, vol. 88, pp. 63–72, 2015.

[31] M. F. Javed, M. N. Amin, M. I. Shah et al., "Applications of gene expression programming and regression techniques for estimating compressive strength of bagasse ash based concrete," *Crystals*, vol. 10, no. 9, p. 737, 2020.

[32] M. Oltean and C. Grosan, "A comparison of several linear genetic programming techniques," *Complex Systems*, vol. 14, no. 4, pp. 285–314, 2003.

[33] M. Oltean and D. Dumitrescu, "Multi expression programming," *Journal of Genetic Programming and Evolvable Machines*, Kluwer, New York, NY, USA, 2002.

[34] A. H. Gandomi, A. Faramarzifar, P. G. Rezaee, A. Asghari, and S. Talatahari, "New design equations for elastic modulus of concrete using multi expression programming," *Journal of Civil Engineering and Management*, vol. 21, no. 6, pp. 761–774, 2015.

[35] S. Sharifi, S. Abrishami, and A. H. Gandomi, "Consolidation assessment using multi expression programming," *Applied Soft Computing*, vol. 86, Article ID 105842, 2020.

[36] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95-international Conference on Neural Networks*, IEEE, Perth, WA, Australia, December 1995.

[37] C. Qi, A. Fourie, and Q. Chen, "Neural network and particle swarm optimization for predicting the unconfined compressive strength of cemented paste backfill," *Construction and Building Materials*, vol. 159, pp. 473–478, 2018.

[38] F. Marini and B. Walczak, "Particle swarm optimization (PSO). A tutorial," *Chemometrics and Intelligent Laboratory Systems*, vol. 149, pp. 153–165, 2015.

[39] R. Srinivasan and K. Sathiya, "Experimental study on bagasse ash in concrete," *International Journal for Service Learning in Engineering, Humanitarian Engineering and Social Entrepreneurship*, vol. 5, no. 2, pp. 60–66, 2010.

[40] J. A. Patel and D. Raijiwala, "Experimental study on use of sugar cane bagasse ash in concrete by partially replacement with cement," *International Journal of Innovative Research in Science, Engineering and Technology*, vol. 4, no. 4, pp. 2228–2232, 2015.

[41] D. Neeraja and K. Kiran, "Experimental study on strength properties of concrete by partial replacement of cement with sugarcane bagasse ash," *Nature Environment and Pollution Technology*, vol. 13, no. 3, p. 629, 2014.

[42] K. Ganesan, K. Rajagopal, and K. Thangavel, "Evaluation of bagasse ash as supplementary cementitious material," *Cement and Concrete Composites*, vol. 29, no. 6, pp. 515–524, 2007.

[43] T. Subramani and M. Prabhakaran, "Experimental study on bagasse ash in concrete," *International Journal of Application or Innovation in Engineering & Management (IJAIEM)*, vol. 4, no. 5, pp. 163–172, 2015.

[44] S. Rukzon and P. Chindaprasirt, "Utilization of bagasse ash in high-strength concrete," *Materials & Design*, vol. 34, pp. 45–50, 2012.

[45] G. C. Cordeiro, R. D. Toledo Filho, L. M. Tavares, and E. d. M. R. Fairbairn, "Ultrafine grinding of sugar cane bagasse ash for application as pozzolanic admixture in concrete," *Cement and Concrete Research*, vol. 39, no. 2, pp. 110–115, 2009.

[46] T. S. Kumar, K. Balaji, and K. Rajasekhar, "Assessment of sorptivity and water absorption of concrete with partial replacement of cement by sugarcane bagasse ash (SCBA) and silica fume," *International Journal of Applied Engineering Research*, vol. 11, no. 3, pp. 5747–5752, 2016.

[47] N.-u. Amin, "Use of bagasse ash in concrete and its impact on the strength and chloride resistivity," *Journal of Materials in Civil Engineering*, vol. 23, no. 5, pp. 717–720, 2011.

[48] B. Hailu and A. Dinku, "Application of sugarcane bagasse ash as a partial cement replacement material," *Zede Journal*, vol. 29, pp. 1–12, 2012.

[49] S. A. Mangi, J. Norwati, H. A. Abdullah et al., "Utilization of sugarcane bagasse ash in concrete as partial replacement of cement," in *IOP Conference Series: Materials Science and Engineering*vol. 271, 2017.

[50] S. W. Dhengare, S. Raut, B. Vijay et al., "Investigation into utilization of sugarcane bagasse ash as supplementary cementitious material in concrete," *International Journal*, vol. 109, 2015.

[51] A. A. E. Hussein, "Compressive strength and microstructure of sugar cane bagasse ash concrete," *Research Journal of*

Applied Sciences, Engineering and Technology, vol. 7, no. 12, pp. 2569–2577, 2014.

[52] M. V. S. Reddy, "Utilization of sugarcane bagasse ash (SCBA) in concrete by partial replacement of cement," *IOSR Journal of Mechanical and Civil Engineering*, vol. 12, no. 6, pp. 12–16, 2015.

[53] K. Ganesan, K. Rajagopal, and K. Thangavel, "Evaluation of bagasse ash as corrosion resisting admixture for carbon steel in concrete," *Anti-Corrosion Methods and Materials*, vol. 54, 2007.

[54] M. Yashwanth, G. B. Avinash, B. G. N. Kumar et al., "An experimental study on alternative cementitious materials: bagasse ash as partial replacement for cement in structural lightweight concrete," *Indian Concrete Journal*, vol. 91, p. 51, 2017.

[55] N. Shafiq, A. Hussein, F. Nuruddin et al., "Effects of sugarcane bagasse ash on the properties of concrete," in *Engineering Sustainability*vol. 171, 2016.

[56] K. L. Priya and R. Ragupathy, "Effect of sugarcane bagasse ash on strength properties of concrete," *International Journal of Research in Engineering and Technology*, vol. 5, no. 4, pp. 159–164, 2016.

[57] S. Praveenkumar and G. Sankarasubramanian, "Mechanical and durability properties of bagasse ash-blended high-performance concrete," *SN Applied Sciences*, vol. 1, no. 12, p. 1664, 2019.

[58] Q. Xu, T. Ji, N. Wu et al., "Characteristics and applications of sugar cane bagasse ash waste in cementitious materials," *Materials*, vol. 12, no. 1, p. 39, 2019.

[59] J. C. Arenas-Piedrahita, P. Montes-García, J. M. Mendoza-Rangel, H. Z. López Calvo, P. L. Valdez-Tamez, and J. Martínez-Reyes, "Mechanical and durability properties of mortars prepared with untreated sugarcane bagasse ash and untreated fly ash," *Construction and Building Materials*, vol. 105, pp. 69–81, 2016.

[60] G. C. Cordeiro, R. D. Toledo Filho, L. M. Tavares, and E. M. R. Fairbairn, "Experimental characterization of binary and ternary blended-cement concretes containing ultrafine residual rice husk and sugar cane bagasse ashes," *Construction and Building Materials*, vol. 29, pp. 641–646, 2012.

[61] D. J. Elwell and G. Fu, *Compression Testing of Concrete: Cylinders vs. Cubes*, TRID, Washington, DC, USA, 1995.

[62] R. Siddique and G. Singh, "Utilization of waste foundry sand (WFS) in concrete manufacturing," *Resources, Conservation and Recycling*, vol. 55, no. 11, pp. 885–892, 2011.

[63] A. ACI, *318-11 ACI Building Code Requirements for Structural Concrete and Commentary*, American Concrete Institute, Farmington Hills, MI, USA, 2011.

[64] Australia, S., Australian standard for concrete structures AS 3600-2001. Australia, 175pp, 2001.

[65] Standard, B., Eurocode 2: Design of Concrete Structures—Part 1-1: General rules and rules for buildings, 2004: p. 230.

[66] JSCE, *Standard Specifications for Concrete Structures*, Structural Performance Verification, Tokyo, Japan, 2007.

[67] New Zealand Standard, *Concrete Structures Standard, NZS 3101: 2006 The Design of Concrete Structures*, New Zealand Standard, Wellington, NZ, Oceania, 2006.

[68] S. M. Mousavi, A. H. Gandomi, A. H. Alavi, and M. Vesalimahmood, "Modeling of compressive strength of HPC mixes using a combined algorithm of genetic programming and orthogonal least squares," *Structural Engineering and Mechanics*, vol. 36, no. 2, pp. 225–241, 2010.

[69] I. Azim, "Prediction model for compressive arch action capacity of RC frame structures under column removal scenario using gene expression programming," in *Structures*Elsevier, Amsterdam, Netherlands, 2020.

[70] A. H. Gandomi, A. H. Alavi, M. R. Mirzahosseini, and F. M. Nejad, "Nonlinear genetic-based models for prediction of flow number of asphalt mixtures," *Journal of Materials in Civil Engineering*, vol. 23, no. 3, pp. 248–263, 2011.

[71] M. Despotovic, V. Nedic, D. Despotovic, and S. Cvetanovic, "Evaluation of empirical models for predicting monthly mean horizontal diffuse solar radiation," *Renewable and Sustainable Energy Reviews*, vol. 56, pp. 246–260, 2016.

[72] J. Zhang, Y. Huang, G. Ma, J. Sun, and B. Nener, "A metaheuristic-optimized multi-output model for predicting multiple properties of pervious concrete," *Construction and Building Materials*, vol. 249, Article ID 118803, 2020.

[73] Y. Sun, "Development of an ensemble intelligent model for assessing the strength of cemented paste backfill," *Advances in Civil Engineering*, vol. 2020, Article ID 1643529, 6 pages, 2020.

[74] S. Xu, X. An, X. Qiao, L. Zhu, and L. Li, "Multi-output least-squares support vector regression machines," *Pattern Recognition Letters*, vol. 34, no. 9, pp. 1078–1084, 2013.

[75] K. S. Raju, "Support Vector Machine with k-fold cross validation model for software fault prediction," *International Journal of Pure and Applied Mathematics*, vol. 118, pp. 321–334, 2018.

[76] R. Kohavi, *A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection*IJCAI, Montreal, Canada, 1995.

[77] G. C. Cordeiro, R. D. Toledo Filho, and E. M. R. Fairbairn, "Effect of calcination temperature on the pozzolanic activity of sugar cane bagasse ash," *Construction and Building Materials*, vol. 23, no. 10, pp. 3301–3303, 2009.

[78] S. A. Memon, S. Khan, W. Israr et al., "Evaluating the effect of calcination and grinding of corn stalk ash on pozzolanic potential for sustainable cement-based materials," *Advances in Materials Science and Engineering*, vol. 2020, Article ID 1619480, 13 pages, 2020.

[79] A. Bahurudeen and M. Santhanam, "Influence of different processing methods on the pozzolanic performance of sugarcane bagasse ash," *Cement and Concrete Composites*, vol. 56, pp. 32–45, 2015.

[80] P. Jagadesh, A. Ramachandramurthy, R. Murugesan, and K. Sarayu, "Micro-Analytical studies on sugar cane bagasse ash," *Sadhana*, vol. 40, no. 5, pp. 1629–1638, 2015.

[81] L. M. S. Souza, "Influence of initial CaO/SiO2 ratio on the hydration of rice husk ash-Ca (OH) 2 and sugar cane bagasse ash-Ca (OH) 2 pastes," *Química Nova*, vol. 37, no. 10, pp. 1600–1605, 2014.

[82] G. C. Cordeiro, R. D. Toledo Filho, L. M. Tavares, and E. M. R. Fairbairn, "Pozzolanic activity and filler effect of sugar cane bagasse ash in Portland cement and lime mortars," *Cement and Concrete Composites*, vol. 30, no. 5, pp. 410–418, 2008.

[83] P. C. Macedo, "Performance of mortars produced with the incorporation of sugar cane bagasse ash," *Revista Ingeniería de Construcción*, vol. 29, no. 2, pp. 187–199, 2014.

[84] A. Gholampour, A. H. Gandomi, and T. Ozbakkaloglu, "New formulations for mechanical properties of recycled aggregate concrete using gene expression programming," *Construction and Building Materials*, vol. 130, pp. 122–145, 2017.

[85] I. E. Frank and R. Todeschini, *The Data Analysis Handbook*, Elsevier, Amsterdam, Netherlands, 1994.

[86] A. Golbraikh and A. Tropsha, "Beware of q2!" *Journal of Molecular Graphics and Modelling*, vol. 20, no. 4, pp. 269–276, 2002.

[87] P. P. Roy and K. Roy, "On some aspects of variable selection for partial least squares regression models," *QSAR & Combinatorial Science*, vol. 27, no. 3, pp. 302–313, 2008.

[88] H. Du, "Properties of ultra-lightweight cement composites with nano-silica," *Construction and Building Materials*, vol. 199, pp. 696–704, 2019.

Hindawi
Advances in Civil Engineering
Volume 2021, Article ID 6618407, 17 pages
https://doi.org/10.1155/2021/6618407

Research Article

Compressive Strength of Fly-Ash-Based Geopolymer Concrete by Gene Expression Programming and Random Forest

Mohsin Ali Khan,[1] **Shazim Ali Memon**,[2] **Furqan Farooq**,[3] **Muhammad Faisal Javed**,[3] **Fahid Aslam**,[4] **and Rayed Alyousef**[4]

[1]*Department of Structural Engineering, Military College of Engineering (MCE), National University of Science and Technology (NUST), Islamabad 44000, Pakistan*
[2]*Department of Civil and Environmental Engineering, School of Engineering and Digital Sciences, Nazarbayev University, Nur-Sultan 010000, Kazakhstan*
[3]*Department of Civil Engineering, COMSATS University Islamabad, Abbottabad Campus, Abbottabad 22060, Pakistan*
[4]*Department of Civil Engineering, College of Engineering in Al-Kharj, Prince Sattam bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia*

Correspondence should be addressed to Mohsin Ali Khan; moak.pg18mce@student.nust.edu.pk and Shazim Ali Memon; shazim.memon@nu.edu.kz

Received 11 November 2020; Revised 3 January 2021; Accepted 16 January 2021; Published 30 January 2021

Academic Editor: Yuantian Sun

Fly ash (FA) is a residual from thermal industries that has been effectively utilized in the production of FA-based geopolymer concrete (FGPC). To avoid time-consuming and costly experimental procedures, soft computing techniques, namely, random forest regression (RFR) and gene expression programming (GEP), are used in this study to develop an empirical model for the prediction of compressive strength of FGPC. A widespread, reliable, and consistent database of compressive strength of FGPC is set up via a comprehensive literature review. The database consists of 298 compressive strength data points. The influential parameters that are considered as input variables for modelling are curing temperature (T), curing time (t), age of the specimen (A), the molarity of NaOH solution (M), percent SiO_2 solids to water ratio ($\%\,S/W$) in sodium silicate (Na_2SiO_3) solution, percent volume of total aggregate ($\%\,A_G$), fine aggregate to the total aggregate ratio (F/A_G), sodium oxide (Na_2O) to water ratio (N/W) in Na_2SiO_3 solution, alkali or activator to the FA ratio (A_L/F_A), Na_2SiO_3 to NaOH ratio (N_s/N_o), percent plasticizer ($\%\,P$), and extra water added as percent FA ($E_W\%$). RFR is an ensemble algorithm and gives outburst performance as compared to GEP. However, GEP proposed an empirical expression that can be used to estimate the compressive strength of FGPC. The accuracy and performance of both models are evaluated via statistical error checks, and external validation is considered. The proposed GEP equation is used for sensitivity analysis and parametric study and then compared with nonlinear and linear regression expressions.

1. Introduction

Fly ash (FA) is considered as waste material resulted from thermal coal production [1]. It is carried by the gases released from the boiler and collected via electrostatic or mechanical separator [2]. The annual production of FA is 375 million tons, and its disposal cost per ton is $20 to $40 [3]. Dumping into landfills without prior treatment causes a malicious effect on the environment [4]. To sustain safe environment, effective management of waste is needed [5]. Fine particles of FA act as poisons when entering the respiratory system. Furthermore, it pollutes underground water, which is harmful to aquatic life and causes diarrhea and skin cancer [6].

Concrete is the second most usable material after water, as three tons of concrete is produced per person [7, 8]. In the world, every year 25 billion tons of concrete is produced that acquires 2.6 billion tons of cement, which will be increased

by 25% in the next ten years [9, 10]. Cement production causes a nasty impact on the atmosphere as one ton of cement emits one ton of CO_2 in the air, which alarms the ecology [11]. Also, limestone is the main source of cement, and its severe shortage may occur after 25–30 years [12, 13]. Therefore, green concrete production is needed to decline its malignant impact on the natural environment [14]. FA is used as supplementary cementitious material to produce green concrete [15]. It is worthy as it reduces the cement utilization and also its harmful effects on the ecology when dumped into landfills.

Since last two decades, the use of FA-based geopolymer concrete (GPC) is rising constantly as it reduces the consumption of cement [16–19]. FA-based GPC has been widely used in construction, but still no empirical model is developed to predict its compressive strength (f_c') on the basis of mix proportion with maximum input parameters. f_c' of FA-dependent GPC varies with different parameters like specimen age (A), curing time (t), initial curing temperature (T), molarity of NaOH solution (M), percent SiO_2 solids to water ratio (%S/W) in the formation of sodium silicate (Na_2SiO_3) solution, ratio of alkali to FA (A_L/F_A), ratio of Na_2SiO_3 to NaOH (N_s/N_o), addition of extra water as percent FA (%E_W), percentage of total aggregate by volume (%A_G), ratio of fine to total aggregate (F/A_G), and percentage of plasticizer (%P) [10, 20–27]. This generates lack of clarity in the prediction of f_c' of FA-dependent GPC. Furthermore, rapid growth of soft computing techniques for the development of empirical equation by using experimental data has been just noticed [28, 29].

Artificial intelligence (AI) techniques have been used widely in the civil engineering field for the prediction of different mechanical properties of concrete. Methods like random forest (RF) [30, 31], support vector machine (SVM) [32], artificial neural networks (ANNs) [33], adaptive neuro fuzzy interface (ANFIS) [34], decision tree (DT) [35], multivariate adaptive regression spline (MARS) [36], genetic programming (GP) [37], and gene programming (GEP) [38] were used vastly by different researchers [39–41]. Recently, ANN was used to accurately predict the mechanical properties of rice husk ash concrete [33] and workability of self-compacting concrete [42]. No empirical equation was provided in these models, which can be used practically, although these models show a strong correlation. This is due to the complex model of ANN structure which limits the wide scale adoption of ANN techniques [43]. The multicollinearity is the main issue in these models [44]. Likewise, an updated ANN technique was used to predict the compressive strength of silica fume concrete and elastic moduli of recycled aggregate concrete. Due to the complications in the proposed relationship, just a graphical interface was developed to make the model functional [45]. The comparative study of ANN and ANFIS was carried out for the prediction of compressive strength of concrete which revealed that ANFIS provides better and strong correlation than ANN [46]. RF is an ensemble machine learning technique which has been effectively used in the prediction of uniaxial compressive strength of rubberized concrete [30]. The RF gives outburst performance in modelling strength of coal grout material [31]. The adamant results were obtained in the prediction of compressive strength of self-compacting concrete with antenna search-based RF algorithm [47].

Genetic programming (GP), one of the strong soft computing techniques, is worthy as it develops a model without considering the previously established relationships [48, 49]. Recently, GP is extended to gene expression programming (GEP), which uses linear chromosomes of fixed length and encodes a small program [50]. GEP is advantageous as it gives a simple and reliable mathematical equation that can be used practically. In civil engineering, it is used as a substitute for ordinary prediction techniques [39, 51–57]. GEP is employed to predict the influence of the strength class of cement on the compressive strength of mortar, the split tensile strength of concrete, and the fresh and hardened properties of the self-compacting mix [39, 51–57]. Farooq et al. [58] predicted the compressive strength of high-strength concrete using RF, ANN, DT, and GEP, providing coefficient of determination equal to 0.96, 0.89, 0.90, and 0.90, respectively. In RF, weak learners are used as base learners. This bagging mechanism of RF provides obstinate results. GEP leads RF as it is an individual model that provides an empirical relation between input and output parameters that can be used in field calculation.

Compressive strength is the major factor to be considered in the design and analysis of concrete [59]. Intensive research is carried out to find the compressive strength of FA-dependent GPC [60, 61]. For the sustainability of FA and to save cost and time, it is needed to develop a reliable and accurate equation that would relate mix proportion and compressive strength of FA-dependent GPC. The comprehensive revision of literature reveals few empirical equations for the prediction of compressive strength of FA-dependent GPC [39, 54, 57]. However, the prediction from these empirical equations are limited to a specified experimental study and is not practicable and reliable beyond the specified dataset. Alkaroosh et al. [62] established a model to predict the compressive strength of FA-dependent GPC using 56 datasets extracted from a particular research paper [63]. The suggested equation uses no variable to counter the formation of sodium silicate solution. Also, the model illustrates a linear relation in the molarity of NaOH and compressive strength. However, other studies reported an inverse relationship between the compressive strength and molarity of NaOH solution [64]. To cover this lack of correspondence, RF and GEP techniques are used to develop a more accurate, effective, and generalized model that predicts the compressive strength of FA-dependent GPC with acceptable error. A comprehensive and detailed dataset file is established from the literature that includes cube samples of size 150 mm and 100 mm and cylindrical samples having size (200 × 100) mm (height × diameter). The ample number of data points guarantees the reliability of the model for data points outside the dataset file. The performance of the RF and GEP model is tested through statistical checks, parametric study, and sensitivity analysis and compared with nonlinear and linear regression models.

2. Research Methodology

This section covers the methodology to develop GEP and RF models for the compressive strength (f'_c) of FA-dependent GPC.

2.1. Brief Overview of Gene Expression Programming. Koza recommended the GP technique as an alternative to genetic algorithm (GA) which uses fixed length binary strings [65]. The use of nonlinear parse tree structure marks the GP as an acceptable technique. It considers the initial nonlinearity of the data. The same nonlinearity has been exercised before [62, 65]. GP is inadequate as it ignores the independent genome. The nonlinear structure of GP works as both the phenotype and genotype. It fails in the development of basic and simple model. To overcome inconsistencies in the GP algorithm, Ferreira suggested its modified version known as GEP technique [65]. It is based on the evolutionary theorem of population. The major change in GEP is the transmission of the genome towards successive generations. Another notable feature is the creation of entities using chromosome which is comprised of different genes [66]. In GEP, each gene originates from terminal set of constants, fixed length parameters, and arithmetic operations used as a function. There is a stabilized and smooth interface between chromosome level and allied functions. Chromosomes record the essential information needed for the establishment of model, and for the deduction of this information, a new language, i.e., Karva, is developed.

The flow diagram of the GEP algorithm is shown in Figure 1. The algorithm begins with the random creation of fixed length chromosomes for each individual. Then, these are similar to the expression trees (ETs). Afterward, the fitness of each individual is evaluated. For many generations, the reiteration begins with different individuals till the development of the finest outcome. For the reiteration of the population, genetic function as mutation, reproduction, and crossover are executed.

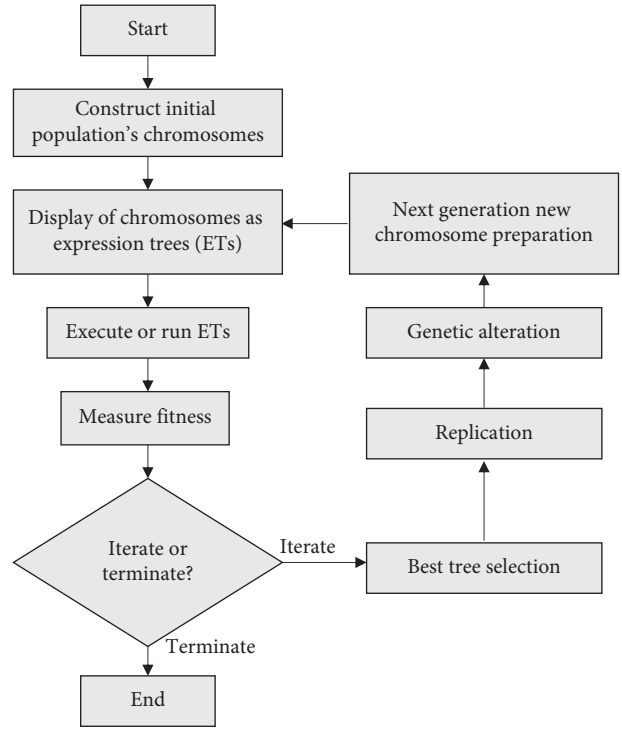


Figure 1: Flow diagram of gene expression programming [29].

2.2. Brief Review of Random Forest Regression. In 2001, Breiman proposed an improved regression technique known as random forest regression (RFR) [67]. The key characteristics of RFR are flexibility and speediness in the development of the relation between output and input parameters. Also, random forest handles large datasets more effectively than other machine learning algorithms. It has been used in different fields like in banking for the prediction of response of customer [68], prices direction in stock exchange [69], in pharmaceutical and medicine production [70], and so on. It has also been used in various engineering fields like potential mapping of ground water using geoinformatics system- (GIS-) based data [71], compressive strength prediction of high-performance concrete [35], self-compacting light-weight concrete [48], high-strength concrete [58], and so on.

The RF technique is comprised of three main steps that include the assembling of trained regression trees via training dataset, calculation of the mean value of single regression tree outcome, and the validation of predicted results via validation dataset. The original trained set is used to calculate a new trained dataset comprising of boot-strap data. In this step, some of the data points are removed and swapped with the present data points. The removed data points assembled in other dataset are called out-of-bag data points. Then, the regression function is estimated using 2/3rd of the data points, and the out-of-bag data points are used in validating the model. This process is continued till the achievement of the required accuracy.

RFR is a built-in process that deletes the data points from out-of-bag data points and uses them for validation. This is the distinctive characteristic of RFR. Finally, for each expression tree, the total error is computed showing the efficiency and accuracy of each expression tree.

2.3. Data Collection. Compressive strength (f'_c) is the key factor to design and analyze concrete. For the sustainability of FA and to save cost and time, it is needed to develop a reliable and accurate model that would relate mix proportion and f'_c of FA-based GPC.

Comprehensive dataset file was compiled from the literature [62, 63, 72–105]. The whole dataset is comprised of 298 experimental results of f'_c of FA-based GPC, which includes 31 and 166 cube samples having 100 mm and 150 mm size respectively and 101 cylindrical samples having size (200×100) mm (height × diameter). f'_c of cylindrical and cube samples is dependent on length to diameter (L/D) ratio [106, 107]. Also, f'_c of 150 mm cubes is 5% lesser than 100 mm cubes. The normalization of cube samples with cylindrical samples is shown in Table 1. The accomplishment

Table 1: Collection of data and normalization of compressive strength.

Type of sample	Number of data points	Normalization factor
Cylindrical (200 × 100) mm)	101	1
Cube (150 mm)	166	1 × 0.8
Cube (100 mm)	31	0.95 × 0.8

of detailed dataset file guarantees the accessibility and reliability of the GEP model to the data which are not utilized for the establishment of the model.

The dataset file contains data of f_c' as a response against input parameters, i.e., samples age (A) in days, initial curing temperature (T) of samples in degree Celsius, molarity of NaOH solution (M), percent SiO_2 solids to water ratio ($\%\, S/W$) in the formation of sodium silicate (Na_2SiO_3) solution, ratio of alkali to FA (A_L/F_A), ratio of Na_2SiO_3 to NaOH (N_s/N_o), addition of extra water as percent FA ($\%\, E_W$), percentage of total aggregate by volume ($\%\, A_G$), ratio of fine to total aggregate (F/A_G), and percentage of plasticizer ($\%\, P$). The collected samples are all heat cured for 24-hour duration at various curing temperatures as the increase in f_c' after 24-hour curing time is insignificant [63]. Due to the geo-polymerization, GPC shows higher early strength; therefore, less research is found in the literature for prolonged curing time. Also, Van Jaarsveld et al. [108] reported no increment in f_c' for prolonged curing duration after 24 hours. The distribution of explanatory variables on wide range guarantees the best performance of the model [109]. For all the selected explanatory parameters, the frequency distribution and cumulative percent are shown in Figure 2.

To develop a more generalized model, both cylindrical and cube samples are considered. The range, mean values, and standard deviation of response and explanatory parameter are presented in Table 2. To achieve a reliable prediction of f_c', it is recommended to use the model within the specified range.

To evaluate the reliability and validity of the data points, several trials were performed. The divergence of data points greater than 20% was excluded in the development of the model and performance evaluation phase. A total of 298 data points were used to establish a reliable model for f_c' of FA-dependent GPC. The data points were randomly divided into two statistically consistent datasets, i.e., training set (30%–90% data points) and a validation set (70%–208% data points) [29]. Training data points are used to train the model, that is, genetic evolution and validation data points are utilized to verify and calibrate the generalization capability of the developed model as described in the literature [56].

2.4. GEP Model Development. In the first step, the most effective parameters for compressive strength (f_c') of FA-dependent GPC were chosen to establish a simplified model. The performance evaluation via multiple initial runs indicates to calculate f_c' of FA based GPC as a function of the following equation.

$$f_c' = f\left(T, A, M, \%\, \frac{S}{W}, \frac{A_L}{F_A}, \frac{N_S}{N_O}A_G, \frac{F}{A_G}, \%\, P, \%\, E_W\right). \quad (1)$$

The number of genes, chromosomes, and expression trees (ETs) are distinguished to develop the GEP expression. The execution time of the program is controlled using population size (number of chromosomes). The combination of genes leads to chromosomes that are used in coding the sub-expression trees (sub-ETs). The complexity of the predictive GEP model reflects to use population size of 150. The configuration of the model in the program relies on the head size, the number of genes that decide the complexity of each term, and the sum of sub-ETs of the model. Hence, the genes and head size which are 3 and 10, respectively, are used for the establishment of the reliable model. The genetic operators are used for the genetic variation of chromosomes. During mutation, the random selection of tail or head of genes is executed and substituted with component of function or terminal sets randomly. The transposition performs the substitution of insertion sequence (IS) and the root insertion sequence (RIS) inside the chromosome. Then, in recombination, chromosomes are combined and divided into two to replace their components. To obtain good algorithm, the suggested setting in the previous study has been exercised [39]. GeneXproTool is used for the execution of the GEP algorithm. Table 3 presents the settings of the parameters used in the execution of the GEP algorithm, to develop a good model.

2.5. Criteria for Evaluation of Model Performance. To verify the performance of the developed models, the coefficient of correlation (R) is usually used. Because of its insensitivity to division and multiplication of output to a constant, it cannot be merely utilized for studying the performance of the model [110]. Therefore, root mean squared error (RMSE), relative square error (RSE), mean absolute error (MAE), and relative root mean square error (RRMSE) are also checked. The performance index (ρ) covers the function of both RRMSE and R, so the performance evaluation of the predictive models using ρ is highly recommended [109]. The error checks equations are given as equations (2)–(7).

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(\exp_i - \text{pred}_i\right)^2}{n}}, \quad (2)$$

$$\text{MAE} = \frac{\sum_{i=1}^{n}\left|\exp_i - \text{pred}_i\right|}{n}, \quad (3)$$

$$\text{RSE} = \frac{\sum_{i=1}^{n}\left(\exp_i - \text{pred}_i\right)^2}{\sum_{i=1}^{n}\left(\overline{e} - te_i\right)^2}, \quad (4)$$

$$\text{RRMSE} = \frac{1}{|\overline{e}|}\sqrt{\frac{\sum_{i=1}^{n}\left(\exp_i - \text{pred}_i\right)^2}{n}}, \quad (5)$$

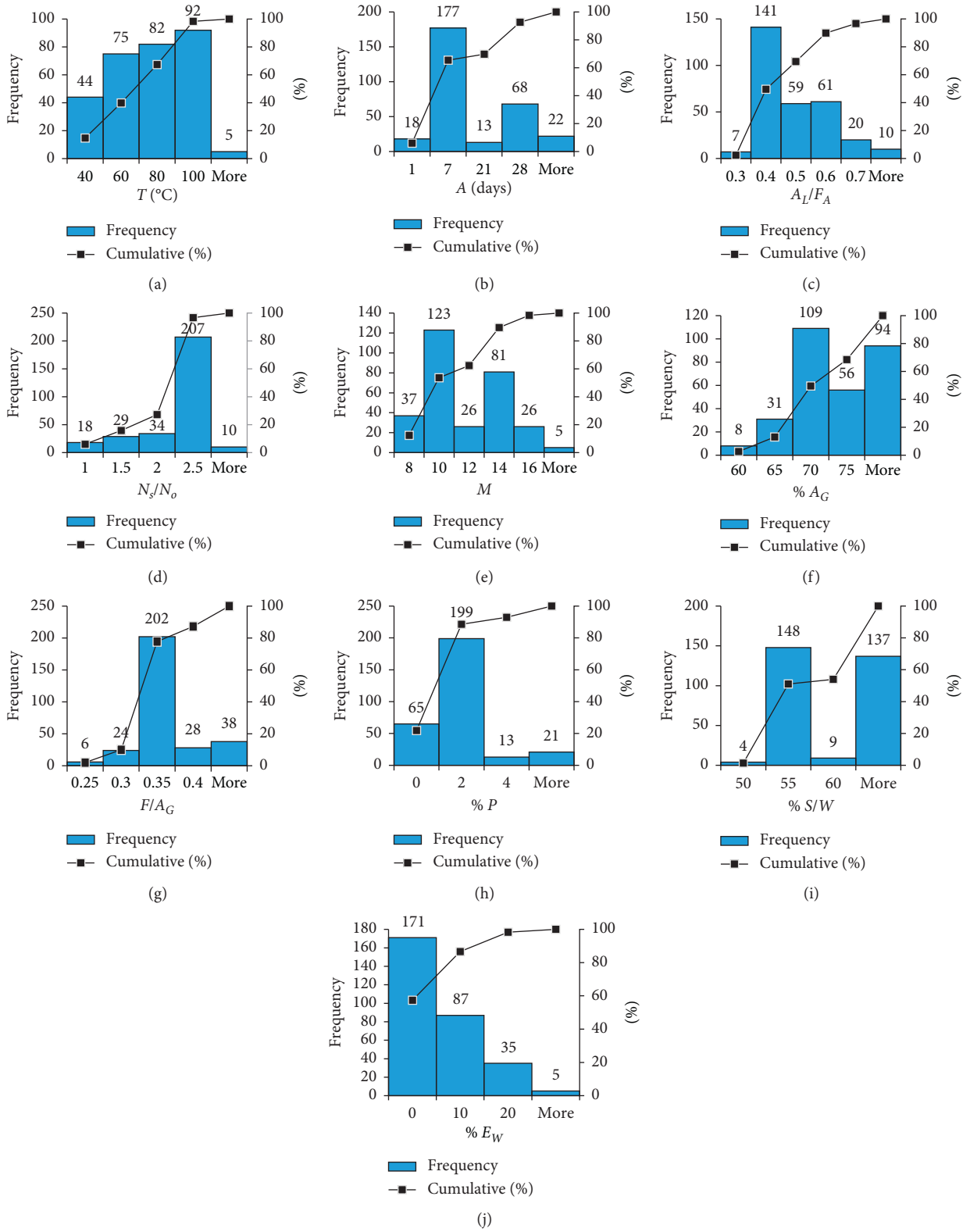


Figure 2: Frequency and cumulative percent of selected explanatory variables.

Table 2: Range, mean, and standard deviation of response and explanatory variable.

Parameters	Maximum value	Minimum value	Mean value	Standard deviation
Output variables				
T (^{O}C)	120	23	71.57	24.61
A (days)	540	1	20.87	45.73
A/F	0.3	1	0.4545	0.1187
N_S/N_O	4	0.4	2.275	0.5168
M	20	8	11.68	2.6415
(% A_G)	80	60	72	4.753
F/A_G	0.5	0.2	0.3568	0.0493
% P	11.3	0	1.998	2.326
% S/W	81.4	43.4	61.68	10.167
% E_W	35	0	3.889	6.341
Response				
f_c'(MPa)	63	8.2	37	11.154

Table 3: The setting of parameters of the GEP algorithm.

Parameters	Adjusted GEP setting
General parameters	
Number of chromosomes	150
Number of genes	3
Head size	10
Linking function	Multiplication
Functions set	+, −, /, ×, 3
Arithmetical operators	
Constants/gene	10
Type of data	Floating data
Upper bound value	10
Lower bound value	−10
Gene operator	
Mutation rate	0.001380
Inversion rate	0.005460
IS transportation rate	0.005460
RIS transportation rate	0.005460
Gene recombination rate	0.007550
Gene transportation rate	0.002770

$$R = \frac{\sum_{i=1}^{n}\left(\exp_i - \overline{\exp}_i\right)\left(\text{pred}_i - \overline{\text{pred}}_i\right)}{\sqrt{\sum_{i=1}^{n}\left(\exp_i - \overline{\text{pred}}_i\right)^2 \sum_{i=1}^{n}\left(\text{pred}_i - \overline{\text{pred}}_i\right)^2}}, \quad (6)$$

$$\rho = \frac{\text{RRMSE}}{1+R}, \quad (7)$$

where $\exp_i$, pred_i, $\overline{\exp}_i$, and $\overline{\text{pred}}_i$ are the i^{th} experimental outcome, predicted model outcome, experimental average value, and average predicted model outcome, respectively while n indicates the total number of data samples. The higher R value and lower MAE, RMSE, RRMSE, and RSE values replicate the fineness of models. For a strong correlation, the R values should be higher than 0.8 (1 for the ideal model) [111]. Also, the ρ value would be nearly equal to zero.

3. Results and Discussion

3.1. GEP Expression for Compressive Strength of FA-Dependent Geopolymer Concrete. The expression tree given by the GEP algorithm is shown in Figure 3, which is further decoded to get an empirical equation for the compressive strength of FA-dependent GPC. The ETs are comprised of five arithmetic operators, i.e., −, +, /, ×, $\sqrt[3]{\cdot}$.

d_o: curing temperature (T) in degree Celsius, d_1: age of the specimen (A), d_2: alkali or activator to the FA ratio (A_L/F_A), d_3: Na_2SiO_3 to NaOH ratio (N_s/N_o), d_4: molarity of NaOH solution (M), d_5: percent volume of total aggregate (%A_G), d_6: fine aggregate to total aggregate ratio (F/A_G) d_7: percent plasticizer (% P), d_8: percent SiO_2 solids to water ratio (%S/W), and d_9: extra water added as percent FA (%E_W).

Equation (8) can be used for the prediction of compressive strength (f_c') of FA-dependent GPC (MPa). It consists of four variables, i.e.,A, B, C, and D, presented as equations (9)–(12), which are extracted from sub-ETs 1, 2, 3, and 4, respectively, as presented in Figure 3.

$$f_c'(\text{MPa}) = A \times B \times C \times D, \quad (8)$$

where

$$A = \sqrt[3]{\%\,\frac{S}{W}} - \%\,P + \left(M \times \frac{F}{A_G} \times \frac{A_L}{F_A} \times 6.6\right) + \%\,E_W - \%\,A_G, \quad (9)$$

$$B = -\sqrt[3]{\frac{A+80}{0.08\,(T-18)} + \frac{N_S}{N_O} + M}, \quad (10)$$

$$C = \frac{F}{A_G} - \left(\left(M \times \%\,E_W\right) - \frac{0.0003}{\left(\left(N_S/N_O\right) - \%\,E_W\right)} - 0.0003, \quad (11)$$

$$D = \sqrt[3]{\frac{1.2\,(\%\,P - \%\,(S/W))}{T}} + \sqrt[3]{\frac{0.2}{\left((F/A_G)\right)}} + 0.8. \quad (12)$$

3.2. Evaluation of the Performance of the GEP Model. Figure 4(a) shows the comparison of two regression lines, namely, the GEP model output values and experimental values for both the validation set data and training set data.

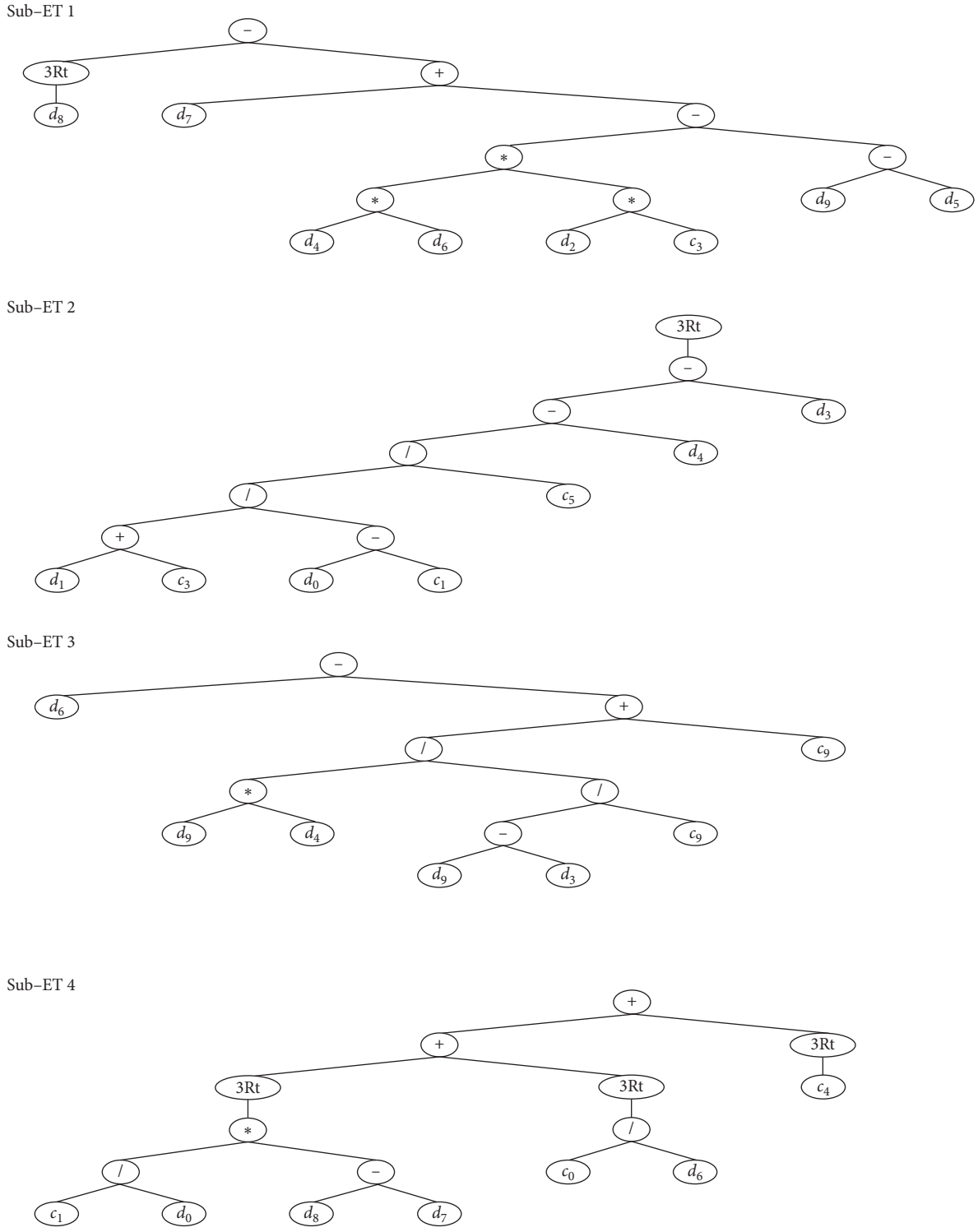


Figure 3: Expression trees (ETs) provided by the GEP algorithm.

The slope of the regression lines shows a strong correlation, i.e., 0.9892 and 1.000 for validation set data and training set data, respectively.

The absolute error between the output of GEP model and experimental values is shown in Figure 4(b). It provides an idea of maximum percent error in the GEP model. The maximum error percentage and mean error percentage are computed as 8.32% and 6.47%, respectively, which approves the similarity between GEP model outcomes and experimental values. Also, the frequency of the maximum error is less. Nearly 90% of GEP model outcomes of the validation dataset have an error of less than 10%, and the average percent error is below 5.56%. This confirms the reliability and generalization capability of the GEP model.

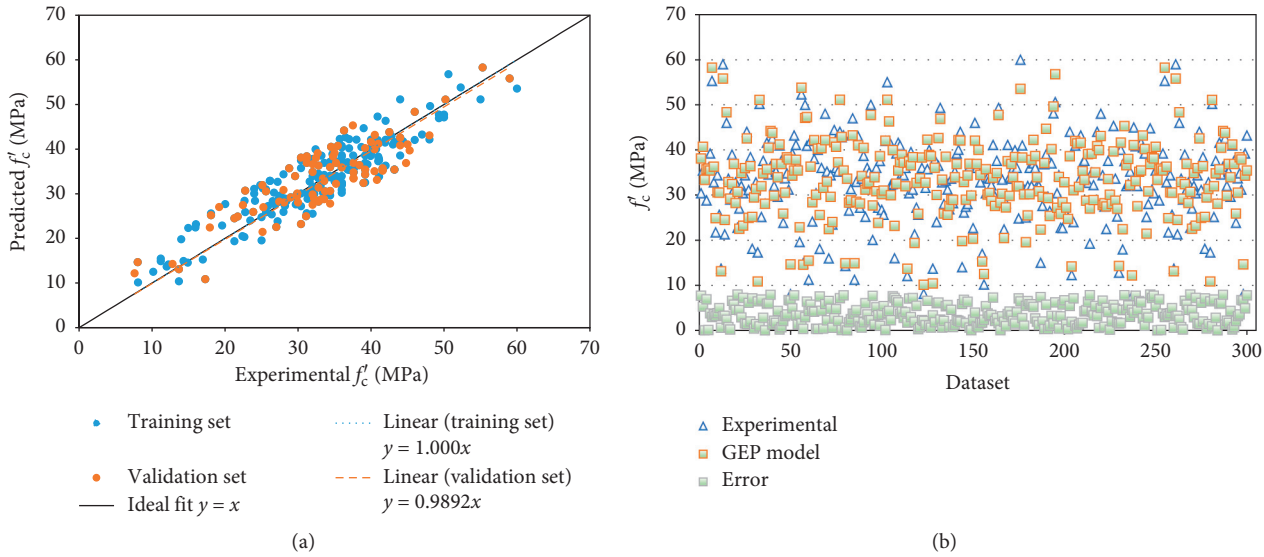


Figure 4: Performance evaluation of the GEP model. (a) Comparison between model and experimental outcomes for compressive strength from training and validation set data. (b) Absolute error plot of GEP predicted outcomes and experimental values.

For the reliable and accurate GEP model, the ratio of total data points to the total input variables should be minimum three [109]. This research uses a higher value equal to 30. The statistical checks for both validation data points and training data points are listed in Table 4. For the GEP model, MAE, RMSE, and RSE of training data points are calculated as 5.832, 5.971, and 0.325, respectively, and 2.057, 2.643, and 0.0675 for validation data points. The similarity in the statistical checks guarantees the generalization capability of the GEP model. Table 4 also shows that ρ for both sets reaches zero. So, the presented GEP model could be valid for new data points.

Different statistical measures are also considered for the external validation of the GEP model. The literature recommended that the inclination (slope) of one of the regression lines (k' or k) crossing the origin should be nearly equal to 1 [38]. Table 5 shows that the slope of regression lines is 0.995 and 1.001, which verifies the correlation and correctness. The literature also recommended that the square of correlation coefficient between the experimental and model predictive output ($R_o'^2$) or between model predictive output and experimental values (R_o^2) should come near 1 [112]. Table 5 confirms the validity of the GEP model. So, the proposed GEP model is not just a correlation.

3.3. Evaluation of Random Forest Regression Model. Random forest regression technique is an ensemble algorithm that utilizes weak learner as a supervised learner and provides a best-performed model based on the coefficient of correlation (R) as shown in Figure 5. This algorithm divides the model into twenty submodels based on different n-estimator and gives model with maximum R. The mean ensemble R is equal to 0.9732 which depicts that all the twenty submodels strongly correlate with the predicted and experimental values. Amongst all these, the submodel with 40 estimators gives outburst performance with maximum R equal to 0.9826. It is attributable to the use of weak learners as a decision which is used in ensemble algorithm [58].

The relation between the response and the predictor is shown via the slope of regression lines in Figure 6(a). The RF algorithm gives noticeable slope of the regression line as 1.000 and 0.9913 for training set data and validation set data, respectively, which proves the superiority of the RF algorithm.

The absolute error plot between the RF algorithm predicted values and experimental values is presented in Figure 6(b). In comparison with the GEP model, the RF model shows less error as the maximum percent error and average percent error are calculated as 4.89% and 2.14%. The RF algorithm yields outstanding results but does not provide an empirical equation like GEP.

Furthermore, the performance of the RF algorithm-based model is also verified through statistical error checks. Table 4 shows that statistical error checks for RF algorithm-predicted values are lesser than those of the GEP model predicted outputs, in both the training and validation stage. This confirms that the RF algorithm gives good performance than GEP model as it is an ensemble one that uses the decision trees as weak learners [58]. Also, $R_o'^2$ and R_o^2 are used for its external validation of RF model as tabulated in Table 5. Their values are calculated near to 1, which verifies that RF algorithm does not work as simple correlation.

3.4. GEP Model Comparison with Linear and Nonlinear Regression Models. The past research reveals that for f_c' of FA-dependent GPC, no GEP model has been developed using the influential input parameters considered in this study. So, it is needed to develop nonlinear and linear regression expressions, for the same dataset, and compare it with the GEP model presented as equation (8). Equations (13) and (14) present the linear and nonlinear regression equations, respectively.

Table 4: Comparison of statistical measures amongst GEP, RF, nonlinear, and linear regression models.

Model	RMSE		RSE		MAE		RRMSE (%)		R		ρ	
	T_{RNG}[1]	V_{LDN}[2]	T_{RNG}	V_{LDN}	T_{RNG}	V_{LDN}	T_{RNG}	V_{LDN}	T_{RNG}	V_{LDN}	T_{RNG}	V_{LDN}
RF	3.034	1.986	0.193	0.0350	2.876	1.862	10.084	4.163	0.9826	0.9943	0.0546	0.02087
GEP	5.971	2.643	0.325	0.0675	5.823	2.057	16.949	4.949	0.8586	0.9643	0.0911	0.02519
Linear	6.986	5.546	0.589	0.3040	6.543	4.967	19.20	10.21	0.8074	0.8976	0.1062	0.05382
Nonlinear	6.593	5.054	0.497	0.2980	6.053	4.875	18.53	9.021	0.8357	0.9247	0.1009	0.04687

[1]$TRNG$ shows training set data. [2]$VLDN$ shows validation set data.

Table 5: External validity of the proposed GEP and RF models.

Expression	Constraint	GEP model	RF model
$k = \sum_{i=1}^{n} (\exp_i \times \text{pred}_i) / \sum_{i=1}^{n} (\exp_i^2)$	$0.85 < k < 1.15$	1.001	1.000
$k' = \sum_{i=1}^{n} (\exp_i \times \text{pred}_i) / \sum_{i=1}^{n} (\text{pred}_i^2)$	$0.85 < k' < 1.15$	0.995	0.9995
$R_o^2 = 1 - (\sum_{i=1}^{n} (\text{pred}_i - \exp_i^o)^2 / \sum_{i=1}^{n} (\text{pred}_i - \overline{\text{pred}_i^o})^2), \exp_i^o = k \times \text{pred}_i$	$R_o^2 \cong 1.0$	0.9998	0.9965
$R_o'^2 = 1 - (\sum_{i=1}^{n} (\exp_i - \text{pred}_i^o)^2 / \sum_{i=1}^{n} (\exp_i - \overline{\exp_i^o})^2), \text{pred}_i^o = k' \times \exp_i$	$R_o'^2 \cong 1.0$	0.9849	0.9994

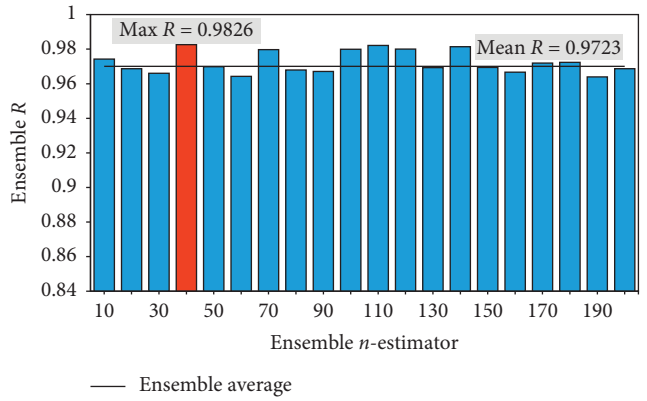


Figure 5: A random forest regression model with twenty submodels.

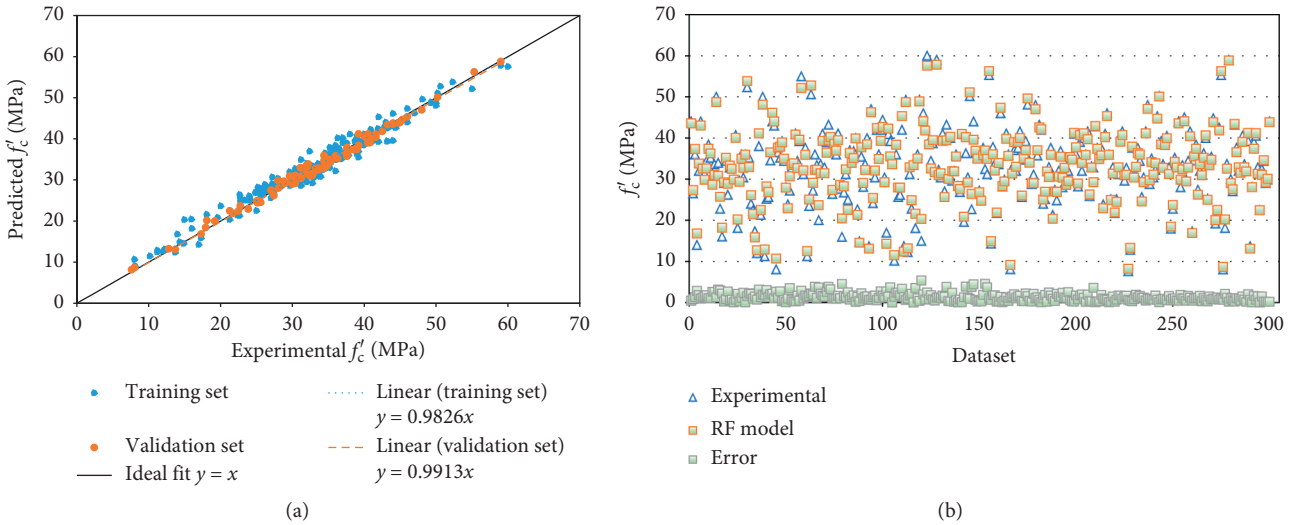


Figure 6: Performance evaluation of RF model. (a) Comparison between model and experimental outcomes for compressive strength from training and validation set data. (b) Absolute error plot of RF predicted outcomes and experimental values.

$$f_c' = 12.8 + 0.23T + 0.04A - 27\frac{A_L}{F_A} + 1.13\frac{N_S}{N_o} - 0.4M + 0.64A_G\% - 0.4\frac{F}{A_G} + 1.3P\% - 0.45\frac{S}{W}\% - 0.7E_W\%, \tag{13}$$

$$\begin{aligned} f_c' = {} & -7.6 + 1.18T^{0.68} + 0.35A^{0.63} - 25.8\left(\frac{A_L}{F_A}\right)^{2.9} + 1.8\left(\frac{N_S}{N_o}\right)^{0.44} \\ & - 0.009M^{2.24} + 0.76\left(A_G\%\right)^{0.93} - 0.37\left(\frac{F}{A_G}\right)^{1.06} \\ & + 2.25\left(P\%\right)^{0.72} - 0.08\left(\frac{S}{W}\%\right)^{1.34} - 0.27\left(E_W\%\right)^{1.32}. \end{aligned} \tag{14}$$

Figure 7 compares the results of the GEP model and nonlinear and linear regression models. For all three models, the statistical checks like RSE, MAE, RMSE, RMSE%, R, and ρ are mentioned in Table 4. ρ and RMSE of the GEP model for both validation set and training set are lesser than those of the linear and nonlinear regression models. ρ_{training} and $\text{RMSE}_{\text{training}}$ for the GEP model are 14% and 14.5% lower than those of the linear expression, respectively. Also, in the validation phase, the GEP model performs better than nonlinear regression expression as $\rho_{\text{validation}}$ differs by 44%. Figure 7 illustrates that linear and nonlinear regression models fail to cover a large range of f_c' effectively. Hence, the application of regression expression is restricted.

Some limitation of regression analysis like the use of predefined equations either nonlinear or linear and pre-assumption of residuals normality restricts its application [111], while GEP modelling chooses the nonlinear relation between input and output parameters effectively and provides a higher generalized model, which significantly reduces the error as compared to regression analysis.

3.5. Sensitivity and Parametric Analysis. Sensitivity analysis (SA) checks the relative contribution of input parameters considered to predict the compressive strength (f_c') of FA-dependent GPC, via equation (15) and (16). SA shows the reliance of output on input parameters.

$$N_j = f_{\max}\left(y_j\right) - f_{\min}\left(y_j\right), \tag{15}$$

$$\text{SA} = \frac{N_j}{\sum_n^{i=1} N_j}, \tag{16}$$

where $f_{\min}(y_j)$ and $f_{\max}(y_j)$ are the j^{th} minimum and maximum predictive model output, respectively while input values are kept constant at mean value. N_j gives the range of j^{th} input variable by taking the difference between $f_{\max}(y_j)$ and $f_{\min}(y_j)$. Both training data points and validation points are consistent; therefore, SA and a parametric study were carried out for only training data points [39, 111]. The result of the sensitivity analysis is presented in Figure 8. It verifies that the relative contribution of input variables is similar in the perspective of material engineering.

The GEP empirical equation, i.e., equation (8), is used to evaluate the effectiveness of influential input parameters by conducting parametric study. The parametric analysis of the GEP model is presented in Figure 9. The changes in compressive strength were noted against the change in the value of only one input parameter from maximum to minimum, and the rest of all input variables are kept at mean value.

The curing temperature in the most important parameter to control the compressive strength (f_c') of FA-dependent GPC, as shown in Figure 8 which reflects that curing temperature comparatively contributes 25.3%. Figure 9 illustrates an increase in f_c' at different rates with an increase in A, T, (N_s/N_o), $\%A_G$, (F/A_G), and $\%P$ while it decreases with (A_L/F_A), $\%E_W$, $(\%S/W)$, and M.

The alkali-activating solution being used in the GPC liberates silicates and hydroxides that form strong alumina silicate polymeric structure. As to speed up its reaction process with the source material, the GPC needs additional heat: to improve the mechanical properties of GPC. f_c' increases as curing temperature increases up to 100°C as shown in Figure 9. After 100°C, the loss in moisture from concrete decreases its strength [64]. Wardhono et al. [77] showed through scanning electron microscopy (SEM) that after 240 days, the gel fills out the interior voids, which results in the formation of semihomogenous, but compacted, microstructure. Therefore, after 240 days, the decline in the incremental rate is noted. The change in total aggregate is related to the fine aggregate to total aggregate ratio. f_c' increases with increment in total aggregate amount as shown in Figure 9.

The ratio of alkali-to-FA and sodium silicate-to-sodium hydroxide and molarity of NaOH are all linked. Sodium silicate changes the microstructure and significantly increases the compressive strength. Therefore, preparation of sodium silicate solution with high ratio of percent silica-to-water is needed. Low alkali-to-FA ratio combines with high sodium silicate-to-sodium hydroxide ratio, and less molar solution of NaOH will result in greater f_c'. However, the NaOH solution should be sufficient to finish the dissolution process. Same results have also been reported in the literature [78].

Total water used in GPC is the combination of water needed for the preparation of sodium hydroxide solution and sodium silicate solution and the extra water added for adjusting the workability. For the workable GPC mix and to

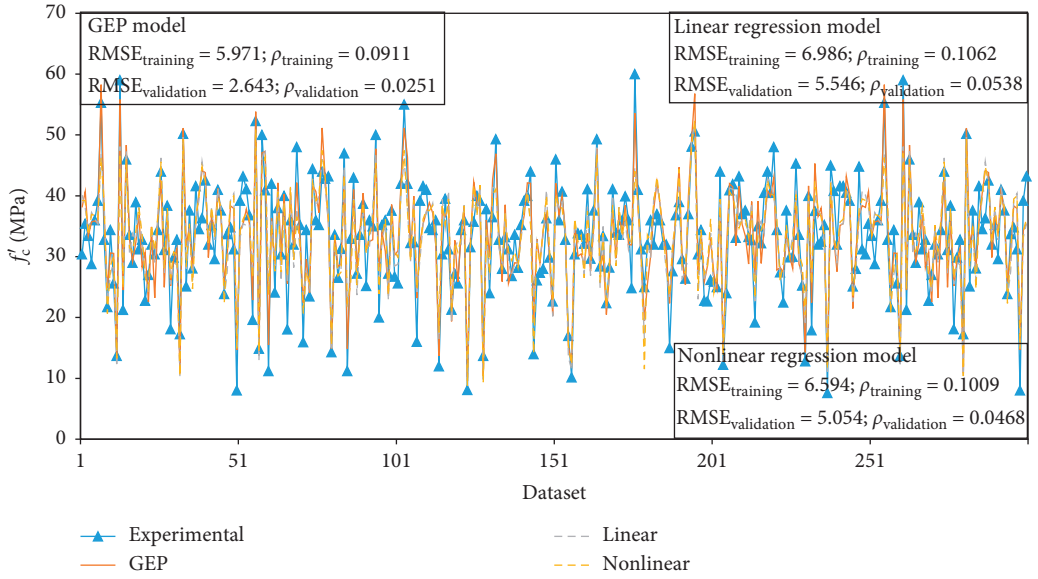


Figure 7: The divergence between GEP and nonlinear and linear regression models.

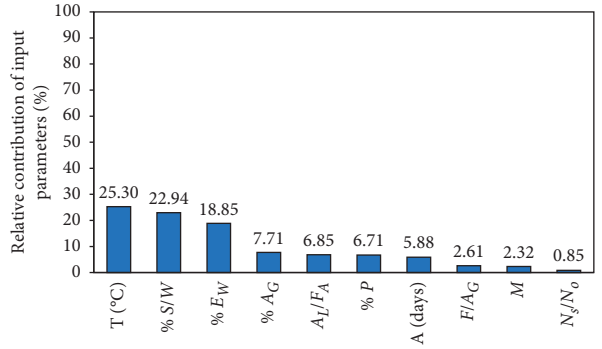


Figure 8: Percent contribution of chosen input parameters.

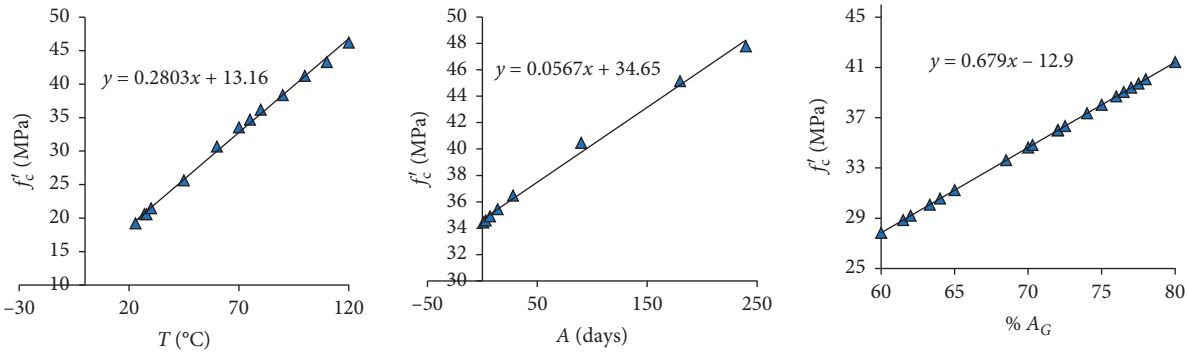


Figure 9: Continued.

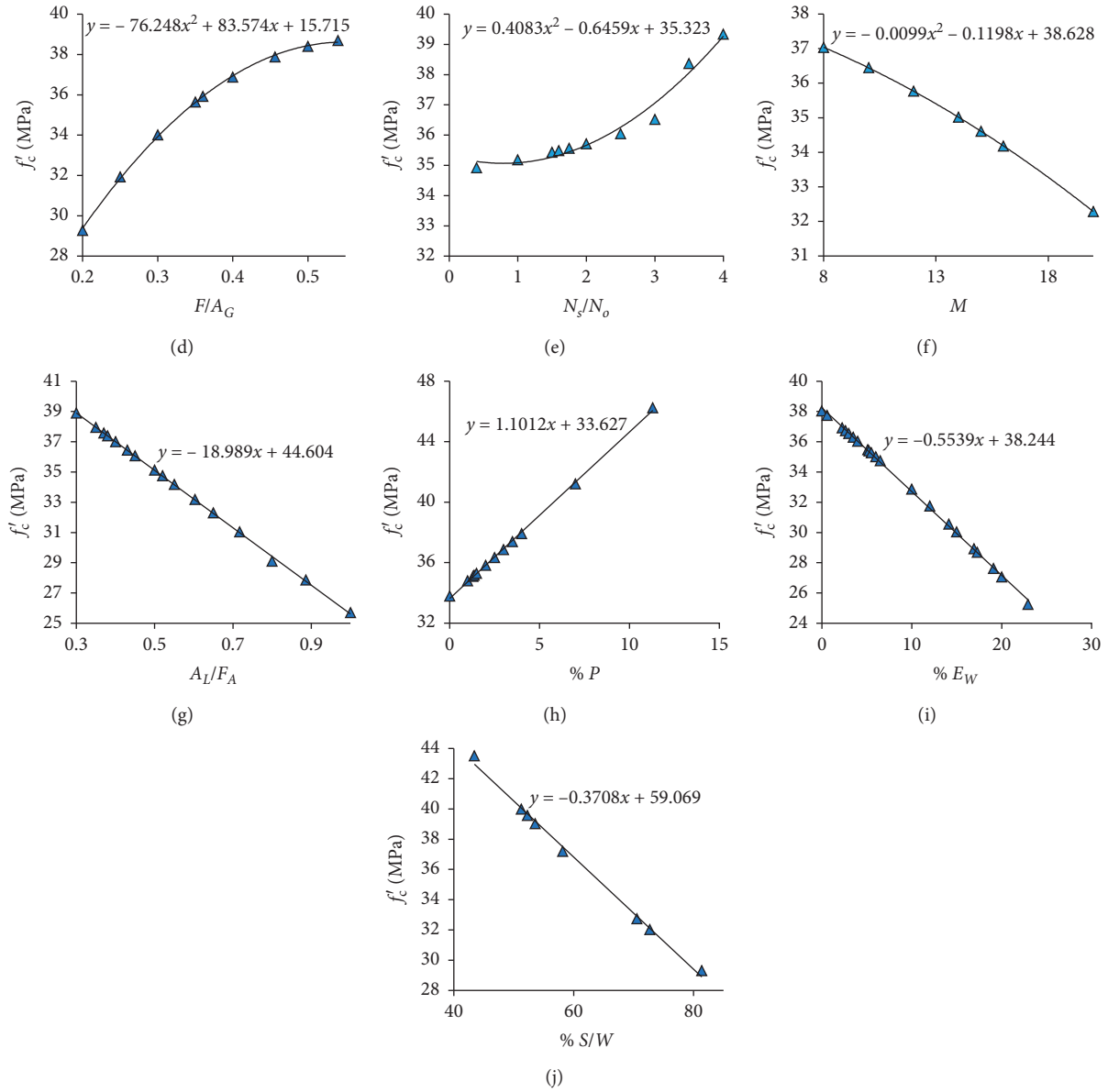


Figure 9: The trend of compressive strength with chosen input parameters.

avoid cracks, it is essential to add a plasticizer and extra water [95]. Figure 9 shows that the relative contribution of plasticizer or extra added water to f_c' is 6.71% and 18.85%, respectively. The extra added water beyond certain limit leads to segregation and bleeding of green concrete.

The results in Figure 9 are linked with previous literature [78, 95]. The parametric analysis accurately shows the effect of input parameters to predict f_c' of FA-dependent GPC.

4. Limitations and Recommendation for Future Work

The research work performed in this article does have certain drawbacks; however, it can be counted as data-mining-based research. The broadness and comprehensiveness of the data is essential for the reliability and proficiency of the predictive models. The range of the datasets used in this research was restricted to 298 experimental data points. This research did not consider the compressive strength of fly-ash-based geopolymer concrete at elevated temperature. Also, this study lacks in providing the empirical relation for other mechanical properties of FGPC like split tensile strength and flexural strength as limited research is available in the literature for both the mechanical properties. In fact, an appropriate testing dataset should be completed as it is essential part in engineering viewpoint. However, this research considered a wide range dataset with ten most influential parameters for modelling compressive strength of FA-dependent GPC.

Furthermore, it is also recommended that the new database developed should be investigated with various supervised machine learning techniques like artificial neural network (ANN), recurrent neural network (RNN), support

vector machine learning (SVM), adaptive neuro fuzzy interface (ANFIS), and multivariate adaptive regression spline (MARS).

5. Conclusions

In this study, random forest (RF) and gene expression programming (GEP) are used to develop a mathematical expression for the compressive strength f_c' of fly-ash- (FA-) dependent geopolymer concrete (GPC). The RF and GEP models are developed on the data collected from the past research, and the most effective variables are considered as input parameters. The proposed GEP empirical expression can be used for the utilization of toxic FA in place of dumping into landfills. This would eventually lead to sustainable green construction. Following are the conclusions deducted via a supervised machine learning algorithm.

(1) The highest R and lowest error checks are observed in the RF model as compared to GEP, nonlinear, and linear regression models. The RF as ensemble machine learning algorithm gives a remarkable performance with R, MAE, RMSE, RSE, and ρ equal to 0.9826, 2.896, 3.034, 0.193, and 0.0546 for training dataset, respectively, and 0.9943, 1.862, 1.986, 0.0350, and 0.02087 for validation dataset, respectively. Also, RF and GEP model accurately meets the specifications for external validation.

(2) RF model outburts performance but lacks in providing an empirical equation. In comparison with nonlinear and linear regression models, the GEP model gives outburst performance and provides an empirical expression, which is suitable for the preliminary design of FA-dependent GPC.

(3) The sensitivity analysis reveals that curing temperature is the most sensitive and dominant parameter in handling the production of FA-dependent GPC. The parametric study of the GEP model shows that the model correctly covers the effect of all explanatory variables.

(4) Furthermore, it is recommended to perform a leachate study before the addition of FA as geopolymer material.

Data Availability

The detailed excel 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.

Acknowledgments

This study was sponsored by the Deanship of Scientific Research at Prince Sattam Bin Abdulaziz University under research project no. 2020/01/16810.

References

[1] S. E. Aprianti, "A huge number of artificial waste material can be supplementary cementitious material (SCM) for concrete production-a review part II," *Journal of Cleaner Production*, vol. 142, pp. 4178–4194, 2017.

[2] A. Akbar, F. Farooq, M. Shafique, F. Aslam, R. Alyousef, and H. Alabduljabbar, "Sugarcane bagasse ash-based engineered geopolymer mortar incorporating propylene fibers," *Journal of Building Engineering*, vol. 33, Article ID 101492, 2021.

[3] A. Dwivedi and M. K. Jain, "Fly ash–waste management and overview: a review," *Recent Research in Science and Technology*, vol. 6, pp. 30–35, 2014.

[4] O. A. Abdulkareem, A. M. Mustafa Al Bakri, H. Kamarudin, I. Khairul Nizar, and A. e. A. Saif, "Effects of elevated temperatures on the thermal behavior and mechanical performance of fly ash geopolymer paste, mortar and lightweight concrete," *Construction and Building Materials*, vol. 50, pp. 377–387, 2014.

[5] N. Nordin, M. M. A. B. Abdullah, M. F. M. Tahir, A. V. Sandu, and K. Hussin, "Utilization of fly ash waste as construction material," *International Journal of Conservation Science*, vol. 7, pp. 161–166, 2016.

[6] N. Ghazali, K. Muthusamy, and S. Wan Ahmad, "Utilization of fly ash in construction,"in Proceedings of the IOP ConferenceSeries: Materials Science and Engineering, PostgraduateSymposium in Civil and Environmental Engineering 2019, Batu Pahat, Malaysia, Universiti Tun Hussein Onn Malaysia, IOP Publishing, pp. 012023, March 2019.

[7] F. Farooq, A. Akbar, R. A. Khushnood, W. L. B. Muhammad, S. K. U. Rehman, and M. F. Javed, "Experimental investigation of hybrid carbon nanotubes and graphite nanoplatelets on rheology, shrinkage, mechanical, and microstructure of SCCM," *Materials*, vol. 13, p. 230, 2020.

[8] K. M. Liew and A. Akbar, "The recent progress of recycled steel fiber reinforced concrete," *Construction and Building Materials*, vol. 232, Article ID 117232, 2020.

[9] K. A. Baumert, *Navigating the Numbers: Greenhouse Gas Data and International Climate Policy*, World Resources Institute, Washington, DC, USA, 2005, http://http//www.wri.org/publication/navigating-the-numbers.

[10] A. Wongsa, A. Siriwattanakarn, P. Nuaklong, V. Sata, P. Sukontasukkul, and P. Chindaprasirt, "Use of recycled aggregates in pressed fly ash geopolymer concrete," *Environmental Progress & Sustainable Energy*, vol. 39, no. 2, 2020.

[11] A. Akbar and K. M. Liew, "Influence of elevated temperature on the microstructure and mechanical performance of cement composites reinforced with recycled carbon fibers," *Composites Part B: Engineering*, vol. 198, Article ID 108245, 2020.

[12] B. Sumanth Kumar, A. Sen, and D. Rama Seshu, "Shear strength of fly ash and GGBS based geopolymer concrete," in *Lecture Notes in Civil Engineering*, vol. 68, pp. 105–117, Springer, Berlin, Germany, 2020.

[13] F. Farooq, S. K. U. Rahman, A. Akbar et al., "A comparative study on performance evaluation of hybrid GNPs/CNTs in conventional and self-compacting mortar," *Alexandria Engineering Journal*, vol. 59, no. 1, pp. 369–379, 2020.

[14] K. Snehal, B. B. Das, and M. Akanksha, "Early age, hydration, mechanical and microstructure properties of nano-silica blended cementitious composites," *Construction and Building Materials*, vol. 233, Article ID 117212, 2020.

[15] Q. Wang, D. Wang, and H. Chen, "The role of fly ash microsphere in the microstructure and macroscopic properties

of high-strength concrete," *Cement and Concrete Composites*, vol. 83, pp. 125–137, 2017.

[16] H. Y. Zhang, G. H. Qiu, V. Kodur, and Z. S. Yuan, "Spalling behavior of metakaolin-fly ash based geopolymer concrete under elevated temperature exposure," *Cement and Concrete Composites*, vol. 106, Article ID 103483, 2020.

[17] C. R. Meesala, N. K. Verma, and S. Kumar, "Critical review on fly-ash based geopolymer concrete," *Structural Concrete*, vol. 21, no. 3, pp. 1013–1028, 2020.

[18] M. E. Gülşan, R. Alzeebaree, A. A. Rasheed, A. Niş, and A. E. Kurtoğlu, "Development of fly ash/slag based self-compacting geopolymer concrete using nano-silica and steel fiber," *Construction and Building Materials*, vol. 211, pp. 271–283, 2019.

[19] K. Kondepudi and K. V. L. Subramaniam, "Rheological characterization of low-calcium fly ash suspensions in alkaline silicate colloidal solutions for geopolymer concrete production," *Journal of Cleaner Production*, vol. 234, pp. 690–701, 2019.

[20] N. Li, C. Shi, Z. Zhang, H. Wang, and Y. Liu, "A review on mixture design methods for geopolymer concrete," *Composites Part B: Engineering*, vol. 178, Article ID 107490, 2019.

[21] T. T. Tran, T. M. Pham, and H. Hao, "Rectangular stress-block parameters for fly-ash and slag based geopolymer concrete," *Structures*, vol. 19, pp. 143–155, 2019.

[22] P. Zhang, Z. Gao, J. Wang, J. Guo, S. Hu, and Y. Ling, "Properties of fresh and hardened fly ash/slag based geopolymer concrete: a review," *Journal of Cleaner Production*, vol. 270, 2020.

[23] W. Prachasaree, S. Limkatanyu, A. Hawa, P. Sukontasukkul, and P. Chindaprasirt, "Manuscript title: development of strength prediction models for fly ash based geopolymer concrete," *Journal of Building Engineering*, vol. 32, Article ID 101704, 2020.

[24] H. Zhang, L. Li, P. K. Sarker et al., "Investigating various factors affecting the long-term compressive strength of heat-cured fly ash geopolymer concrete and the use of orthogonal experimental design method," *International Journal of Concrete Structures and Materials*, vol. 13, no. 1, 2019.

[25] D. Dao, H.-B. Ly, S. Trinh, T.-T. Le, and B. Pham, "Artificial intelligence approaches for prediction of compressive strength of geopolymer concrete," *Materials*, vol. 12, no. 6, p. 983, 2019.

[26] S. Luhar, S. Chaudhary, and I. Luhar, "Development of rubberized geopolymer concrete: strength and durability studies," *Construction and Building Materials*, vol. 204, pp. 740–753, 2019.

[27] Y. Wang, S. Hu, and Z. He, "Mechanical and fracture properties of fly ash geopolymer concrete addictive with calcium aluminate cement," *Materials*, vol. 12, no. 18, p. 2982, 2019.

[28] M. F. Javed, M. N. Amin, M. I. Shah et al., "Applications of gene expression programming and regression techniques for estimating compressive strength of bagasse ash based concrete," *Crystals*, vol. 10, no. 9, p. 737, 2020.

[29] M. F. Javed, F. Farooq, S. A. Memon et al., "New prediction model for the ultimate axial capacity of concrete-filled steel tubes: an evolutionary approach," *Crystals*, vol. 10, no. 9, p. 741, 2020.

[30] Y. Sun, G. Li, J. Zhang, and D. Qian, "Prediction of the strength of rubberized concrete by an evolved random forest model," *Advances in Civil Engineering*, vol. 2019, p. 1, 2019.

[31] Y. Sun, G. Li, N. Zhang, Q. Chang, J. Xu, and J. Zhang, "Development of ensemble learning models to evaluate the strength of coal-grout materials," *International Journal of Mining Science and Technology*, 2020.

[32] A. M. Abd and S. M. Abd, "Modelling the strength of lightweight foamed concrete using support vector machine (SVM)," *Case Studies in Construction Materials*, vol. 6, pp. 8–15, 2017.

[33] M. A. Getahun, S. M. Shitote, and Z. C. Abiero Gariy, "Artificial neural network based modelling approach for strength prediction of concrete incorporating agricultural and construction wastes," *Construction and Building Materials*, vol. 190, pp. 517–525, 2018.

[34] B. Vakhshouri and S. Nejadi, "Predicition of compressive strength in light-weight self-compacting concrete by ANFIS analytical model," *Archives of Civil Engineering*, vol. 61, no. 2, pp. 53–72, 2015.

[35] Q. Han, C. Gui, J. Xu, and G. Lacidogna, "A generalized method to predict the compressive strength of high-performance concrete by improved random forest algorithm," *Construction and Building Materials*, vol. 226, pp. 734–742, 2019.

[36] P. Samui, "Multivariate adaptive regression spline (Mars) for prediction of elastic modulus of jointed rock mass," *Geotechnical and Geological Engineering*, vol. 31, no. 1, pp. 249–253, 2013.

[37] M. Sonebi and A. Cevik, "Genetic programming based formulation for fresh and hardened properties of self-compacting concrete containing pulverised fuel ash," *Construction and Building Materials*, vol. 23, no. 7, pp. 2614–2622, 2009.

[38] F. Aslam, F. Farooq, M. N. Amin et al., "Applications of gene expression programming for estimating compressive strength of high-strength concrete," *Advances in Civil Engineering*, vol. 2020, Article ID 8850535, 23 pages, 2020.

[39] M. F. Iqbal, Q.-f. Liu, I. Azim et al., "Prediction of mechanical properties of green concrete incorporating waste foundry sand based on gene expression programming," *Journal of Hazardous Materials*, vol. 384, Article ID 121322, 2020.

[40] M. Jalal, Z. Grasley, C. Gurganus, and J. W. Bullard, "Experimental investigation and comparative machine-learning prediction of strength behavior of optimized recycled rubber concrete," *Construction and Building Materials*, vol. 256, Article ID 119478, 2020.

[41] J.-S. Chou and A.-D. Pham, "Enhanced artificial intelligence for ensemble approach to predicting high performance concrete compressive strength," *Construction and Building Materials*, vol. 49, pp. 554–563, 2013.

[42] H. Mashhadban, S. S. Kutanaei, and M. A. Sayarinejad, "Prediction and modeling of mechanical properties in fiber reinforced self-compacting concrete using particle swarm optimization algorithm and artificial neural network," *Construction and Building Materials*, vol. 119, pp. 277–287, 2016.

[43] H. Sebaaly, S. Varma, and J. W. Maina, "Optimizing asphalt mix design process using artificial neural network and genetic algorithm," *Construction and Building Materials*, vol. 168, pp. 660–670, 2018.

[44] R. Sudin and N. Swamy, "Bamboo and wood fibre cement composites for sustainable infrastructure regeneration," *Journal of Materials Science*, vol. 41, no. 21, pp. 6917–6924, 2006.

[45] 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.

[46] Z. Keshavarz and H. Torkian, "Application of ANN and ANFIS models in determining compressive strength of

concrete," *Journal of Soft Computing in Civil Engineering*, vol. 2, no. 1, pp. 62–70, 2018.

[47] J. Zhang, G. Ma, Y. Huang, J. Sun, F. Aslani, and B. Nener, "Modelling uniaxial compressive strength of lightweight self-compacting concrete using random forest regression," *Construction and Building Materials*, vol. 210, pp. 713–719, 2019.

[48] A. H. Gandomi, 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, no. 9, pp. 1132–1143, 2012.

[49] A. H. Gandomi, G. J. Yun, and A. H. Alavi, "An evolutionary approach for modeling of shear strength of RC deep beams," *Materials and Structures*, vol. 46, no. 12, pp. 2109–2119, 2013.

[50] C. Ferreira, "Gene Expression Programming: Mathematical Modeling by an Artificial Intelligence," *Springer*, Vol. 21, Berlin, Germany, 2006.

[51] L. Chen, C.-H. Kou, and S.-W. Ma, "Prediction of slump flow of high-performance concrete via parallel hyper-cubic gene-expression programming," *Engineering Applications of Artificial Intelligence*, vol. 34, pp. 66–74, 2014.

[52] S. B. Beheshti Aval, H. Ketabdari, and S. Asil Gharebaghi, "Estimating shear strength of short rectangular reinforced concrete columns using nonlinear regression and gene expression programming," *Structures*, vol. 12, pp. 13–23, 2017.

[53] I. F. Kara, "Empirical modeling of shear strength of steel fiber reinforced concrete beams by gene expression programming," *Neural Computing and Applications*, vol. 23, no. 3-4, pp. 823–834, 2013.

[54] E. Sadrossadat, B. Ghorbani, M. Hamooni, and M. H. Moradpoor Sheikhkanloo, "Numerical formulation of confined compressive strength and strain of circular reinforced concrete columns using gene expression programming approach," *Structural Concrete*, vol. 19, no. 3, pp. 783–794, 2018.

[55] F. Özcan, "Gene expression programming based formulations for splitting tensile strength of concrete," *Construction and Building Materials*, vol. 26, no. 1, pp. 404–410, 2012.

[56] A. Gholampour, A. H. Gandomi, and T. Ozbakkaloglu, "New formulations for mechanical properties of recycled aggregate concrete using gene expression programming," *Construction and Building Materials*, vol. 130, pp. 122–145, 2017.

[57] D. Behnia, K. Ahangari, A. Noorzad, and S. R. Moeinossadat, "Predicting crest settlement in concrete face rockfill dams using adaptive neuro-fuzzy inference system and gene expression programming intelligent methods," *Journal of Zhejiang University Science A*, vol. 14, no. 8, pp. 589–602, 2013.

[58] F. Farooq, M. Nasir Amin, K. Khan et al., "A comparative study of random forest and genetic engineering programming for the prediction of compressive strength of high strength concrete (HSC)," *Applied Sciences*, vol. 10, no. 20, pp. 7330–7418, 2020.

[59] A. Akbar, K. M. Liew, F. Farooq, and R. A. Khushnood, "Exploring mechanical performance of hybrid MWCNT and GNMP reinforced cementitious composites," *Construction and Building Materials*, Article ID 120721, 2020.

[60] S. Ishak, H.-S. Lee, J. K. Singh et al., "Materials performance of fly ash geopolymer concrete incorporating bamboo ash at elevated temperature," *Materials*, vol. 12, no. 20, p. 3404, 2019.

[61] M. Albitar, P. Visintin, M. S. Mohamed Ali, and M. Drechsler, "Assessing behaviour of fresh and hardened geopolymer concrete mixed with class-F fly ash," *KSCE Journal of Civil Engineering*, vol. 19, no. 5, pp. 1445–1455, 2015.

[62] I. S. Alkroosh and P. K. Sarker, "Prediction of the compressive strength of fly ash geopolymer concrete using gene expression programming," *Computers and Concrete*, vol. 24, pp. 295–302, 2019.

[63] D. Hardjito and B. V. Rangan, "Development and properties of low-calcium fly ash-based geopolymer concrete," *Research report GC1, Curtin Research Publications*, Curtin University of Technology, Perth, Australia, 2005.

[64] B. Joseph and G. Mathew, "Influence of aggregate content on the behavior of fly ash based geopolymer concrete," *Scientia Iranica*, vol. 19, no. 5, pp. 1188–1194, 2012.

[65] J. R. Koza and R. Poli, "Genetic programming," in *Search Methodologies Book*, pp. 127–164, Springer, Boston, MA, USA, 2005.

[66] M. Saridemir, "Genetic programming approach for prediction of compressive strength of concretes containing rice husk ash," *Construction and Building Materials*, vol. 24, pp. 1911–1919, 2010.

[67] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[68] V. Svetnik, A. Liaw, C. Tong, J. C. Culberson, R. P. Sheridan, and B. P. Feuston, "Random forest: a classification and regression tool for compound classification and QSAR modeling," *Journal of Chemical Information and Computer Sciences*, vol. 43, no. 6, pp. 1947–1958, 2003.

[69] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, "Predicting stock market index using fusion of machine learning techniques," *Expert Systems with Applications*, vol. 42, no. 4, pp. 2162–2172, 2015.

[70] H. Jiang, Y. Deng, H.-S. Chen et al., "Joint analysis of two microarray gene-expression data sets to select lung adenocarcinoma marker genes," *BMC Bioinformatics*, vol. 5, no. 1, p. 81, 2004.

[71] O. Rahmati, H. R. Pourghasemi, and A. M. Melesse, "Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: a case study at Mehran Region, Iran," *Catena*, vol. 137, pp. 360–372, 2016.

[72] B. Joseph and G. Mathew, "Behavior of geopolymer concrete exposed to elevated temperatures," *Ph.D. Dissertation.*, Cochin University of Science and Technology, Kochi, India, 2015.

[73] P. K. Sarker, R. Haque, and K. V. Ramgolam, "Fracture behaviour of heat cured fly ash based geopolymer concrete," *Materials & Design*, vol. 44, pp. 580–586, 2013.

[74] T. Long, X. S. Shi, Q. Y. Wang, and L. Li, "Mechanical properties and microstructure of fly ash based geopolymeric polymer recycled concrete," *Sichuan Daxue Xuebao (Gongcheng Kexue Ban)/Journal of Sichuan University (Engineering Science Edition)*, vol. 45, pp. 43–47, 2013.

[75] T. Sujatha, K. Kannapiran, and S. Nagan, "Strength assessment of heat cured geopolymer concrete slender column," *Asian Journal of Civil Engineering*, vol. 13, pp. 635–646, 2012.

[76] P. R. Vora and U. V. Dave, "Parametric studies on compressive strength of geopolymer concrete," *Procedia Engineering*, vol. 51, pp. 210–219, 2013.

[77] A. Wardhono, C. Gunasekara, D. W. Law, and S. Setunge, "Comparison of long term performance between alkali activated slag and fly ash geopolymer concretes," *Construction and Building Materials*, vol. 143, pp. 272–279, 2017.

[78] W. Lokuge, A. Wilson, C. Gunasekara, D. W. Law, and S. Setunge, "Design of fly ash geopolymer concrete mix proportions using Multivariate Adaptive Regression Spline

model," *Construction and Building Materials*, vol. 166, pp. 472–481, 2018.

[79] A. Mehta and R. Siddique, "Properties of low-calcium fly ash based geopolymer concrete incorporating OPC as partial replacement of fly ash," *Construction and Building Materials*, vol. 150, pp. 792–807, 2017.

[80] K. Ramujee and M. Potharaju, "Mechanical properties of geopolymer concrete composites," *Materials Today: Proceedings*, vol. 4, pp. 2937–2945, 2017.

[81] T. Sathanandam, P. O. Awoyera, V. Vijayan, and K. Sathishkumar, "Low carbon building: experimental insight on the use of fly ash and glass fibre for making geopolymer concrete," *Sustainable Environment Research*, vol. 27, no. 3, pp. 146–153, 2017.

[82] P. Nuaklong, V. Sata, and P. Chindaprasirt, "Influence of recycled aggregate on fly ash geopolymer concrete properties," *Journal of Cleaner Production*, vol. 112, pp. 2300–2307, 2016.

[83] M. Fareed Ahmed, M. Fadhil Nuruddin, and N. Shafiq, "Compressive strength and workability characteristics of low-calcium fly ash-based self-compacting geopolymer concrete," *World Academy of Science, Engineering and Technology*, vol. 74, pp. 8–14, 2011.

[84] A. Wongsa, Y. Zaetang, V. Sata, and P. Chindaprasirt, "Properties of lightweight fly ash geopolymer concrete containing bottom ash as aggregates," *Construction and Building Materials*, vol. 111, pp. 637–643, 2016.

[85] F. U. A. Shaikh, "Mechanical and durability properties of fly ash geopolymer concrete containing recycled coarse aggregates," *International Journal of Sustainable Built Environment*, vol. 5, no. 2, pp. 277–287, 2016.

[86] H. K. Shehab, A. S. Eisa, and A. M. Wahba, "Mechanical properties of fly ash based geopolymer concrete with full and partial cement replacement," *Construction and Building Materials*, vol. 126, pp. 560–565, 2016.

[87] A. A. Aliabdo, A. E. M. Abd Elmoaty, and H. A. Salem, "Effect of cement addition, solution resting time and curing characteristics on fly ash based geopolymer concrete performance," *Construction and Building Materials*, vol. 123, pp. 581–593, 2016.

[88] F. N. Okoye, J. Durgaprasad, and N. B. Singh, "Mechanical properties of alkali activated flyash/Kaolin based geopolymer concrete," *Construction and Building Materials*, vol. 98, pp. 685–691, 2015.

[89] N. Ganesan, R. Abraham, and S. Deepa Raj, "Durability characteristics of steel fibre reinforced geopolymer concrete," *Construction and Building Materials*, vol. 93, pp. 471–476, 2015.

[90] L. N. Assi, E. Deaver, M. K. Elbatanouny, and P. Ziehl, "Investigation of early compressive strength of fly ash-based geopolymer concrete," *Construction and Building Materials*, vol. 112, pp. 807–815, 2016.

[91] F. U. A. Shaikh and V. Vimonsatit, "Compressive strength of fly-ash-based geopolymer concrete at elevated temperatures," *Fire and Materials*, vol. 39, no. 2, pp. 174–188, 2015.

[92] S. V. Joshi and M. S. Kadu, "Role of alkaline activator in development of eco-friendly fly ash based geo polymer concrete," *International Journal of Environmental Science and Development*, vol. 3, no. 5, p. 417, 2012.

[93] G. Lavanya and J. Jegan, "Evaluation of relationship between split tensile strength and compressive strength for geopolymer concrete of varying grades and molarity," *International Journal of Applied Engineering Research*, vol. 10, pp. 35523–35529, 2015.

[94] P. S. Deb, P. Nath, and P. K. Sarker, "Properties of fly ash and slag blended geopolymer concrete cured at ambient temperature," in *Proceedings of the ISEC 2013-7th International Structural Engineering and Construction Conference*, pp. 571–576, New Developments in Structural Engineering and Construction, Honolulu, HI, USA, 2013.

[95] M. F. Nuruddin, S. Demie, and N. Shafiq, "Effect of mix composition on workability and compressive strength of self-compacting geopolymer concrete," *Canadian Journal of Civil Engineering*, vol. 38, no. 11, pp. 1196–1203, 2011.

[96] S. Ishak, H. S. Lee, J. K. Singh, M. A. M. Ariffin, N. H. A. S. Lim, and H. M. Yang, "Performance of fly ash geopolymer concrete incorporating bamboo ash at elevated temperature," *Materials*, vol. 12, no. 20, p. 3404, 2019.

[97] S. V. Patankar, Y. M. Ghugal, and S. S. Jamkar, "Mix design of fly ash based geopolymer concrete," in *Advances in Structural Engineering: Materials*, pp. 1619–1634, Springer India, New Delhi, India, 2015, ISBN 9788132221876.

[98] S. V. Patankar, S. S. Jamkar, and Y. M. Ghugal, "Effect of water-to-geopolymer Binder ratio on the production of fly ash based geopolymer concrete," *International Journal of Advanced Technology in Civil Engineering*, pp. 79–83, 2013.

[99] M. D. J. Sumajouw and B. V. Rangan, "Low-calcium fly ash-based geopolymer concrete: reinforced beams and columns," *Research report GC 3, Curtin Research Publications*, Curtin University of Technology, Perth, Australia, 2006.

[100] P. S. Deb, P. K. Sarker, and P. Nath, "Sulphate resistance of slag blended fly ash based geopolymer concrete," in *Proceedings of the 26th Biennial National Conference of the Concrete Institute of Australia, Gold Coast, Australia*, Concrete Institute of Australia, October 2013.

[101] B. Galvin and N. Lloyd, "Fly ash based geopolymer concrete with recycled concrete aggregate," in *Proceedings of the Concrete 2011 Conference. The Concrete Institute of Australia*, Perth, Australia, October 2011.

[102] A. Kusbiantoro, M. F. Nuruddin, N. Shafiq, and S. A. Qazi, "The effect of microwave incinerated rice husk ash on the compressive and bond strength of fly ash based geopolymer concrete," *Construction and Building Materials*, vol. 36, pp. 695–703, 2012.

[103] M. F. Nuruddin, S. A. Qazi, A. Kusbiantoro, and N. Shafiq, "Utilisation of waste material in geopolymeric concrete," *Proceedings of the Institution of Civil Engineers-Construction Materials*, vol. 164, no. 6, pp. 315–327, 2011.

[104] P. Nath and P. K. Sarker, "Flexural strength and elastic modulus of ambient-cured blended low-calcium fly ash geopolymer concrete," *Construction and Building Materials*, vol. 130, pp. 22–31, 2017.

[105] M. Olivia and H. Nikraz, "Properties of fly ash geopolymer concrete designed by Taguchi method," *Materials & Design*, vol. 36, pp. 191–198, 2012.

[106] A. J. Hamad, "Size and shape effect of specimen on the compressive strength of HPLWFC reinforced with glass fibres," *Journal of King Saud University - Engineering Sciences*, vol. 29, no. 4, pp. 373–380, 2017.

[107] J. R. del Viso, J. R. Carmona, and G. Ruiz, "Shape and size effects on the compressive strength of high-strength concrete," *Cement and Concrete Research*, vol. 38, no. 3, pp. 386–395, 2008.

[108] J. G. S. Van Jaarsveld, J. S. J. Van Deventer, and G. C. Lukey, "The effect of composition and temperature on the properties of fly ash- and kaolinite-based geopolymers," *Chemical Engineering Journal*, vol. 89, pp. 63–73, 2002.

[109] A. H. Gandomi and D. A. Roke, "Assessment of artificial neural network and genetic programming as predictive tools," *Advances in Engineering Software*, vol. 88, pp. 63–72, 2015.

[110] S. K. Babanajad, A. H. Gandomi, and A. H. Alavi, "New prediction models for concrete ultimate strength under true-triaxial stress states: an evolutionary approach," *Advances in Engineering Software*, vol. 110, pp. 55–68, 2017.

[111] A. H. Gandomi, A. H. Alavi, M. R. Mirzahosseini, and F. M. Nejad, "Nonlinear genetic-based models for prediction of flow number of asphalt mixtures," *Journal of Materials in Civil Engineering*, vol. 23, no. 3, pp. 248–263, 2011.

[112] A. H. Alavi, M. Ameri, A. H. Gandomi, and M. R. Mirzahosseini, "Formulation of flow number of asphalt mixes using a hybrid computational method," *Construction and Building Materials*, vol. 25, no. 3, pp. 1338–1355, 2011.

Hindawi
Advances in Civil Engineering
Volume 2020, Article ID 8863181, 11 pages
https://doi.org/10.1155/2020/8863181

Research Article

Predicting the Permeability of Pervious Concrete Based on the Beetle Antennae Search Algorithm and Random Forest Model

Jiandong Huang,[1,2] Tianhong Duan ,[1,2] Yi Zhang ,[3] Jiandong Liu,[1] Jia Zhang,[2] and Yawei Lei[4]

[1]*State Key Laboratory of Coal Resources and Safe Mining, China University of Mining and Technology, Xuzhou 221116, China*
[2]*School of Mines, China University of Mining and Technology, Xuzhou 221116, China*
[3]*Key Laboratory of Road and Traffic Engineering of the Ministry of Education, College of Transportation Engineering, Tongji University, 4800 Caoan Road, Shanghai 201804, China*
[4]*China Construction Second Engineering Bureau Ltd., Beijing 100160, China*

Correspondence should be addressed to Tianhong Duan; passionduan@cumt.edu.cn and Yi Zhang; zhangyi1990@tongji.edu.cn

Received 16 September 2020; Revised 30 October 2020; Accepted 6 December 2020; Published 29 December 2020

Academic Editor: Junfei Zhang

Pervious concrete is an environmentally friendly material that improves water permeability, skid resistance, and sound absorption characteristics. Permeability is the most important functional performance for the pervious concrete while limited studies have been conducted to predict permeability based on mix-design parameters. This study proposed a method to combine the beetle antennae search (BAS) and random forest (RF) algorithm to predict the permeability of pervious concrete. Based on the 36 samples designed in the laboratory and 4 key influencing variables, the permeability of pervious concrete can be obtained by varying mix-design parameters by RF. BAS algorithm was used to tune the hyperparameters of RF, which were then verified by the so-called 10-fold cross-validation. Furthermore, the model to combine the BAS and RF was validated by the correlation parameters. The results showed that the hyperparameters of RF can be tuned by the BAS efficiently; the BAS can combine the conventional RF algorithm to construct the evolved model to predict the permeability of pervious concrete; the cement/aggregate ratio was the most significant variable to determine the permeability, followed by the coarse aggregate proportions.

1. Introduction

Pervious concrete is similar to conventional concrete but designed without fine aggregates (i.e., sand) and has a porosity and median pore diameter in the range of 0.15–0.3 and 2–4 mm, respectively [1–4]. Pervious concrete is an environmentally friendly material that improves skid resistance and sound absorption characteristics and reduces the "heat island effect" [5–10]. Further, pervious concrete displays better water permeability characteristics due to connected pore structure through the fluid [2, 11].

Permeability is the most important functional performance for the pervious concrete, and it has been confirmed to be closely related to the pore structure, which determines the permeation rate per unit area. The permeability is traditionally characterized by the so-called permeability coefficient, and its value is typically between 0.1 and 2 cm/s [12, 13]. According to whether it is directly related to permeability performance of pervious concrete, pore-structure parameters can be divided into two categories: nonconnected related parameters and connected related parameters. Nonconnected related parameters include total porosity, pore diameter, and distribution, and connected related parameters include connected porosity and pore tortuosity [14–16]. The total porosity of pervious concrete can be defined as the ratio of the voids volume to the specimen volume, which mainly depends on the ratio of mortar to aggregate and the compactness degree of the concrete, usually between 15% and 25% [17]. Studies have shown that the total porosity of pervious concrete decreases with the increase of the ratio of mortar to aggregate [18]. Besides, as the pressure load or vibration load increases, the

skeleton of pervious concrete tends to be dense, and the total porosity gradually decreases [19]. Connected porosity is also called effective porosity, that is, pores that can effectively pass air and liquid. Cosic et al. used X-ray tomography technology to study the effects of aggregate type and size on the pore structure and found that the interconnected porosity is a function of aggregate size, accounting for about 50% to 70% of the total porosity [20]. Also, another study confirmed that the connected porosity decreases with the increase of the amount of mortar [13]. Kuang et al. defined the ratio of the effective length of the pore to the total length of the pervious concrete specimen as the pore tortuosity, which can more intuitively reflect the characteristics of the pore structure [21]. Zhong et al. defined the pore tortuosity as a function of the average pore diameter and the aggregate size and believed that the fluidity of the mortar and the aggregate size directly affects the pore tortuosity. The pore tortuosity increases with the decrease of aggregate size and the increase of mortar fluidity [22]. It can be confirmed from the previous research that the parameters affecting the permeability of pervious concrete have been studied, including aggregate size and mortar content [14, 17, 19]. However, from the author's knowledge, few studies can predict the water permeability based on these variable parameters. Although some studies can propose models for predicting permeability from a microscopic perspective (pore structure, the effective length of pores, etc.) [21, 22], these models often require the acquisition of concrete cross-section information firstly (e.g., CT scan), and it is difficult to predict from the perspective of mixture design. Therefore, systematical investigations are required to evaluate the permeability of pervious concrete in a way of more economic and efficient technology as per the permeability database including varying parameters.

Machine learning methods are gradually applied to the evaluation and prediction of the mechanical performance of cement materials [22–32]. The punching shear capacity of steel fiber reinforced concrete slabs was predicted by using the sequential piecewise multiple linear regression and artificial neural network [33]. Jamal et al evaluated the possibility to predict the strength of recycled aggregate concrete using machine learning methods, multiple linear regression, and adaptive neurofuzzy inference system [34]. The same method was also used by Khademi et al. in a follow-up study for the prediction of compressive strength after 28 days [35]. In the above-mentioned methods, the agreement between the experimental results and the predicted results indicates the feasibility of the machine learning algorithm for strength prediction of cement materials. However, within the scope of the author's knowledge, the limited literature can provide accurate and widely used machine learning algorithms specifically for the functional performance (permeability, workability, etc.) of cement materials.

Furthermore, the above machine learning methods have been used for the specific predictions in cement-based materials, but limitations still existed in these studies, such as uncertain structure, time-consuming, and low efficiency. Consequently, more efficient and simple machine learning models need to be proposed and used to predict the permeability of pervious concrete. In recent years, due to the good performance of the random forest (RF) method in nonlinear regression and classification, it has been used to predict the mechanical and functional properties of concrete. Specifically, the coefficients of thermal expansion and other properties of concrete were confirmed to be accurately predicted using the RF method [36]. The same method was also employed to predict and evaluate the compressive strength of high-performance concrete. However, no corresponding studies were reported to use RF to predict the permeability of pervious concrete so far. Besides, as far as the RF model employed in the previous studies, the hyperparameters were still required to be optimized to arrive at their optimized predictive ability [37].

2. Research Objective and Overview

The present study aims to propose a robust machine learning technique to be used as a tool to predict the permeability of pervious concrete. An efficient global optimization algorithm (called the beetle antennae search, BAS) proposed by Jiang et al. was adopted in this study to obtain the optimized parameters of RF [37]. In this way, the random forest (RF) and beetle antennae search (BAS) algorithms were combined to build a robust machine learning technique, named as BRF method. To plant the database applied to the proposed BRF method, varying mixes of pervious concrete were designed considering four parameters (aggregate proportion %: 9.5~13.2 mm; aggregate proportion: 4.75~9.5 mm; aggregate proportion %: 2.36~4.75 mm; cement-aggregate ratio) that have significant effects on the permeability coefficient according to the investigation based on known literature. Using the obtained database of the permeability of the pervious concrete, the training subset and the testing subset were developed for machine learning, and finally, the prediction of permeability can be realized. The above research process can be overviewed in Figure 1.

3. Methodology

3.1. BRF Model. BRF model combines BAS and RF, where RF is used to determine the nonlinear relationship of the dataset, and BAS is applied to adjust the hyperparameters of RF. The detailed introduction for the BRF model is described as follows.

3.1.1. Random Forest (RF) Model. The RF model is a modeling method that combines multiple independent classification trees. The algorithm can improve the prediction accuracy on the premise that the calculation is not significantly increased. The principle framework of the RF model is presented in Figure 2.

RF is a classifying method that uses a collection of classification trees, and every classification tree is constructed by using guided samples of data. For the tree construction, variables are randomly selected in each partition as the candidate variable set.

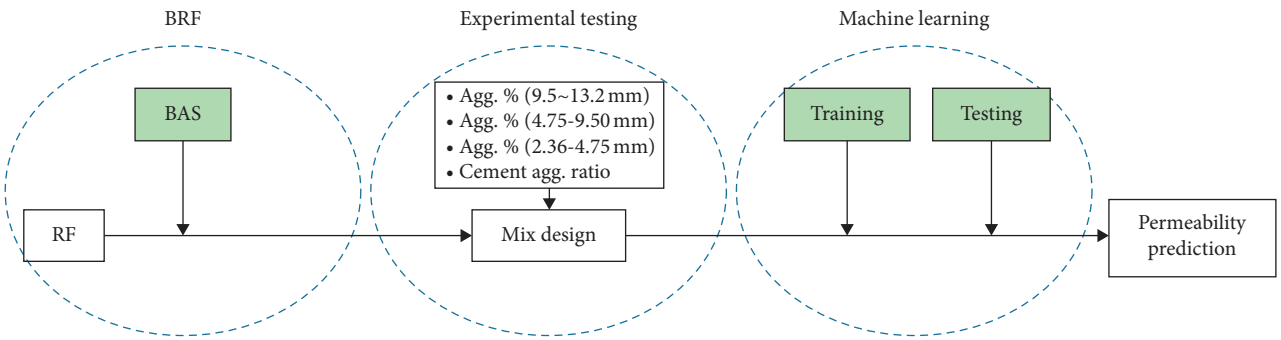


Figure 1: Research overview.

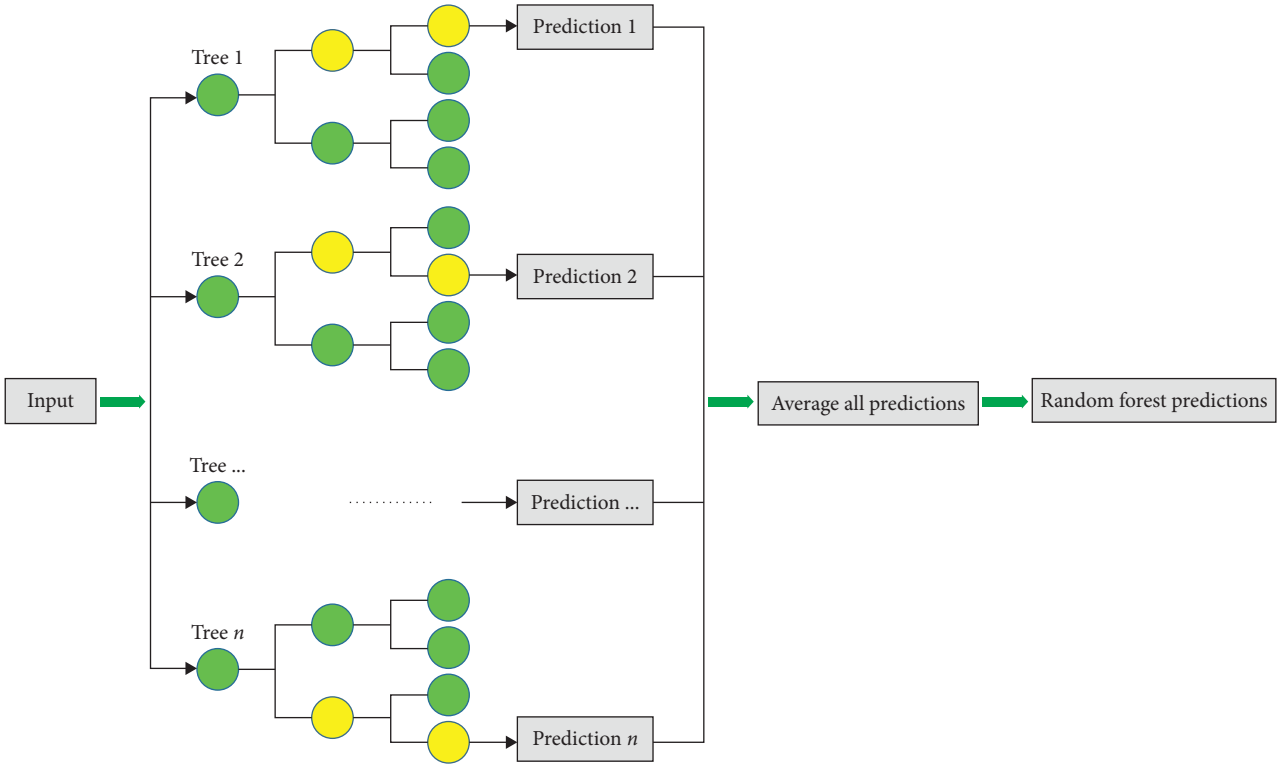


Figure 2: Principle of the RF model.

3.1.2. Beetle Antennae Search (BAS). The design of the BAS algorithm is inspired by the behavior of the beetle when looking for a mate. Similar to intelligent optimization algorithms such as genetic algorithm, particle swarm algorithm, and simulated annealing, BAS does not need to know the specific form of the function and does not need gradient information to achieve efficient optimization. Compared with the particle swarm algorithm, BAS only requires one individual, that is, a long beetle, which greatly reduces the amount of calculation. It simulates the behavior of beetles, which can use the two antennae to randomly explore nearby areas and transform them into a higher concentration of odor. The performance of the BAS algorithm has been evaluated in various applications [39, 40]. Figure 3 gives the work chart of the BAS algorithm.

4. Experimental Testing and Model Validation

4.1. Experimental Testing. The purpose of the laboratory testing is to evaluate the influence of different variables in the design of pervious concrete on the permeability coefficient and then provide enough data to be assembled as the training set and testing set. The raw materials, mixture design, sample preparation, and permeability test methods used in the laboratory testing are introduced as follows.

4.1.1. Raw Materials. Cement and aggregate were used as the raw materials to prepare the pervious concrete specimen. The ordinary Portland cement grade 42.5 was selected as the cemented material. Table 1 gives the physical properties of the cement.

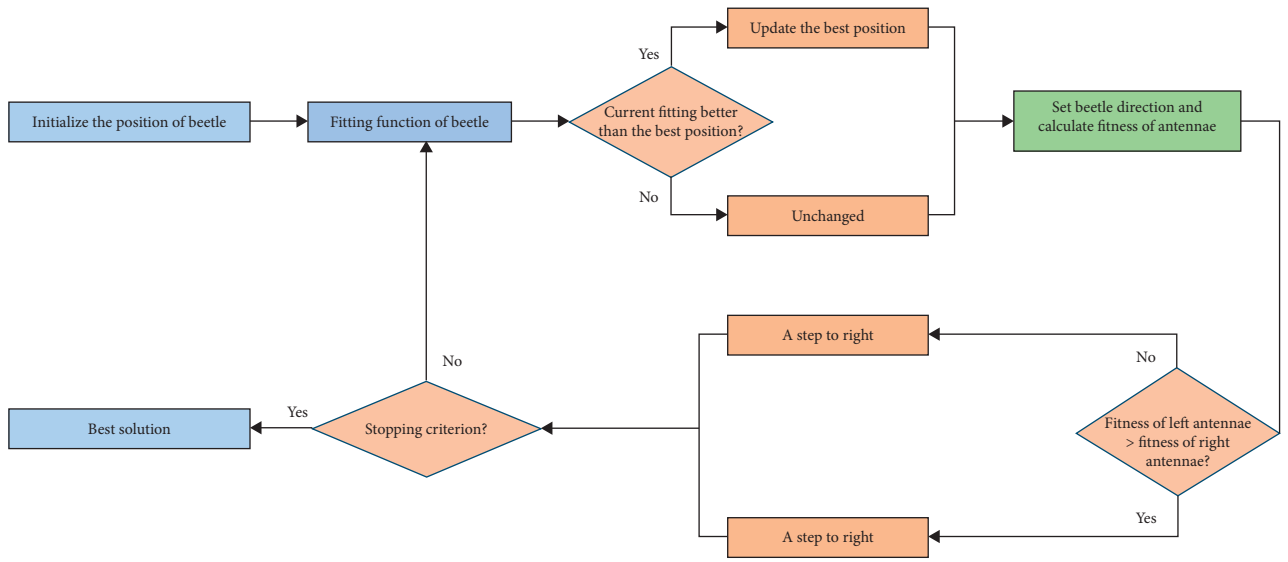


Figure 3: Work chart of the BAS algorithm.

Table 1: Physical properties of the cement.

Coagulation time (min)		Flexural strength (MPa)		Compressive strength (MPa)	
Initial solidification	Finial solidification	3d	28d	3d	28d
165	330	5.1	7.1	20.9	44.3

Table 2: Physical properties of coarse aggregate.

Aggregates	Size (mm)	Apparent density (kg/m^3)	Bulk density (kg/m^3)
G_1	2.36–4.75	2787	1406
G_2	4.75–9.50	2844	1513
G_3	9.50–13.2	2840	1537

According to the standard CJJ/T135-200, four different sizes of limestone gravel aggregates (G_1, G_2, and G_3) were adopted in the present study. The physical indexes are given in Table 2. To ensure that the performance of the prepared pervious concretes to meet the requirements, the aggregate used should be placed in a constant temperature oven in advance to keep it clean and dry. A liquid modifier for the pervious concrete provided by a local supplier in Jiangsu Province was used.

4.1.2. Mix Design and Sample Preparation. To evaluate the effects of different variables on the permeability coefficient, the designed mix was divided into 69 groups according to different aggregate proportions and cement-aggregate ratio (C/A), as shown in Table 3.

Generally, the reasonable water-cement ratio (W/C) of pervious concrete ranges from 0.29 to 0.33, and 0.3 was determined in the present study for the mixing [7]. Three possible C/A values (0.22, 0.24, and 0.26) were chosen to determine their effects on water permeability, based on the results of previous studies conducted [18, 39, 40]. After stirring and weighing, the mixture of pervious concrete was poured into a cylindrical mold with a diameter of 100 mm and a thickness of 50 mm to obtain samples for permeability testing. Before demolding, cover the sample and mold with damp geotextile in the laboratory at a temperature of 20°C for 24 h. Then, the samples were placed in a standard curing room with a temperature of 20°C and humidity of 95% for 28 days. It should be noted that three replicate samples were prepared for each mixture to reduce the deviation.

4.1.3. Experimental Methods. As far as the permeability testing methods concerned, the falling-head and constant-head method were typically used in the previous studies [41, 42]. For materials with poor permeability (permeability coefficient $< 10^{-3}$ cm/s) such as cohesive soil and fine-grained soil, permeability is typically measured by the falling-head method since the flow is too small to be measured. However, for pervious concrete, its permeability coefficient is relatively high (>0.35 cm/s) and the constant-head method is more suitable to measure the permeability coefficient. Therefore, the constant-head method was selected for the determination of the permeability coefficient in the study. The permeability coefficient was determined based on Darcy's law, where the amount of water passing through the concrete per unit time is proportional to the surface area and inversely proportional to the length of the permeable path, as given in equation (1):

$$k_t = \frac{Q \cdot D}{A \cdot H \cdot \Delta t}, \quad (1)$$

where k_t (mm/s) represents the coefficient of permeability at the given temperature t (°C), Q (mm^3) is the amount of water flowing through concrete in Δt (s) time; D (mm) and A (mm^2) are the thickness and area of the pervious concrete

Table 3: Mix proportions.

Mixes	Aggregate proportion			Cement-aggregate ratio (C/A)	Water-cement ratio
	G_1	G_2	G_3		
M1	1	0	0	0.22	
M2	1	0	0	0.24	
M3	1	0	0	0.26	
M4	0	1	0	0.22	
M5	0	1	0	0.24	
M6	0	1	0	0.26	
M7	0	0	1	0.22	
M8	0	0	1	0.24	
M9	0	0	1	0.26	
M10	0.1	0.9	0	0.22	
M11	0.1	0.9	0	0.24	
M12	0.1	0.9	0	0.26	
M13	0.2	0.8	0	0.22	
M14	0.2	0.8	0	0.24	
M15	0.2	0.8	0	0.26	
M16	0	0.9	0.1	0.22	
M17	0	0.9	0.1	0.24	
M18	0	0.9	0.1	0.26	
M19	0	0.8	0.2	0.22	
M20	0	0.8	0.2	0.24	
M21	0	0.8	0.2	0.26	
M22	0.1	0.8	0.1	0.22	
M23	0.1	0.8	0.1	0.24	
M24	0.1	0.8	0.1	0.26	
M25	0.1	0.7	0.2	0.22	
M26	0.1	0.7	0.2	0.24	
M27	0.1	0.7	0.2	0.26	
M28	0.1	0.6	0.3	0.22	
M29	0.1	0.6	0.3	0.24	
M30	0.1	0.6	0.3	0.26	
M31	0.1	0.5	0.4	0.22	
M32	0.1	0.5	0.4	0.24	
M33	0.1	0.5	0.4	0.26	
M34	0.1	0.4	0.5	0.22	
M35	0.1	0.4	0.5	0.24	0.3
M36	0.1	0.4	0.5	0.26	
M37	0.15	0.7	0.15	0.22	
M38	0.15	0.7	0.15	0.24	
M39	0.15	0.7	0.15	0.26	
M40	0.15	0.6	0.25	0.22	
M41	0.15	0.6	0.25	0.24	
M42	0.15	0.6	0.25	0.26	
M43	0.15	0.5	0.35	0.22	
M44	0.15	0.5	0.35	0.24	
M45	0.15	0.5	0.35	0.26	
M46	0.15	0.4	0.45	0.22	
M47	0.15	0.4	0.45	0.24	
M48	0.15	0.4	0.45	0.26	
M49	0.2	0.7	0.1	0.22	
M50	0.2	0.7	0.1	0.24	
M51	0.2	0.7	0.1	0.26	
M52	0.2	0.6	0.2	0.22	
M53	0.2	0.6	0.2	0.24	
M54	0.2	0.6	0.2	0.26	
M55	0.2	0.5	0.3	0.22	
M56	0.2	0.5	0.3	0.24	
M57	0.2	0.5	0.3	0.26	
M58	0.2	0.4	0.4	0.22	
M59	0.2	0.4	0.4	0.24	
M60	0.2	0.4	0.4	0.26	
M61	0.3	0.6	0.1	0.22	
M62	0.3	0.6	0.1	0.24	
M63	0.3	0.6	0.1	0.26	
M64	0.3	0.5	0.2	0.22	
M65	0.3	0.5	0.2	0.24	
M66	0.3	0.5	0.2	0.26	
M67	0.3	0.4	0.3	0.22	
M68	0.3	0.4	0.3	0.24	
M69	0.3	0.4	0.3	0.26	

specimen, respectively; H (mm) represents the hydraulic head difference.

The equipment used in the permeability tests was self-made in the laboratory, and the schematic diagram is presented in Figure 4. During the test, water was injected above the steel mold, then flooded the pervious concrete specimen below, and entered the tank, while the excess water overflowed from the outlet pipe. When the flow was too large, water can flow out through the vent pipe above the steel mold. It should be noted that the water level inside the steel mold should be kept constant and then the head difference can be recorded to calculate the permeability coefficient of the specimen.

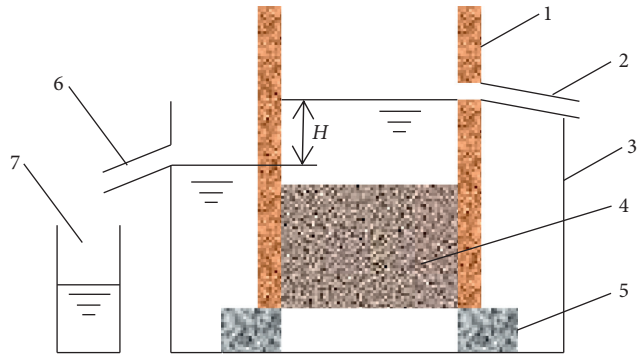


Figure 4: Schematic diagram of the permeable test device.*Notes: 1: steel mold; 2: vent pipe; 3: water tank; 4: concrete specimen; 5: pad; 6: outlet pipe; 7: measuring cylinder.

4.2. Model Validation

4.2.1. Methods for Model Validation. In the present study, the evolved random forest model was trained for the 70% dataset while the remaining 30% was for the testing dataset. It should be noted that all datasets should be split randomly during the training and validating process. Besides, two parameters (correlation, R; root-mean-square error, E_{RMS}) that were widely used in the previous studies were selected to assess the predictive ability of the dataset. The two parameters can be given as follows:

$$R = \frac{\sum_{i=1}^{N}\left(y_i^* - \overline{y^*}\right)\left(y_i - \overline{y}\right)}{\sqrt{\sum_{i=1}^{N}\left(y_i^* - \overline{y^*}\right)^2} \cdot \sqrt{\sum_{i=1}^{N}\left(y^i - \overline{y}\right)^2}},$$
$$E_{\mathrm{RMS}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i^* - y_i\right)^2}, \quad (2)$$

where N represents the collected numbers of the dataset; y_i and y_i^* represent actual values and predicted values, respectively; $\overline{y}$ and $\overline{y}^*$ represent the mean of the actual values and the predicted values. In addition, in order to minimize the deviation, the so-called 10-fold cross-validation method was adopted in the present study [43]. Under such a validation system, the samples as training dataset are divided into 10 subsets, one of these 10 subsets is applied to verify the predicted results of the BRF method, and the remaining 9 subsets are applied for training. Such a process should be repeated 10 times as described above.

4.2.2. Procedures of Hyperparameter Tuning. In order to obtain the optimized RF structure, hyperparameter tuning should be conducted. In the present study, the BAS algorithm was applied to tune two important parameters (the number of the trees, named as tree_num; the minimum number of samples required at a leaf node in RF, named as min_sample_leaf). Through the 10-fold cross-validation method introduced above, the 9 subsets as the training set were applied to search for the idealized hyperparameters of RF. This process should be performed 50 times by BAS [37]. For the validation dataset, the smallest E_{RMS} was selected after 50 iterations and in this fold, it can represent the optimized RF model. Consequently, the optimized RF model and the corresponding optimized hyperparameters (tree_num and min_sample_leaf) can be selected after 10 times. Due to the possibility of overfitting, the performance of the RF model should be verified by evaluating the testing set. Figure 5 summarizes the flowchart of the hyperparameters for RF tuning by BAS during training and testing.

4.2.3. Dataset Description. The dataset of pervious concrete from the experimental testing is applied to establish and validate the proposed BRF model for permeability prediction. A total of 36 mixes were used for the verification and prediction of the proposed model. It should be noted that the test results of the water permeability of each mixture are derived from the average of three parallel samples. Four key influencing variables (aggregate proportion of G_1 and G_1%; the aggregate proportion of G_2 and G_2%; the aggregate proportion of G_3 and G_3%; cement-aggregate ratio, C/A) were determined in the present study, as given in Table 3. These four variables have been confirmed in previous studies to have significant effects on the water permeability of pervious concrete and its clogging behavior [44–46]. Table 3 gives the influencing variables and their values in the design of pervious concrete, which were used to construct the dataset.

The main goal is to predict the permeability coefficient of pervious concrete, which is determined by the mix-design parameters. The relative importance of the mix-design parameters used for the input variables needs further analysis. It should be noted that the collected dataset should be normalized to [0, 1] in order to improve the efficiency of the proposed model. According to the proportion of the dataset, 25 mixes (70%) were randomly selected as the training dataset, and the other 11 mixes (30%) were used as the testing dataset (Figure 5).

5. Results and Discussion

5.1. Results of Hyperparameter Tuning. In order to obtain the optimized RF structure, the hyperparameters were adjusted on the testing set based on E_{RMS} obtained from the 10-fold

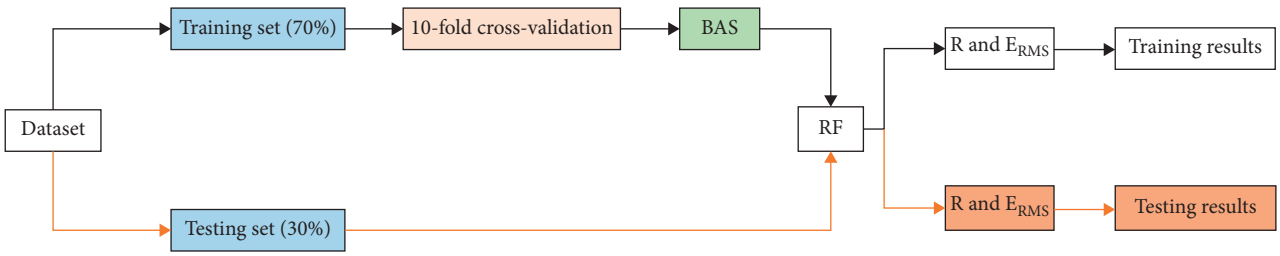


Figure 5: Flowchart of the hyperparameters for RF tuning by BAS during training and testing.

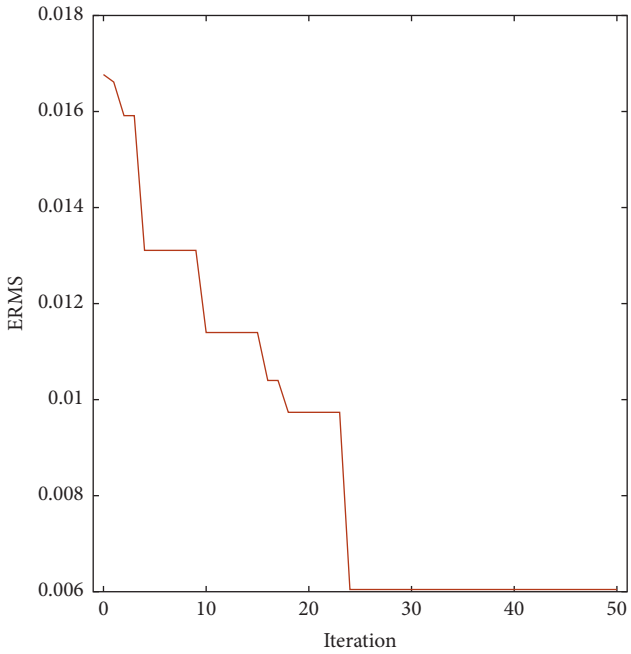


Figure 6: Relationship between the iteration and E_{RMS}.

cross-validation. Figure 6 gives the relationship between E_{RMS} and iterations during BAS tuning, which was performed 50 times in the present study.

It can be seen that E_{RMS} was greatly reduced with the increase of iteration, indicating that BAS can effectively tune the RF structure. In addition, E_{RMS} converged and reached the minimum value after 25 iterations, indicating that the optimized RF model was obtained in this fold. During the 10-fold cross-validation process, the optimized RF model in the whole calculation was determined after 10-fold and the corresponding optimized hyperparameters can be obtained. It should be noted that the prediction results of the RF should be verified through the way of evaluating the testing set. Table 4 gives the final hyperparameters of RF.

5.2. Assessing the Established Model. Figure 7 gives the comparison of the predicted permeability of the pervious concrete by the proposed method and the actual one in datasets.

Good agreement can be seen between the predicted permeability of pervious concrete and the actual

Table 4: Final hyperparameters of RF.

Parameters	Initial	Results	Empirical scope
tree_num	6	9	[1, 10]
min_sample_leaf	6	1	[1, 10]

permeability, indicating that the proposed method can well establish the nonlinear relationship between the permeability of pervious concrete and the input variables.

Furthermore, the statistical parameters of these comparisons for training and testing datasets were obtained, as shown in Table 5. The low E_{RMS} values of 0.0059 and 0.0131 can be observed for the training and testing dataset, respectively. Also, the high R values for the training set and test set were 0.9258 and 0.9208, respectively. All the above results indicated that the proposed RF model has no overfitting

5.3. Variable Importance of Pervious Concrete. Figure 8 gives the relative importance of the 4 design parameters, which were used as the input variables in the machine learning process.

Obviously, C/A was the most important design variable to determine the permeability of pervious concrete, since the highest importance score of 1.7762 can be observed. The mechanism of the effect of C/A on water permeability is that as C/A increases, the compaction resistance provided by the reduced aggregates decreases, resulting in a decrease in the volume of intercrystalline voids. This leads to lower porosity and permeability. The results of this study indicated that the effect mechanism of C/A on permeability exceeded the role of the aggregate proportion in determining permeability. Contemporaneously, the results were consistent with the findings noted by Zhang et al., Chandrappa et al., and Wang et al. [18, 47, 48]. As the largest aggregate used in the testing, the proportion of $G3$ (9.50 mm to 13.2 mm) also played an important role in determining the permeability of pervious concrete. This result was in line with the permeability properties of pervious concrete developed in the previous studies to understand the effects of aggregate sizes [44, 49, 50]. Also, it can be observed that the importance score of the $G1$ (2.36 mm–4.75 mm) aggregate is 0.8440, which ranks third among all variables by virtue of this score, indicating that it was more sensitive than the $G2$ (4.75 mm–9.50 mm) aggregate in determining the permeability of pervious

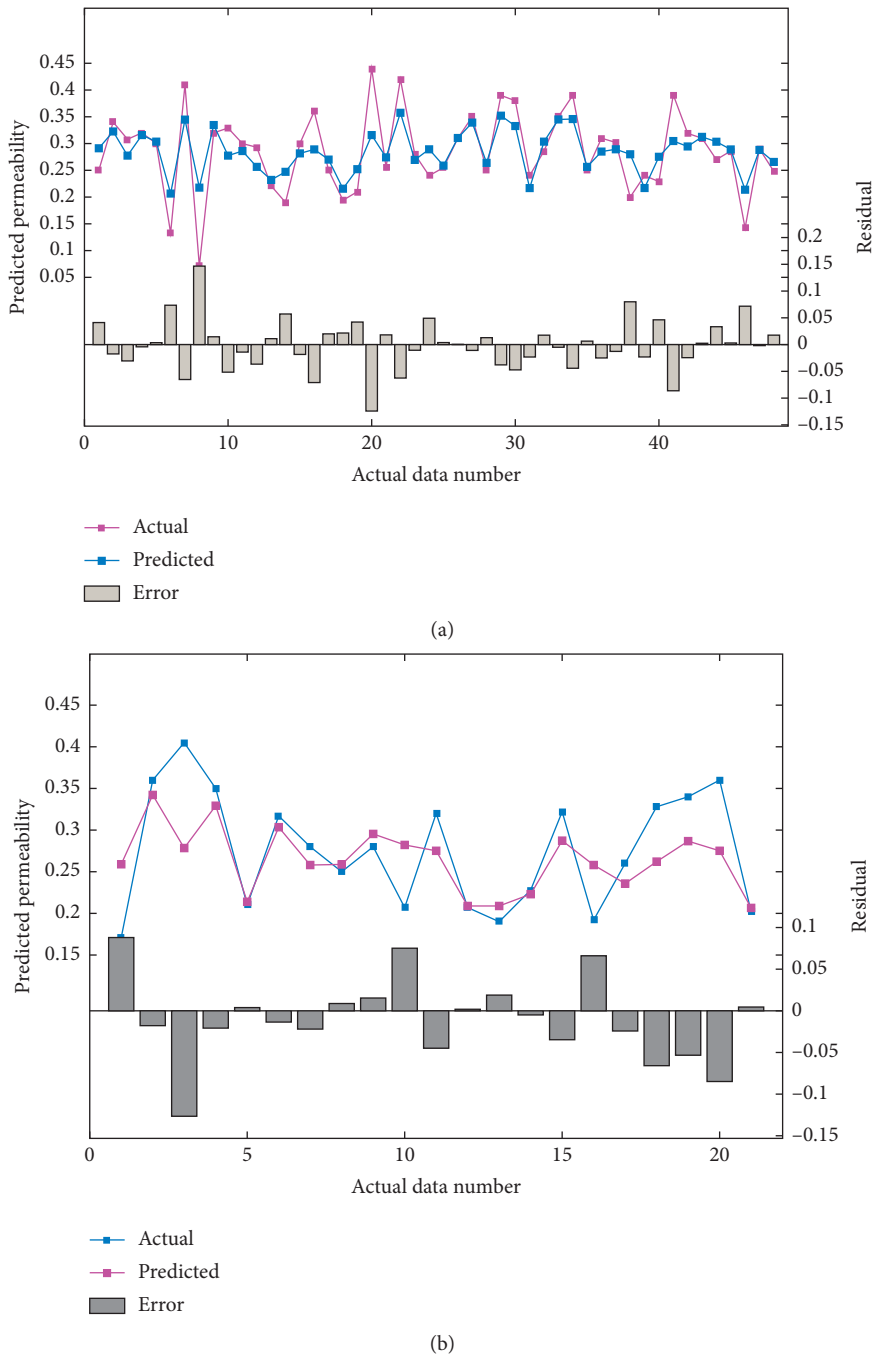


Figure 7: Comparison of permeability. (a) Training dataset. (b) Testing dataset.

Table 5: Statistical parameters of actual and predicted permeability in the datasets.

Datasets	E_{RMS}	R
Training dataset	0.0059	0.9258
Testing dataset	0.0131	0.9208

concrete. Therefore, in the future tests of pervious concrete, more C/A and $G3$ aggregate proportion combinations should be selected to optimize the target permeability. The results obtained can effectively guide the design of pervious concrete and select appropriate parameters to optimize C/A and aggregate gradation.

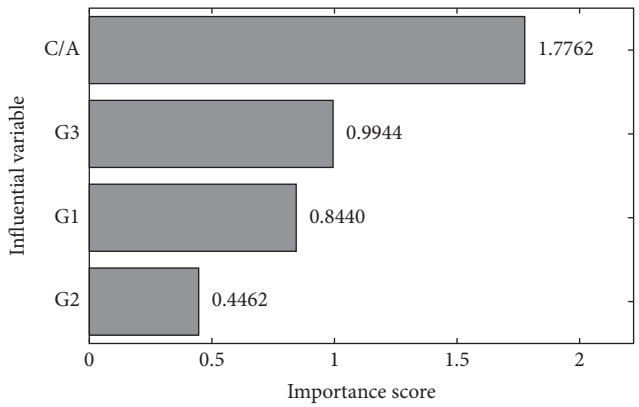


Figure 8: Importance score of the four design parameters.

6. Conclusions

The present study aims to propose a method to combine the beetle antennae search (BAS) and random forest (RF) algorithm to predict the permeability of pervious concrete. Based on the 36 samples designed in the laboratory and 4 key influencing variables, the permeability of pervious concrete can be determined by the independent variables under the RF model. The BAS algorithm was used to tune the hyperparameters of RF and the results were verified by 10-fold cross-validation. The prediction results of the optimized BRF were evaluated through R and E_{RMS}. The importance of the variables that determine permeability is also revealed and discussed. The following are the conclusions drawn from the above research process:

(i) The BAS algorithm is effective for adjusting the hyperparameters of RF and can be applied in evolved RF to construct predictive models; it has higher reliability and effectiveness than random hyperparameter selection.

(ii) The proposed RF model can accurately predict the permeability of pervious concrete, which can guide the functional designing for pervious concrete; for the testing set, R and E_{RMS} were 0.9223 and 0.0123, respectively, indicating that the proposed RF model showed good predictive ability on the collected dataset.

(iii) The C/A (cement/aggregate ratio) can be considered as the most significant variable for determining the permeability of pervious concrete, followed by the coarse aggregate proportions. Among them, the proportion of $G3$ and $G1$ aggregates is considered to be the most significant variable affecting the permeability for pervious concrete. However, the proportion of $G2$ aggregate has an almost negligible influence on the permeability. The results obtained can effectively guide the design of pervious concrete and select appropriate parameters to optimize C/A and aggregate gradation.

It should be noted that the results obtained in this study were limited by the number of samples. If more datasets and more variables are considered, the predicted permeability closer to the actual ones can be obtained. In the future, more samples with different combinations will be designed, mixed, and tested to obtain larger data sets for analysis through machine learning methods to more widely and effectively apply pervious concrete in the field of green construction.

Data Availability

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

Conflicts of Interest

There are no conflicts of interest.

Authors' Contributions

Jiandong Huang and Yi Zhang contributed to conceptualization; Jiandong Liu and Jiandong Huang contributed to methodology; Jiandong Huang, Yi Zhang, and Lin Wang wrote the original draft; Tianhong Duan and Yawei Lei reviewed and edited the article; Tianhong Duan supervised the study. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

The authors sincerely acknowledge the support from Independent Research Projects of State Key Laboratory of Coal Resources and Safe Mining, China University of Mining and Technology (SKLCRSM18X013).

References

[1] L. Haselbach, V. F. P. Dutra, P. Schwetz, and L. C. P. da Silva Filho, "Laboratory evaluations of long-term hydraulic performance and maintenance requirements for pervious concrete mixes: A case study in Southern Brazil," in *Proceedings of the International Conference on Transportation and Development*, vol. 2016, pp. 309–317, Houston, TX, USA, June 2016.

[2] J. Sanslone, X. Kuang, and V. Ranieri, "Permeable pavement as a hydraulic and filtration interface for urban drainage," *Journal of Irrigation and Drainage Engineering*, vol. 134, pp. 666–674, 2008.

[3] M. S. Sumanasooriya and N. Neithalath, "Stereology-and morphology-based pore structure descriptors of enhanced porosity (pervious) concretes," *ACI Materials Journal*, vol. 106, 2009.

[4] J. Huang, R. Alyousef, M. Suhatril, S. Baharom et al., "Influence of porosity and cement grade on concrete mechanical properties," *Advances in Concrete Construction*, vol. 10, no. 5, pp. 393–402, 2011.

[5] L. Haselbach, M. Boyer, J. T. Kevern, and V. R. Schaefer, "Cyclic heat island impacts on traditional versus pervious concrete pavement systems," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2240, no. 1, pp. 107–115, 2011.

[6] A. Mohajerani, J. Bakaric, and T. Jeffrey-Bailey, "The urban heat island effect, its causes, and mitigation, with reference to

the thermal properties of asphalt concrete," *Journal of Environmental Management*, vol. 197, pp. 522–538, 2017.

[7] Y. Qin, H. Yang, Z. Deng, and J. He, "Water permeability of pervious concrete is dependent on the applied pressure and testing methods," *Advances in Materials Science and Engineering*, vol. 2015, Article ID 404136, 6 pages, 2015.

[8] Y. Zhang, H. Li, A. Abdelhady, and H. Du, "Laboratorial investigation on sound absorption property of porous concrete with different mixtures," *Construction and Building Materials*, vol. 259, p. 120414, 2020.

[9] Y. Zhang, H. Li, A. Abdelhady, and J. Yang, "Effect of different factors on sound absorption property of porous concrete," *Transportation Research Part D: Transport and Environment*, vol. 87, p. 102532, 2020.

[10] J. Huang, J. Zhang, J. Ren, and H. Chen, "Anti-rutting performance of the damping asphalt mixtures (DAMs) made with a high content of asphalt rubber (AR)," *Construction and Building Materials*, vol. 271, Article ID 121878, 2020 In press.

[11] P. D. Tennis, M. L. Leming, and D. J. Akers, *Pervious Concrete Pavements*, Portland Cement Association Skokie, Skokie, IL, USA, 2004.

[12] P. Chindaprasirt, S. Hatanaka, T. Chareerat, N. Mishima, and Y. Yuasa, "Cement paste characteristics and porous concrete properties," *Construction and Building Materials*, vol. 22, no. 5, pp. 894–901, 2008.

[13] M. S. Sumanasooriya and N. Neithalath, "Pore structure features of pervious concretes proportioned for desired porosities and their performance prediction," *Cement and Concrete Composites*, vol. 33, no. 8, pp. 778–787, 2011.

[14] A. K. Chandrappa and K. P. Biligiri, "Pervious concrete as a sustainable pavement material - research findings and future prospects: a state-of-the-art review," *Construction and Building Materials*, vol. 111, pp. 262–274, 2016.

[15] J. Huang and Y. Sun, "Effect of modifiers on the rutting, moisture-induced damage, and workability properties of hot mix asphalt mixtures," *Applied Sciences*, vol. 10, no. 20, p. 7145, 2020.

[16] J. Huang and Y. Sun, "Viscoelastic analysis of the damping asphalt mixtures (dams) made with a high content of asphalt rubber (ar)," *Advances in Civil Engineering*, vol. 2020, Article ID 8826926, 12 pages, 2020.

[17] F. Montes, S. Valavala, and L. M. Haselbach, "A new test method for porosity measurements of portland cement pervious concrete," *Journal of ASTM International*, vol. 2, pp. 1–13, 2005.

[18] H. Wang, H. Li, X. Liang, H. Zhou, N. Xie, and Z. Dai, "Investigation on the mechanical properties and environmental impacts of pervious concrete containing fly ash based on the cement-aggregate ratio," *Construction and Building Materials*, vol. 202, pp. 387–395, 2019.

[19] A. Torres, J. Hu, and A. Ramos, "The effect of the cementitious paste thickness on the performance of pervious concrete," *Construction and Building Materials*, vol. 95, pp. 850–859, 2015.

[20] K. Cosic, L. Korat, V. Ducman, and I. Netinger, "Influence of aggregate type and size on properties of pervious concrete," *Construction and Building Materials*, vol. 78, pp. 69–76, 2015.

[21] X. Kuang, J. Sansalone, G. Ying, and V. Ranieri, "Pore-structure models of hydraulic conductivity for permeable pavement," *Journal of Hydrology*, vol. 399, no. 3-4, pp. 148–157, 2011.

[22] R. Zhong, M. Xu, R. Vieira Netto, and K. Wille, "Influence of pore tortuosity on hydraulic conductivity of pervious concrete: characterization and modeling," *Construction and Building Materials*, vol. 125, pp. 1158–1168, 2016.

[23] K. Xie, Y. Du, and C. Sun, "Application of the mind-evolution-based machine learning in mixture-ratio calculation of raw materials cement," in *Proceedings of the 3rd World Congress on Intelligent Control and Automation (Cat. No. 00EX393)*, pp. 132–134, IEEE, Hefei, China, July 2000.

[24] G. Ozcan, Y. Kocak, and E. Gulbandilar, "Estimation of compressive strength of bfs and wtrp blended cement mortars with machine learning models," *Computers and Concrete*, vol. 19, no. 3, pp. 275–282, 2017.

[25] A. Shadravan, M. Tarrahi, and M. Amani, "intelligent cement design: utilizing machine learning algorithms to assure effective long-term well integrity," in *Proceedings of the Carbon Management Technology Conference*, Sugarland, TX, USA, November 2015.

[26] A. Shadravan, M. Tarrahi, and M. Amani, "Intelligent tool to design fracturing, drilling, spacer and cement slurry fluids using machine learning algorithms," in *Proceedings of the SPE Kuwait Oil and Gas Show and Conference*, Society of Petroleum Engineers, Mishref, Kuwait, October 2015.

[27] T. Oey, S. Jones, J. W. Bullard, and G. Sant, "Machine learning can predict setting behavior and strength evolution of hydrating cement systems," *Journal of the American Ceramic Society*, vol. 103, no. 1, pp. 480–490, 2020.

[28] A. Menon, C. M. Childs, B. Poczos, N. R. Washburn, and K. E. Kurtis, "Molecular engineering of superplasticizers for metakaolin-portland cement blends with hierarchical machine learning," *Advanced Theory and Simulations*, vol. 2, no. 4, p. 1800164, 2019.

[29] N.-D. Hoang, C.-T. Chen, and K.-W. Liao, "Prediction of chloride diffusion in cement mortar using multi-gene genetic programming and multivariate adaptive regression splines," *Measurement*, vol. 112, pp. 141–149, 2017.

[30] G. Konstantopoulos, E. P. Koumoulos, and C. A. Charitidis, "Testing novel portland cement formulations with carbon nanotubes and intrinsic properties revelation: nano-indentation analysis with machine learning on microstructure identification," *Nanomaterials*, vol. 10, no. 4, p. 645, 2020.

[31] J. Huang, P. G. Asteris, S. M. K. Pasha, A. S. Mohammed, and M. Hasanipanah, "A new auto-tuning model for predicting the rock fragmentation: a cat swarm optimization algorithm," *Engineering with Computers*, pp. 1–12, 2020.

[32] J. Huang, M. Koopialipoor, and D. J. Armaghani, "A combination of fuzzy delphi method and hybrid ann-based systems to forecast ground vibration resulting from blasting," *Scientific Reports*, vol. 10, p. 19397, 2020.

[33] N.-D. Hoang, "Estimating punching shear capacity of steel fibre reinforced concrete slabs using sequential piecewise multiple linear regression and artificial neural network," *Measurement*, vol. 137, pp. 58–70, 2019.

[34] S. M. Jamal, N. Deshpande, and S. Londhe, "Predicting strength of recycled aggregate concrete using artificial neural network, adaptive neuro-fuzzy inference system and multiple linear regression," *International Journal of Sustainable Built Environment*, vol. 5, pp. 355–369, 2016.

[35] F. Khademi, M. Akbari, S. M. Jamal, and M. Nikoo, "Multiple linear regression, artificial neural network, and fuzzy logic prediction of 28 days compressive strength of concrete," *Frontiers of Structural and Civil Engineering*, vol. 11, no. 1, pp. 90–99, 2017.

[36] V. Nilsen, L. T. Pham, M. Hibbard, A. Klager, S. M. Cramer, and D. Morgan, "Prediction of concrete coefficient of thermal expansion and other properties using machine learning,"

Construction and Building Materials, vol. 220, pp. 587–595, 2019.

[37] Y. Sun, G. Li, J. Zhang, and D. Qian, "Prediction of the strength of rubberized concrete by an evolved random forest model," *Advances in Civil Engineering*, vol. 2019, Article ID 5198583, 7 pages, 2019.

[38] X. Jiang and S. Li, "BAS: beetle antennae search algorithm for optimization problems," 2017, http://arxiv.org/abs/1710.10724.

[39] T. Joshi and U. Dave, "Evaluation of strength, permeability and void ratio of pervious concrete with changing w/c ratio and aggregate size," *International Journal of Civil Engineering and Technology*, vol. 7, pp. 276–284, 2016.

[40] Z. Dai, H. Li, W. Zhao et al., "Multi-modified effects of varying admixtures on the mechanical properties of pervious concrete based on optimum design of gradation and cement-aggregate ratio," *Construction and Building Materials*, vol. 233, p. 117178, 2020.

[41] N. Neithalath, *Development And Characterization Of Acoustically Efficient Cementitious Materials*, Portland Cement Association, Newyork, NY, USA, 2004.

[42] F. Imai and N. Shinnishi, *Permeability of No-Fines Concrete*, 1998.

[43] J. Sun, J. Zhang, Y. Gu, Y. Huang, Y. Sun, and G. Ma, "Prediction of permeability and unconfined compressive strength of pervious concrete using evolved support vector regression," *Construction and Building Materials*, vol. 207, pp. 440–449, 2019.

[44] A. K. Chandrappa and K. P. Biligiri, "Comprehensive investigation of permeability characteristics of pervious concrete: a hydrodynamic approach," *Construction and Building Materials*, vol. 123, pp. 627–637, 2016.

[45] A. A. Aliabdo, A. E. M. Abd Elmoaty, and A. M. Fawzy, "Experimental investigation on permeability indices and strength of modified pervious concrete with recycled concrete aggregate," *Construction and Building Materials*, vol. 193, pp. 105–127, 2018.

[46] B. Huang, H. Wu, X. Shu, and E. G. Burdette, "Laboratory evaluation of permeability and strength of polymer-modified pervious concrete," *Construction and Building Materials*, vol. 24, no. 5, pp. 818–823, 2010.

[47] Y. Zhang, H. Li, A. Abdelhady, and J. Yang, "Comparative laboratory measurement of pervious concrete permeability using constant-head and falling-head permeameter methods," *Construction and Building Materials*, vol. 263, p. 120614, 2020.

[48] A. K. Chandrappa and K. P. Biligiri, "Effect of pore structure on fatigue of pervious concrete," *Road Materials and Pavement Design*, vol. 20, pp. 1525–1547, 2019.

[49] S. O. Ekolu, S. Diop, and F. Azene, "Properties of pervious concrete for hydrological applications," *Concrete Beton*, vol. 144, pp. 18–25, 2016.

[50] H. Liu, G. Luo, H. Wei, and H. Yu, "Strength, permeability, and freeze-thaw durability of pervious concrete with different aggregate sizes, porosities, and water-binder ratios," *Applied Sciences*, vol. 8, no. 8, p. 1217, 2018.

Hindawi
Advances in Civil Engineering
Volume 2020, Article ID 8864766, 8 pages
https://doi.org/10.1155/2020/8864766

Research Article

Prediction of Low-Temperature Rheological Properties of SBS Modified Asphalt

Qian Chen,[1] Chaohui Wang,[1] and Liang Song[2]

[1]*School of Highway, Chang'an University, Xi'an 710064, China*
[2]*Xinjiang Transportation Planning Surveying and Design Institute, Urumqi 830006, China*

Correspondence should be addressed to Chaohui Wang; wchh0205@chd.edu.cn and Liang Song; 3359559@qq.com

Received 30 September 2020; Revised 9 November 2020; Accepted 20 November 2020; Published 2 December 2020

Academic Editor: B. Binici

The extreme learning machine (ELM) algorithm optimized by genetic algorithm (GA) was used to quickly predict the low-temperature rheological properties of styrenic block copolymer (SBS) modified asphalt through the properties of the raw materials. In this work, one hundred groups of survey data and test data were collected and analyzed. Fourteen vital raw material parameters, such as chemical composition indexes of matrix asphalt and technical indexes of SBS modifier, were selected as the input parameter. The stiffness modulus and *m*-value of SBS modified asphalt were taken as the output parameter. Then, the GA-ELM prediction model of low-temperature rheological properties was established. According to comparison and analysis with other prediction models, the accuracy and output stability of the GA-ELM prediction model were verified. The results show that the GA-ELM model had obvious accuracy and efficiency. It can be used to predict the low-temperature rheological properties of SBS modified asphalt. Compared with the traditional prediction models, the error of the GA-ELM model was reduced by 68.97–81.48%.

1. Introduction

At present, the main research methods of low-temperature performance of styrenic block copolymer (SBS) modified asphalt are the force ductility method, microscopic observation method, and rheological test method [1]. Strategic Highway Research Program (SHRP) mainly uses the bending beam rheological test (BBR) to evaluate the low-temperature performance of asphalt [2]. The stiffness modulus and creep rate (*m*-value) of asphalt are the core indexes. Some scholars have carried out a large number of macroexperiments and microanalysis. Shan et al. evaluated the effect of SBS on the linear and nonlinear rheological behavior of asphalt binder [3]. Chen et al. analyzed the characteristics and reasons for the rheological behavior of SBS modified asphalt with different SBS dosages [4]. Ren et al. studied the effect of trans-polyactenamer on rheological properties, microstructure, and thermal stability of crumb rubber (CR)/SBS modified asphalt [5]. The effects of stiffness modulus and *m*-value on low-temperature performance of SBS modified asphalt were determined [6, 7]. However, due to the complex source of matrix asphalt and SBS modifier and the small number of test samples, there are some differences in the research conclusions of different scholars. For SBS modified asphalt, the chemical composition of matrix asphalt and the physicochemical index of modifier have different effects on the performance of modified asphalt. As a result, many factors need to be considered in the performance evaluation of SBS modified asphalt [8], and the amount and time of the experiment are large relatively, limiting the promotion and application of SBS modified asphalt in the engineering field to a certain extent.

In order to reduce the amount of test and get more accurate results quickly, the energy coefficient method (ECM), grey target decision-making method (GTDM), backpropagation algorithm (BP), radial basis function algorithm (RBF), and other evaluation methods and algorithms were used to predict the performance of asphalt materials [9–13]. Yan et al. predicted the change of modulus

and phase angle of SBS polymer modified asphalt by exponential regression and linear regression [14]. Xu et al. established the performance prediction model of SBS modified asphalt by using principal component analysis (PCA) and partial least squares (PLS) [15]. Diab et al. predicted the viscosity and rheological behavior of SBS modified asphalt, according to Vinogradov Malkin and Phillips Deutsch models [16]. However, there are some problems with these methods [17, 18]. For example, it is easy to fall into the local optimum solution in the process of solving; the parameters are not easy to determine; the requirements of training samples are high relatively; and the training is difficult.

Therefore, many vital parameters, such as the chemical composition of matrix asphalt and physicochemical index of modifier, were selected. The extreme learning machine (ELM) algorithm was optimized by the genetic algorithm (GA) to form the GA-ELM algorithm. The prediction model for the low-temperature rheological properties of SBS modified asphalt based on the GA-ELM algorithm was established. Compared with the BP model and the ELM model, the accuracy and output stability of the GA-ELM prediction model were verified. It provided a new pathway for the study of low-temperature rheological properties of SBS modified asphalt.

2. Experiment

2.1. Test Method. According to Test Method for Separation of Asphalt into Four Fractions (NB/SH/T 0509-2010), CN, the component of asphalt, was measured [19]. The specific steps were as follows. The asphaltenes were precipitated from the sample with normal heptane. After filtration, the soluble inclusions in the precipitation were removed by reflux of normal heptane. The asphaltenes were obtained by dissolving precipitation with toluene reflux. Then, the deasphalting part was adsorbed on the alumina chromatographic column. In turn, normal heptane, toluene, and toluene-ethanol were used to obtain saturates, aromatics, and colloids.

2.2. Data Preparation. The raw material parameters of SBS modified asphalt were selected as the input parameter. They mainly consisted of the following: PG-grade, stiffness modulus, m-value, asphaltenes, colloids, aromatics, saturates of matrix asphalt and structure, block ratio (S/B), tensile strength, elongation, permanent deformation, hardness (HSD), and the content of SBS modifier. The stiffness modulus and m-value of SBS modified asphalt were taken as the output parameter.

In order to collect more sample data, on the basis of the data provided in the literature [8], more series of experiments were completed. The test parameters and indexes are shown in Tables 1 and 2. Finally, 100 sets of test data were collected. Among them, 80 sets of data were used for training, and 20 sets were used for testing. According to equation (1), the sample data were normalized.

$$X_i = \frac{X - X_{\min}}{X_{\max} - X_{\min}}, \tag{1}$$

where X_i are the normalized sample data, X are the initial sample data, $X_{\max}$ is the maximum of the initial sample data, and $X_{\min}$ is the minimum of initial sample data.

2.3. GA-ELM Prediction Model. ELM is a new feed-forward neural network. Compared with the traditional neural network with a single hidden layer, its hidden layer does not need iteration and has an obvious faster learning speed. However, the input layer weight matrix and the hidden layer threshold matrix of the ELM model are random. GA has a strong global optimization ability. The fitting accuracy of the ELM model can be improved by optimizing the above matrix with the GA. The GA-ELM model was established in the following steps [20].

(1) For input sample X_i, which had been normalized, the output matrix (H) of hidden layer neurons was calculated according to the following equation:

$$H = g\left(WX^T + b\right), \tag{2}$$

where W is the weight matrix of the input layer, b is the threshold matrix of the implicit layer, and g is the neuronal activation function of the hidden layer, which was a "sigmoid" function in this study.

(2) According to equation (3), the output value (P) of the ELM neural network was calculated.

$$P = \left(H^T\beta\right), \tag{3}$$

where β is the weight matrix from the implicit layer to the output layer, and the ELM neural network can be determined by calculating β.

(3) The given training output sample (Y) was used to replace the output value of the neural network. β can be obtained by solving the least square solution of the following equation:

$$\min_{\beta}\left\|H^T\beta - Y\right\|. \tag{4}$$

(4) The genetic algorithm was used to find the optimal initial W and b of the ELM neural network. Through fitness function, the genetic algorithm found the corresponding individuals of minimum fitness value through selection, crossover, and mutation operation.

(5) The optimal initial weight and threshold assignment of the ELM neural network were obtained by the genetic algorithm. The number of the hidden layers

Table 1: Test parameters and their indexes of matrix asphalt.

Type	Index						
	PG-grade	Stiffness modulus (–18°C) (MPa)	m-value (–18°C) (MPa · s^{-1})	Asphaltenes (%)	Colloids (%)	Aromatics (%)	Saturates (%)
SK 90	58-22	270–275	0.295–0.300	8.73–8.78	19.35–19.40	46.65–46.70	25.20–25.24
Shell 90	58-22	325–320	0.270–0.274	8.10–8.14	33.17–33.21	35.55–35.59	23.08–23.13
Kunlun 90	58-22	270–275	0.292–0.295	10.54–10.57	30.14–30.17	39.80–39.83	19.19–19.22
Zhenhai 90	58-22	265–269	0.288–0.293	7.24–7.28	30.08–30.12	39.62–39.65	23.00–23.03
Esso 70	58-16	311–315	0.190–0.195	15.50–15.52	20.14–20.17	44.20–44.26	20.09–20.13

Table 2: Test parameters and their indexes of SBS modifier.

Type	Index						
	Structure	Block ratio (S/B)	Tensile strength (MPa)	Elongation (%)	Permanent deformation (%)	Hardness	Content (%)
DG2	Star	40/60	>12.0	>650	<30	>79	3.6–5.2
T161B	Star	30/70	>18.0	>630	<25	>78	3.6–5.2
1320-115	Linear	30/70	>18.0	>750	<40	>70.5	3.6–5.2
YS4303	Star	30/70	>12.0	>590	<45	>65	3.6–5.2
YH791	Linear	30/70	>18.0	>700	<45	>60	3.6–5.2

Note. S/B is the abbreviation of styrene and butadiene. Test temperature was –18°C.

was determined, and the GA-ELM model was established.

(6) According to equations (5)–(7), mean absolute error (MAE), mean absolute percent error (MAPE), and root mean squared error (RMSE) were used as error criteria. The GA-ELM model was tested and evaluated by the test set samples. The algorithm flow is shown in Figure 1 [20].

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|P - P'\right|, \tag{5}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{P - P'}{P}\times 100\%\right|, \tag{6}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(P - P'\right)^2}, \tag{7}$$

where P is the true value, P' is a predictive value, and n is the number of test sample data.

3. Results and Discussion

3.1. Parameter Optimization of the ELM Model. The key operation parameter of the ELM model is the number of hidden layers. The operation parameters of the genetic algorithm include population size, crossover probability, mutation probability, and maximum iterations. In practical applications, it is often necessary to undergo a large number of tests before the reasonable range of these parameters is determined. Based on the test set data, the superior hidden layer number of the ELM prediction model was determined through multiple tests, as shown in Figure 2.

Figures 2(a) and 2(b) show that with the increase of the hidden layer number in the ELM model, the error decreased first and then increased. When the number of the hidden layers was in the range of 60–70, the error was small. After many tests, the optimal hidden layer number of the ELM model was determined to be 65. At this time, MAE, MAPE, and RMSE were 369.37%, 1.66%, and 502.77%, respectively. It was noteworthy that MAE and RMSE were large and needed to be reduced in the subsequent optimization process. For m-value, Figures 2(c) and 2(d) show that the optimal hidden layer number was determined to be 55. Its MAE, MAPE, and RMSE were 0.27%, 0.85%, and 0.32%, respectively.

3.2. Parameter Optimization of the GA. After repeated cyclic tests, the optimal values of the four parameters of the GA were determined by using the Sheffield toolbox in MATLAB software, as shown in Figure 3. To simplify the simulation process, when population size, crossover probability, and mutation probability were optimized, the maximum iterations was set to 50.

Figure 3 indicates that with the change of the four parameters of the GA, the corresponding error values also changed regularly. In general, the recommended ranges of population size, crossover probability, mutation probability, and maximum iterations are 40–100, 0.40–0.80, 0.001–0.1, and 100–300, respectively. According to the test results of the stiffness modulus prediction model, the optimal ranges of population size, crossover probability, mutation probability, and maximum iterations were 40–45, 0.40–0.50, 0.0005–0.001, and 300–400, respectively. As for the m-value prediction model, the optimal ranges of population size,

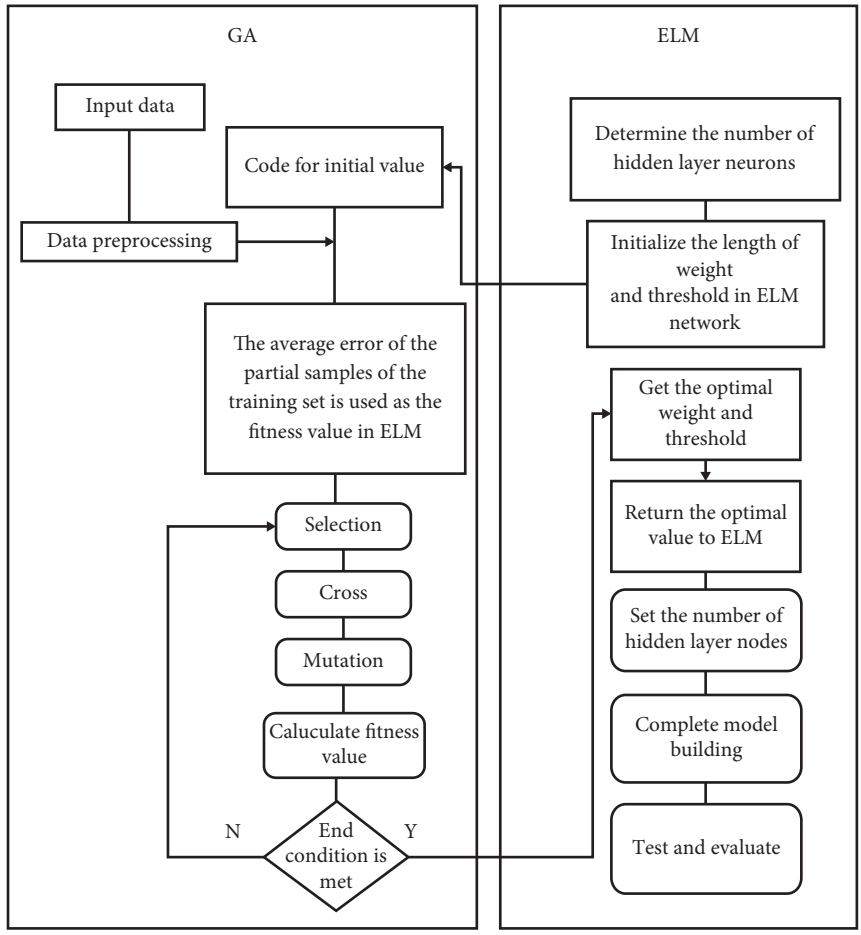


Figure 1: The algorithm flow of the GA-ELM model.

crossover probability, mutation probability, and maximum iterations were 45–50, 0.70–0.80, 0.0005–0.001, and 200–300, respectively. At this time, the genetic algorithm has better convergence to the optimal weights and thresholds. Due to the large scale of the weight matrix and threshold matrix, their specific values are no longer described.

3.3. Prediction Accuracy of the GA-ELM Model. The Pearson correlation test was used to verify the prediction accuracy of the GA-ELM model. The discriminant coefficient (R^2) of the fitting function between the true value and predictive value of the test set sample was calculated, as shown in Figure 4. The accuracy of the GA-ELM prediction model for low-temperature rheological properties was determined by analyzing the goodness of fit between predictive data and true data.

Figure 4 shows that the true value of test set samples is very close to the predictive value. The correlation coefficients (R) of the fitting function of the two models are 0.9997 and 0.9992, respectively. This indicated that the predictive value was strongly correlated with the true value, and the prediction accuracy of the prediction model was high. In addition, the corresponding discriminant coefficients (R^2) are 0.9995 and 0.9985, respectively, which indicates that the GA-ELM prediction model could maintain the stability of output based on high prediction accuracy.

3.4. Contrastive Analysis of Different Models. To further verify the output stability of the GA-ELM model, the BP model, the ELM model, and the GA-ELM model were used to predict 20 sets of test sample data. For stiffness modulus, after many simulation calculations, the neuron node number of hidden layers in the BP model was set to 15. The hidden layer number in the ELM model was set to 65. The population size, crossover probability, mutation probability, and maximum iterations of the GA were set to 45, 0.80, 0.001, and 350, respectively. And for the m-value, the neuron node number of hidden layers in the BP model was set to 15. The hidden layer number in the ELM model was set to 55. The population size, crossover probability, mutation probability, and maximum iterations of the GA were set to 50, 0.50, 0.001, and 200, respectively. The prediction results based on different models are shown in Figure 5.

Figure 5 shows that the three models achieved consistent prediction results. Still, the prediction error of the GA-ELM model was significantly smaller than that of the BP model and the ELM model. For stiffness modulus, the MAE, MAPE, and RMSE of the GA-ELM model were

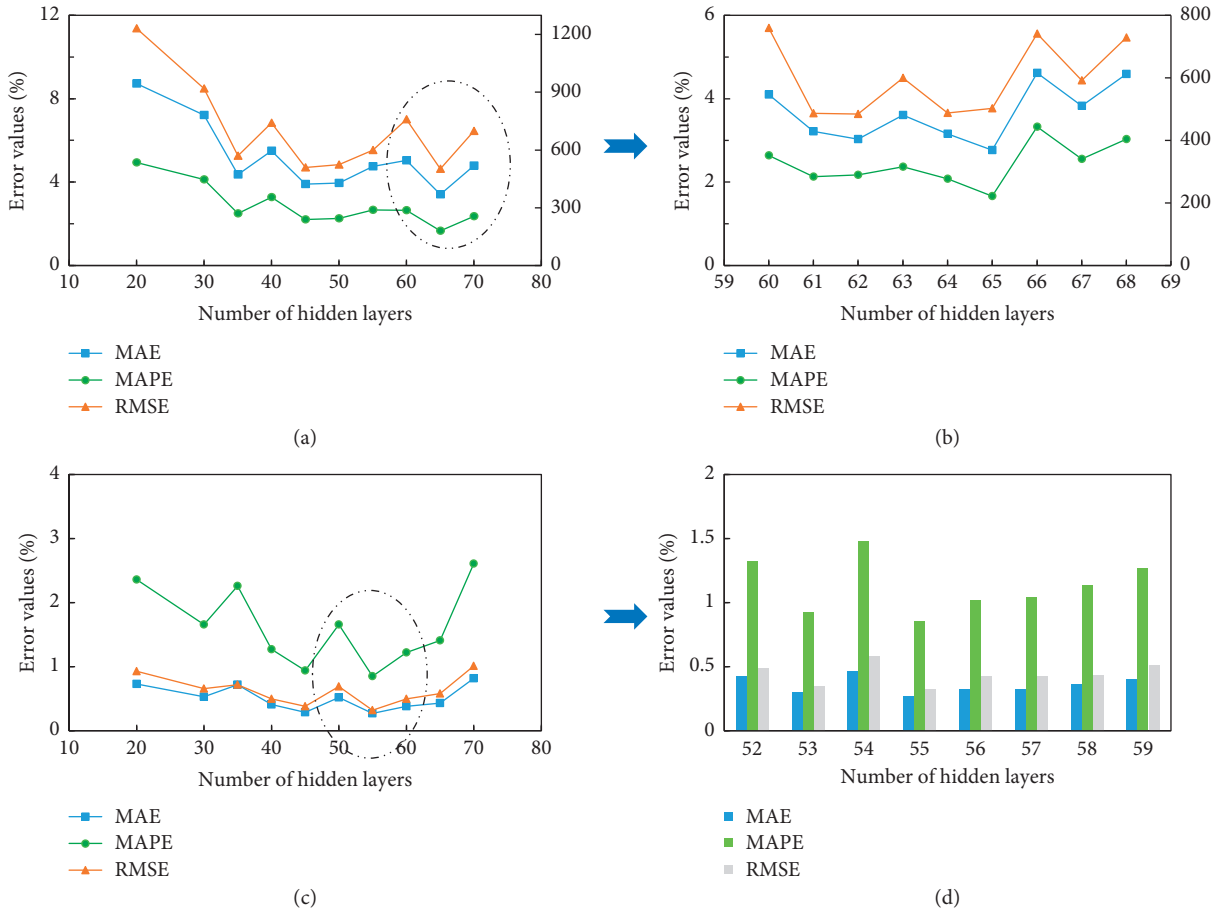


Figure 2: The optimization results of the hidden layer in the ELM model. (a) Stiffness modulus/20–70 layers. (b) Stiffness modulus/60–68 layers. (c) m-value/20–70 layers. (d) m-value/52–59 layers.

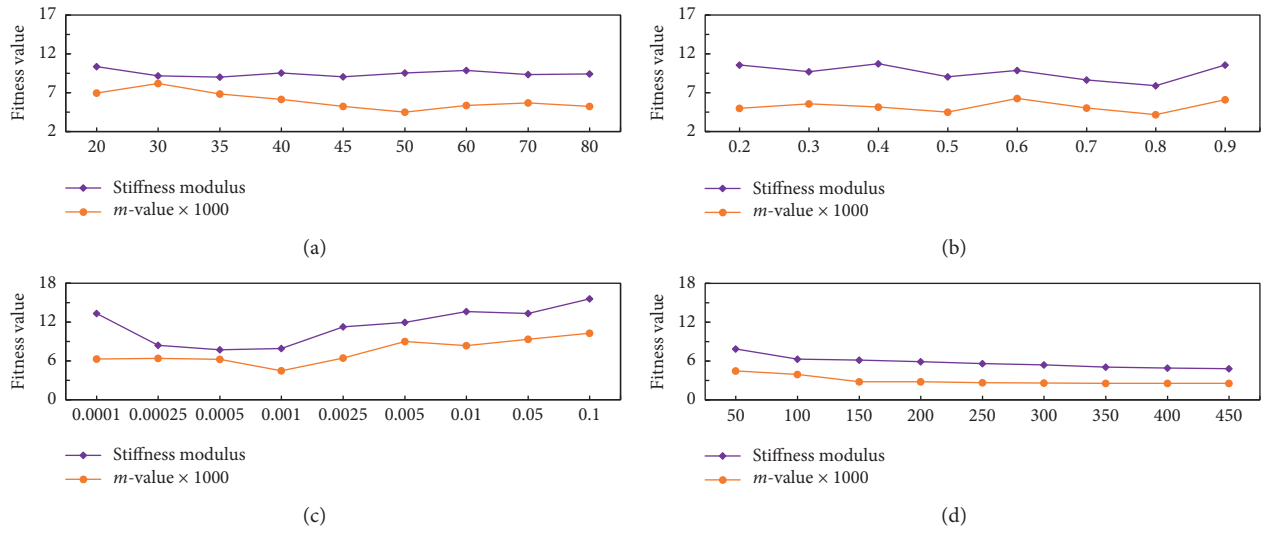


Figure 3: The parameter optimization results of the GA. (a) Population size. (b) Crossover probability. (c) Mutation probability. (d) Maximum iterations.

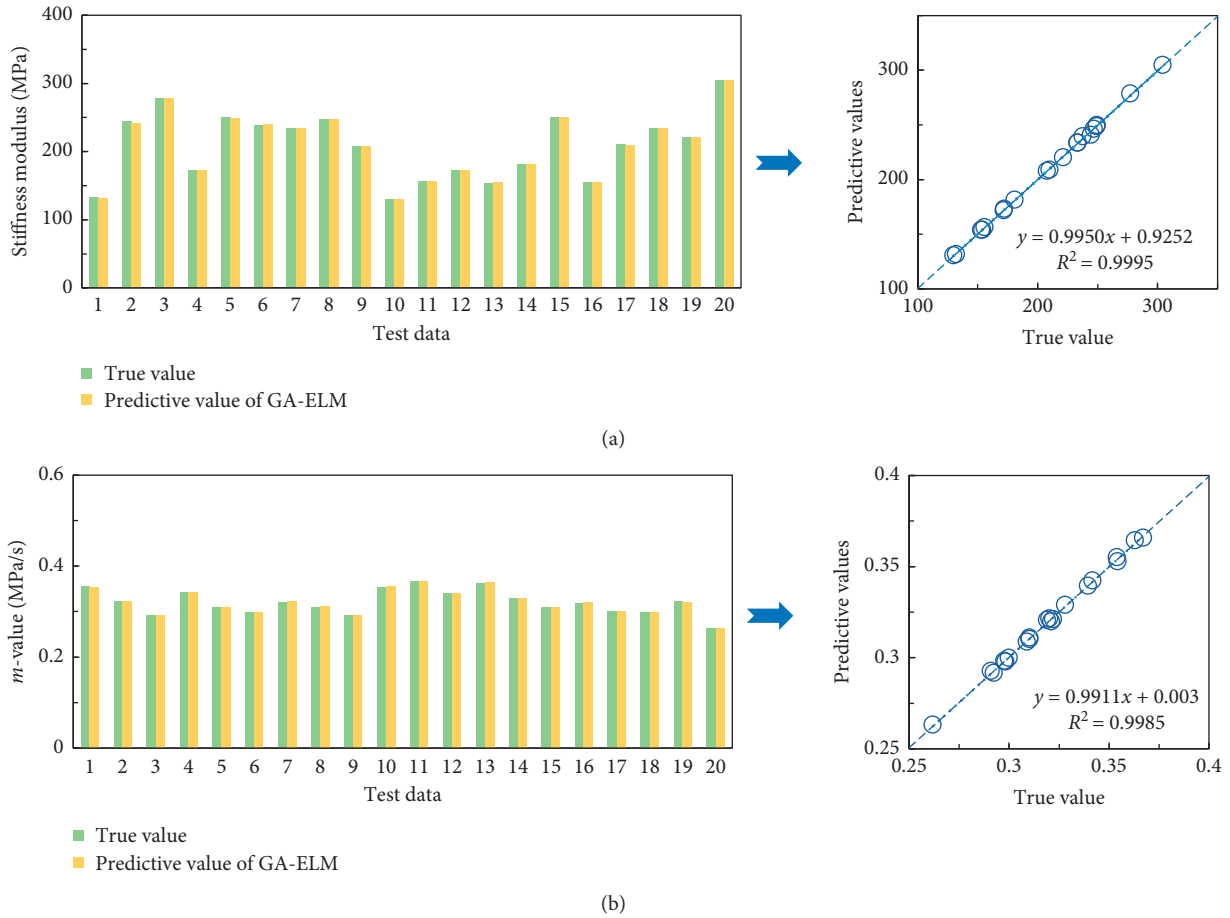


Figure 4: Accuracy of the GA-ELM prediction model. (a) Stiffness modulus. (b) m-value.

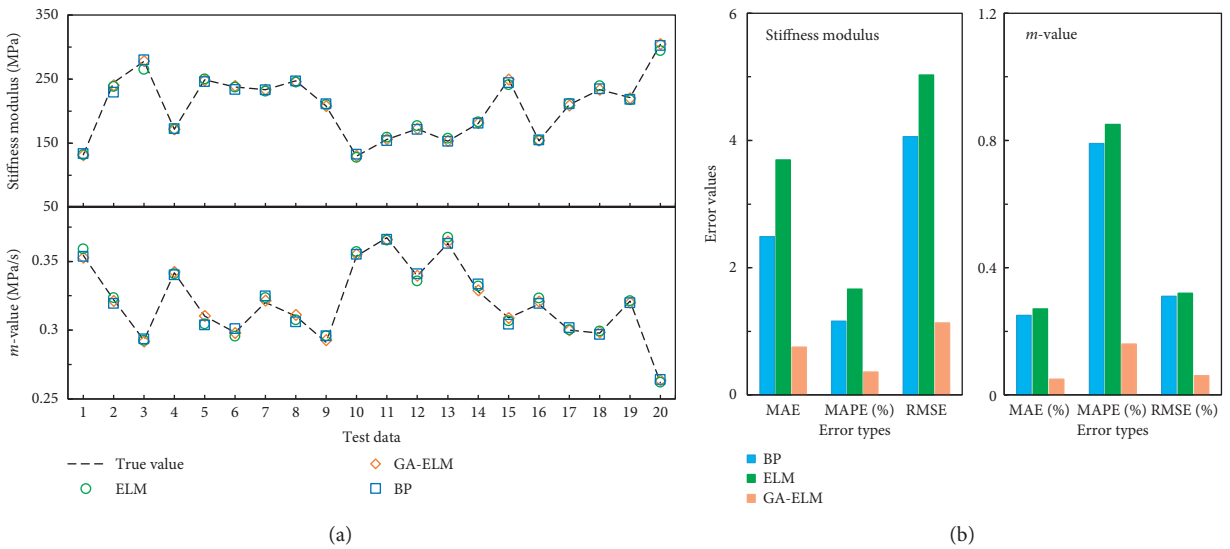


Figure 5: Prediction accuracy comparison of different models. (a) Predictive value comparison. (b) Error comparison.

0.7539, 0.36%, and 1.1323, respectively. Compared with the BP model and the ELM model, the error of the GA-ELM model decreased by 68.97–72.10% and 83.63–87.41%, respectively. And for the *m*-value, the MAE, MAPE, and RMSE of the GA-ELM model were 0.0005, 0.16%, and 0.0006, respectively. Compared with the BP model and the ELM model, the error of the GA-ELM model decreased by 79.75–80.65% and 81.18–81.48%, respectively. This indicated that the GA-ELM model had better advantages in prediction accuracy and efficiency than the traditional prediction model.

4. Conclusion

(1) The GA-ELM prediction model can rapidly predict the low-temperature rheological properties of SBS modified asphalt. Compared with the traditional prediction model, the GA-ELM model had obvious accuracy and efficiency. The error was reduced by 68.97–81.48%.

(2) The applicable boundaries of the input parameter are as follows. For the matrix asphalt, PG-grade is 58-22 and 58-16; stiffness modulus is 265–320 MPa; *m*-value is 0.19–0.3 MPa/s; asphaltenes are 7.24–15.52%; colloids are 19.35–33.21%; aromatics are 35.55–46.70%; and saturates are 19.19–25.24%. For the SBS modifier, block ratio (S/B) is 40/60 and 30/70; tensile strength is over 12%; elongation is over 590%; permanent deformation is less than 45%; hardness is over 60; and the content is 3.6–5.2%.

(3) Due to the use of the genetic algorithm, the training efficiency of the GA-ELM model was lower than that of the original ELM model. In addition, the optimum number of the hidden layers was influenced by the input of data. The more the number and type of data, the less the optimum number of hidden layers. However, the optimum number will remain within a specific range.

(4) The GA-ELM model still belongs to the black-box operation. It should be further improved in future studies. With the updating of the algorithm and software, the prediction model will be established based on input parameters, which is more conducive to regulating the model and the feedback of the results. In addition, more raw material parameters and their corresponding performance indexes should be collected in the future to further expand the prediction range of the model.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was sponsored by the Natural Science Foundation of Xinjiang Uygur Autonomous Region (2020D01A92), Fundamental Research Funds for the Central Universities, CHD (300102219701 and 300102219314), Science and Technology Project of Transportation Industry in Xinjiang Uygur Autonomous Region (2019-ZD1-016), and China Postdoctoral Science Foundation (2020M683709XB).

References

[1] S. Rezaei, H. Ziari, and S. Nowbakht, "Low temperature functional analysis of bitumen modified with composite of nano-SiO_2 and styrene butadiene styrene polymer," *Petroleum Science and Technology*, vol. 34, no. 5, pp. 415–421, 2016.

[2] R. Tarefder and S. Yousefi, "Rheological examination of aging in polymer-modified asphalt," *Journal of Materials in Civil Engineering*, vol. 28, no. 2, Article ID 04015112, 2016.

[3] L. Shan, X. Qi, X. Duan, S. Liu, and J. Chen, "Effect of styrene-butadiene-styrene (SBS) on the rheological behavior of asphalt binders," *Construction and Building Materials*, vol. 231, Article ID 117076, 2020.

[4] Z. Chen, H. Zhang, H. Duan, and C. Wu, "Determination of time-temperature superposition relationship of SBS modified asphalt based on special rheological phenomenon caused by SBS-formed structure in asphalt matrix," *Construction and Building Materials*, vol. 260, Article ID 119835, 2020.

[5] S. Ren, X. Liu, M. Li, W. Fan, J. Xu, and S. Erkens, "Experimental characterization of viscoelastic behaviors, microstructure and thermal stability of CR/SBS modified asphalt with TOR," *Construction and Building Materials*, vol. 261, Article ID 120524, 2020.

[6] B. V. Kök, M. Yilmaz, and A. Geçkil, "Evaluation of low-temperature and elastic properties of crumb rubber- and SBS-modified bitumen and mixtures," *Journal of Materials in Civil Engineering*, vol. 25, no. 2, pp. 257–265, 2013.

[7] W. Huang, X. Fu, Y. Li, and S. Liu, "Evaluation of low temperature performance and correlation analysis on low temperature indexes of SBS modified asphalts," *Journal of Building Materials*, vol. 20, no. 3, pp. 456–463, 2017, in Chinese.

[8] X. Ma, H. Chen, X. Zhang, M. Xing, P. Yang, and Z. Wang, "An analysis over the correlation between low temperature rheological properties of SBS-modified asphalt and raw materials' performance parameters," *Materials Review*, vol. 32, no. 22, pp. 3885–3890, 2018, in Chinese.

[9] Q. Chen, C. Wang, P. Wen, M. Wang, and J. Zhao, "Comprehensive performance evaluation of low-carbon modified asphalt based on efficacy coefficient method," *Journal of Cleaner Production*, vol. 203, pp. 633–644, 2018.

[10] Q. Chen, C. Wang, Z. Qiao, and T. Guo, "Graphene/tourmaline composites as a filler of hot mix asphalt mixture: preparation and properties," *Construction and Building Materials*, vol. 239, Article ID 117859, 2020.

[11] H. Liu, A. Sha, Z. Tong, and J. Gao, "Autonomous microscopic bunch inspection using region-based deep learning for evaluating graphite powder dispersion," *Construction and Building Materials*, vol. 173, pp. 525–539, 2018.

[12] Y. Cao, A. Sha, Z. Liu, J. Li, and W. Jiang, "Energy output of piezoelectric transducers and pavements under simulated

traffic load," *Journal of Cleaner Production*, vol. 279, Article ID 123508, 2021.

[13] Q. Chen, S. Wang, C. Wang et al., "Modified waterborne epoxy as a cold pavement binder: preparation and long-term working properties," *Journal of Materials in Civil Engineering*, vol. 33, 2021.

[14] C. Yan, W. Huang, P. Lin, Y. Zhang, and Q. Lv, "Chemical and rheological evaluation of aging properties of high content SBS polymer modified asphalt," *Fuel*, vol. 252, pp. 417–426, 2019.

[15] M. Xu, Y. Zhang, P. Zhao, and C. Liu, "Study on aging behavior and prediction of SBS modified asphalt with various contents based on PCA and PLS analysis," *Construction and Building Materials*, vol. 265, Article ID 120732, 2020.

[16] A. Diab, Z. You, X. Li, J. C. Pais, X. Yang, and S. Chen, "Rheological models for non-Newtonian viscosity of modified asphalt binders and mastics," *Egyptian Journal of Petroleum*, vol. 29, no. 2, pp. 105–112, 2020.

[17] Q. Chen, C. Wang, X. Sun, Y. Cao, T. Guo, and J. Chen, "Evaluation and prediction for effect of conductive gussasphalt mixture on corrosion of steel bridge deck," *Construction and Building Materials*, vol. 228, Article ID 116837, 2019.

[18] R. Prabu and R. Harikumar, "A performance analysis of GA-ELM classifier in classification of abnormality detection in electrical impednce tomography (EIT) lung images," *Journal of Scientific and Industrial Research*, vol. 75, pp. 404–411, 2016.

[19] NB/SH/T 0509-2010, *Test Method for Separation of Asphalt into Four Fractions*, China Petrochemical Press, Beijing, China, 2010.

[20] Y. Mei, Q. Sun, L. Yu, C. Wang, and H. Xiao, "Grain size prediction of aluminum alloy dies castings based on GA-ELM," *Acta Metallurgica Sinica*, vol. 53, no. 9, pp. 1125–1132, 2017, in Chinese.