Journal of Science and Technology in Civil Engineering, HUCE, 2023, 17 (3): 65–79

INVESTIGATING COMPRESSIVE STRENGTH OF CONCRETE CONTAINING STEEL FIBER BY DATA-DRIVEN APPROACH

Tran Van Quan[a], Nguyen Ngoc Linh[b], Nguyen Ngoc Tan[b,*]

[a] *University of Transport Technology, 54 Trieu Khuc street, Thanh Xuan district, Hanoi, Vietnam*
[b] *Faculty of Building and Industrial Construction, Hanoi University of Civil Engineering, 55 Giai Phong street, Hai Ba Trung district, Hanoi, Vietnam*

Article history:
Received 14/4/2023, Revised 31/5/2023, Accepted 01/6/2023

Abstract

The main objective of this paper is to use the data-driven approach to predict and study the factors affecting the compressive strength of steel fiber concrete. Therefore, six machine learning (ML) models were evaluated against a database of 166 samples and ten input variables, including Cement content, Water content, Silica fume content, Steel fiber content, Coarse aggregate content, Sand content, Superplasticizer content, Fiber diameter, Fiber length, Fly ash content. Six ML models, including Artificial Neural Network (ANN), K-Nearest Neighbor (KNN), Categorical Boosting (CatB), Random Forest (RF), Gradient Boosting (GB), and Extreme Gradient Boosting (XGB), are evaluated against performance metrics and validated by 1000 Monte Carlo simulations. The compressive strength prediction performance by the ML models is arranged in descending order as follows XGB > GB > CatB > RF > ANN > KNN. The two best models can predict the compressive strength of steel fiber concrete to be GB with a coefficient of determination (R^2) of 0.9874 and a root mean square error ($RMSE$) of 2.5763 MPa and XGB with an R^2 of 0.9926 and an $RMSE$ of 1.9814 MPa for the testing dataset. The influence of the ten variables can be arranged in descending order as follows: Cement content > Water content > Silica fume content > Steel fiber content > Coarse aggregate content > Sand content > Superplasticizer content > Fiber diameter > Fiber length > Fly ash content. Among them, Cement content, Silica fume content, and Steel fiber content have a positive effect on improving the compressive strength of concrete. The steel fiber content used should be less than 1.5% of concrete volume to improve the efficiency of steel fiber in concrete. Meanwhile, Fiber diameter and Fiber length have a minimal influence on the compressive strength of steel fiber concrete.

Keywords: concrete; steel fiber; compressive strength; data-driven; machine learning.

https://doi.org/10.31814/stce.huce2023-17(3)-06

1. Introduction

Steel fibers have been popularly used in manufacturing ultra-high and high-performance concretes. Adding steel fibers dispersed to the mixture can improve the mechanical properties of steel fiber concrete compared to ordinary concrete, especially its tensile strength and mechanical behavior after cracking [1, 2]. It is due to the bridging action of steel fibers at the cracking locations that help to delay the crack mouth opening while maintaining the ductility of concrete. Thanks to the above advantages, steel fiber concrete could be used to partially or totally replace reinforcing bars in structural members. Several recent studies stated that fiber steel concrete is an interesting material for producing reinforced concrete structures under aggressive environmental conditions (e.g., chloride attack) because it enhances the load-bearing capacity and avoids brittle failure due to reinforcement corrosion [3, 4].

The highest compressive strength is obtained at a volume of 1.5%, according to Song and Hwang [5], who also noted that the compressive strength increases at the volume fraction of 1.5% steel

*Corresponding author. *E-mail address:* tannn@huce.edu.vn ()

fiber employed. The compressive strength significantly declines as the volume of steel fibers reaches 2%. Also, as the volume proportion of steel fibers rises, so does the high-strength fiber-reinforced concrete's splitting tensile strength. The manufacturing process (e.g., time, order), the volume of steel fiber, and the geometry of steel fiber have all been found to affect how the fiber separates during mixing. The design of steel fiber concrete strength is complicated because it depends on many input factors, such as steel fiber content, length, and diameter. In addition, other design factors, including cement content, water content, sand content, coarse aggregate content, superplasticizer content, silica fume content, and fly ash content, also greatly affect the compressive strength of steel fiber concrete. A large number of design inputs complicates the design of the compressive strength of concrete using steel fibers. This leads to costly and difficult designs using traditional methods such as destructive tests in the laboratory. Therefore, it is necessary to use new approaches to determine the compressive strength of concrete using steel fibers. In addition, analyzing the influence of input factors above also helps to easily design the compressive strength of concrete using steel fibers.

Currently, Big Data Approach combined with ML algorithms is used a lot to solve complex correlations in problems of engineering science, especially in civil engineering. Geotechnical engineering and materials science are two challenging fields of civil engineering that use data-driven or ML methodologies [6–11]. Geotechnical engineering is currently applying some ML techniques, such as ANN and SVR (Support Vector Regression), in order to forecast soil parameters via ML approaches, such as soil strength and the foundation's load-bearing capacity [12, 13]. Also, the coefficient of chloride diffusion was predicted by Tran [11] using a data-driven approach.

Regarding the compressive strength of concrete using steel fiber, Li et al. [14] proposed ML methods, such as individual and ensemble models, that were taken into consideration to estimate the 28-day compressive strength of steel fiber concrete. SVR algorithm, a single technique, and two ensemble approaches (namely SVR-AdaBoost and SVR bagging) were used in this study. To evaluate the effectiveness of each method, k-fold cross-validation, statistical analysis, and coefficient of determination were used. The SVR-AdaBoost approach was the most accurate, with an R^2 value of 0.96. Based on the SVR-Adaboost model, a sensitivity technique was also employed to evaluate the impact of factors on the accuracy of the predictions. Khan et al. [15] also estimated the compressive strength of steel fiber concrete by a high-performance ML model, which is Random Forest, with an $R^2 = 0.96$. Cement content had the most positive effect on the compressive strength of steel fiber concrete, according to the SHapley Additive exPlanations (SHAP) value. These two studies both use ten input variables including cement content, water content, sand content, coarse aggregate content, superplasticizer content, silica fume content, fly ash content, steel fiber content, length of steel fiber, and diameter of steel fiber for building an ML model to predict the compressive strength of steel fiber concrete. However, the influence of these factors on the compressive strength of steel fiber concrete has not been quantified in the study of Li et al. [14] and Khan et al. [15]. In addition, the accuracy of the ML model in assessing the compressive strength of steel fiber concrete can be improved.

Some experimental investigations were performed regarding factors that affect the compressive strength of steel fiber concrete. The water-to-cement ratio (W/C) significantly affects the compressive strength of concrete. The compressive strength decreases as the W/C ratio increases, following Abbass et al. [16]. According to Kim et al. [17], the compressive strength of steel fiber concrete increases when the sand/aggregate ratio increases. The workability and compressive strength are improved when the superplasticizer content reaches 1.5% of cement content, according to Aruntas et al. [18]. According to Nili and Afroughsabet [19], adding more silica fume to concrete increases its compressive strength. Challoob and Srivastava [20] investigated how fly ash and steel fibers affect the strength characteristics of pozzolana cement concrete. Using fiber volume fractions of up to 1% improves the compressive strength of steel fiber concrete, according to Köksal et al. [21]. Each variable effect on the compressive strength of steel fiber content has been individually evaluated, but a simultaneous assessment of multiple variables on the compressive strength of steel fiber concrete has not been thoroughly studied.

Therefore, the main objective of this study is to quantify the influence of the ten inputs mentioned above on the compressive strength of steel fiber concrete based on data-driven approaches combined with sophisticated algorithms such as SHAP and Partial Dependence Plot 1D. These algorithms need to be based on an ML model that has high performance in predicting the compressive strength of steel fiber concrete. The performance of the ML model depends greatly on the statistical distribution of the samples in the database used to build the model. Therefore, the first part of the paper will describe the statistical distribution of variables, including an input variable and an output variable compressive strength of the database gathered from previously published studies by Li et al. [14]. Then, he second part of the paper will be used to briefly describe the basis of data-driven approaches, including using six popular models: ANN, KNN, CatB, RF, GB, and XGB. In this study, six popular and easily accessible algorithms are used in open-source libraries written in Python programming language. In this study, all hyperparameters of the ML models use the default value proposed by the Sklearn library [22]. Therefore, the third part of the paper will evaluate the performance of these available models in predicting the compressive strength of steel fiber concrete based on performance criteria such as R^2 and *RMSE*. Performance will be validated using Monte Carlo simulation. The best-performing ML model will be used in conjunction with the SHAP technique and Partial Dependence Plot 1D in quantifying the influencing factors as well as predicting the compressive strength of steel fiber concrete.

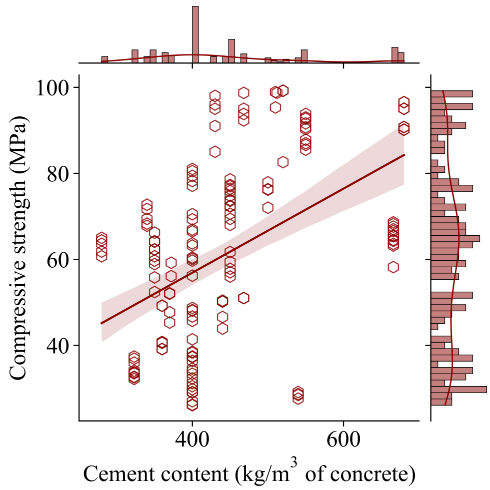
2. Database description and analysis

To select a high-performance ML model for determining the compressive strength of steel fiber concrete, a database of 166 samples, 10 input variables and 1 output compressive strength collected by Li et al. [14]. The database only considers steel fiber concrete compressive strength at 28 days; therefore, curing age will not be considered an input variable for ML model selection.

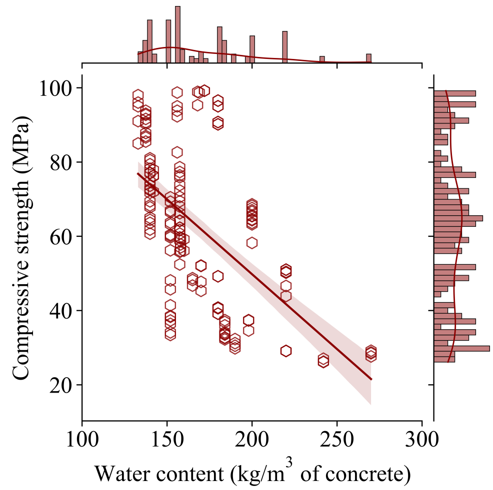
Table 1. Statistical value of space variables

Variable	Unit	Mean	StD	Min	$Q_{25\%}$	Median	$Q_{75\%}$	Max
Cement content	kg/m^3	446.05	105.31	280	370.5	400	500	680
Water content	kg/m^3	171.22	30.88	133.00	152.00	158.00	184.00	270.00
Sand content	kg/m^3	785.83	156.56	582.00	682.00	743.00	835.00	1350.00
Coarse aggregate content	kg/m^3	936.54	260.63	0.00	872.00	1050.50	1100.00	1170.00
Superplasticizer content	%	0.95	1.79	0.00	0.00	0.15	1.00	9.00
Silica fume content	%	5.99	11.74	0.00	0.00	0.00	10.00	43.00
Fly ash content	%	1.36	5.68	0.00	0.00	0.00	0.00	30.00
Steel fiber content	%	0.86	0.61	0.00	0.50	1.00	1.50	2.00
Fiber length	mm	40.67	15.76	0.00	30.00	35.00	60.00	60.00
Fiber diameter	mm	0.60	0.19	0.00	0.50	0.62	0.75	0.90
Compressive strength	MPa	61.39	21.49	26.10	40.93	62.75	76.85	99.20

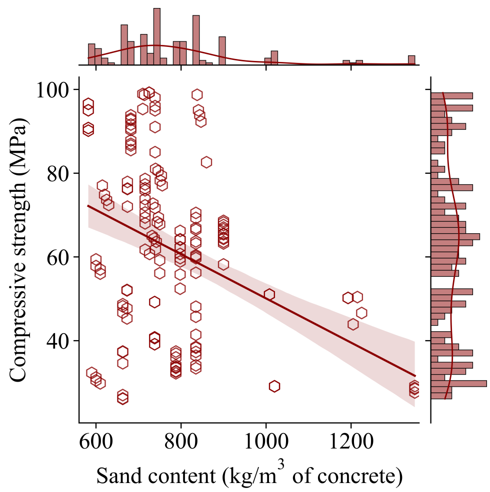
The database description is very important in improving the accuracy as well as determining the applicable domain of the ML model. The results of the data statistics are described in Table 1. The distribution of data between the input variable and the compressive strength is shown in Fig. 1. Based


Compressive strength (MPa)
Cement content (kg/m3 of concrete)

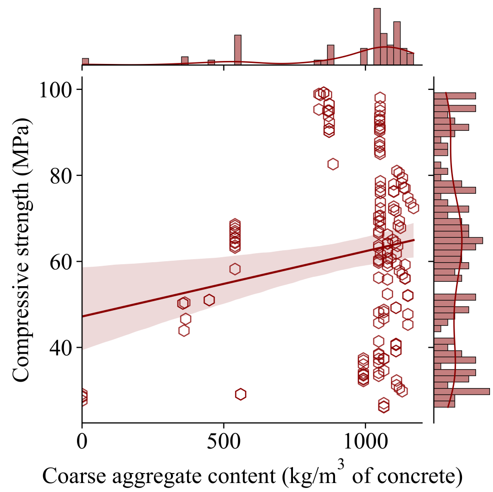
(a)


Compressive strength (MPa)
Water content (kg/m3 of concrete)

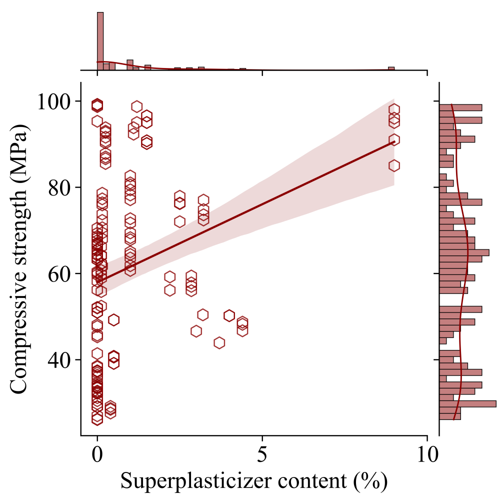
(b)


Compressive strength (MPa)
Sand content (kg/m3 of concrete)

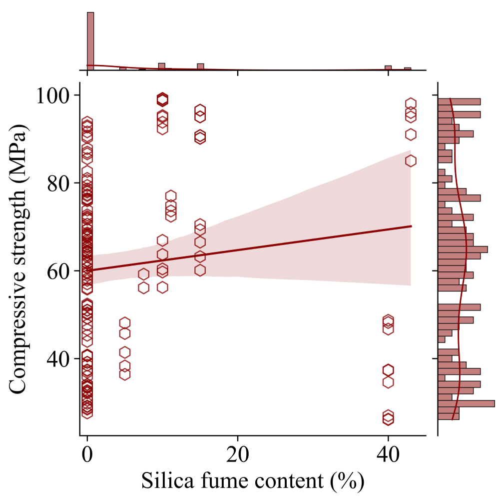
(c)


Compressive strength (MPa)
Coarse aggregate content (kg/m3 of concrete)

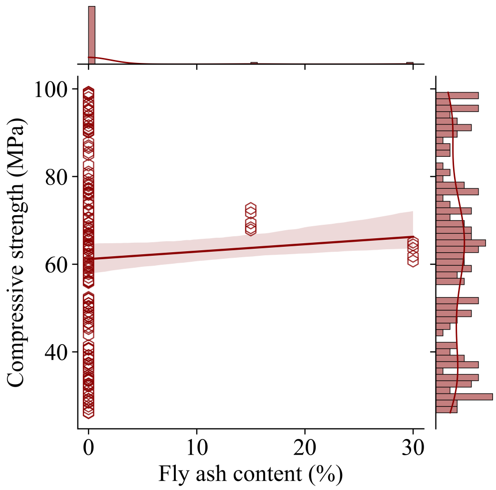
(d)


Compressive strength (MPa)
Superplasticizer content (%)

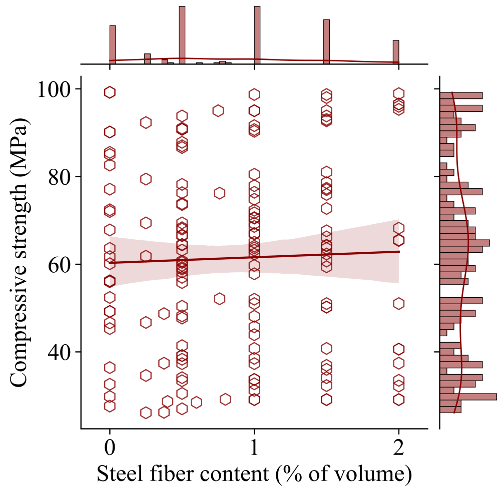
(e)


Compressive strength (MPa)
Silica fume content (%)

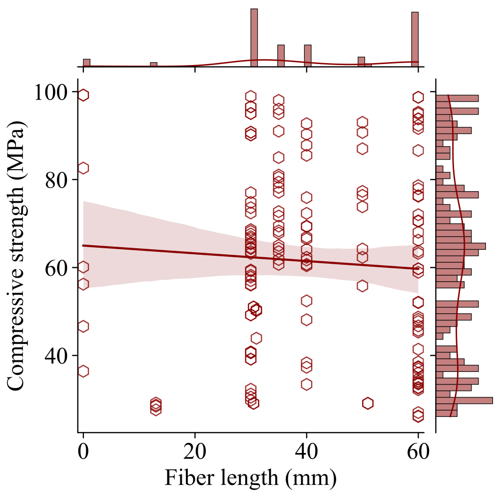
(f)



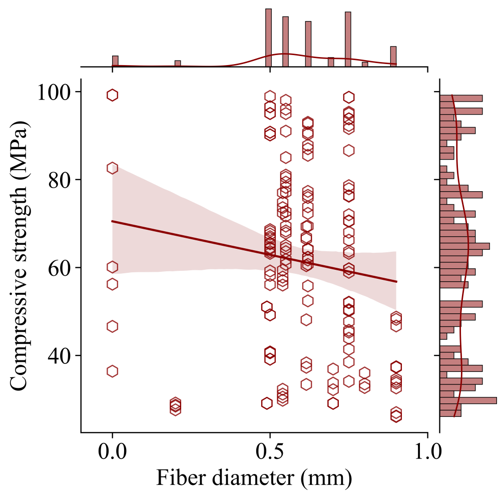
(g)



(h)



(i)



(j)

Figure 1. Description of input variables and compressive strength

on the data distribution in Fig. 1, a linear correlation is established between each input variable and compressive strength. Linear correlation allows preliminary assessment of the influence of factors on the compressive strength of steel fiber concrete. Fig. 1(a) shows that increasing cement content improves the compressive strength of steel fiber concrete. In contrast, increasing water content reduces the compressive strength of steel fiber concrete, as shown in Fig. 1(b). Figs. 1(c) and 1(d) show that increasing sand content reduces compressive strength while increasing coarse aggregate content improves compressive strength. Superplasticizer and increased silica fume content increase compressive strength, as indicated in Figs. 1(e) and 1(f). The linear correlation in Figs. 1(h) – 1(j) has not shown a significant influence of the steel fiber content as well as the geometrical properties of steel fiber, including fiber length and diameter. Fly ash has a weak impact on the compressive strength of steel fiber concrete, as shown in Fig. 1(g).

Linear correlation is difficult to accurately describe the influence of the input factors on the compressive strength of steel fiber concrete. Therefore, techniques to further analyze these correlations are at the end of the study.

The range of values to be able to use the ML models developed in this study effectively is described in Table 1. The compressive strength is limited, ranging from 26.1 MPa to 99.2 MPa. This range corresponds to conventional concrete and high-performance concrete. The steel fiber content varies from 0% to 2% by volume. The geometry of steel fiber ranges from 13 mm to 60 mm for length fiber and 0.2 mm to 0.9 mm for diameter fiber. The binder content, including cement, varies from 280 to 680 kg/m^3, silica fume content is 0 to 43%, and fly ash content is 0 to 30 % of the cement content for binder replacement. The aggregate used includes sand content varying from 582 to 1350 kg/m^3 and coarse aggregate content from 0 to 1170 kg/m^3. Some concrete samples do not use coarse aggregate in the mix composition and use sand entirely. The amount of water varies from 133 to 270 kg/m^3, with the superplasticizer content varying from 0 to 9% of the concrete volume.

3. Data-driven approach

3.1. Machine learning algorithms

In this section, six popular ML algorithms including ANN, KNN, CatB, RF, GB and XGB will be briefly described. The human brain systems are greatly simplified in ANN models. Artificial neurons are computational units similar to those in the biological nervous system. The input, hidden, and output layers make up the majority of the ANN model. First, a signal connects each neuron in the n^{th} layer to the neurons in the $(n + 1)^{th}$ layer. Then, a weight is applied to each link. Next, each input may be multiplied by its appropriate weight to determine the output. Finally, an activation function processes the output to obtain the final ANN output. The ANN could be helpful in resolving many technical and scientific issues. As a result, the ANN has a wide range of applications, including signal processing, picture compression, function approximation, differential equations, stock market prediction, and medical diagnostics. According to Tamura and Tateishi [23], the number of neurons for one hidden layer can be calculated by formulation $N_i - 1$ with N_i being the number of input variables. In this study, $N_i = 10$, the number of neurons for the hidden layer is considered equal to 10.

Because of its great effectiveness and simple interpretability, the KNN approach has been widely employed in classification challenges [24]. In KNN, a new sample is categorized in accordance with its k nearest neighbors in the neighbor space that have known labels. KNN is a lazy learning method that delays computation till later while approximating the function locally. The output of a sample is determined for regression issues by averaging the outputs of its k nearest neighbors.

A startup called Yandex created the open-source ML algorithm Categorical Boosting in 2017 [25]. One of CatB core edges is its ability to integrate a variety of different data types, such as images, audio, or text features, into one framework. However, CatB also offers an idiosyncratic way of handling categorical data, requiring a minimum of categorical feature transformation, unlike most ML algorithms that cannot handle non-numeric values.

Breiman [26] originally introduced Random Forest, which groups several classification or regression trees. One of the decision tree-based algorithms is RF, where each tree serves as a vote and forms the basis of the algorithm's decision-making. As a result, group learning techniques with the unique outcomes of each tree frequently produce better outcomes. RF is a modified version of bagging or bootstrap aggregation that creates multiple regression trees without pruning by using random (repeated) training data samples and summing their averages.

Gradient Boosting is a synthesis process that draws inspiration from gradient descent and builds an enhanced prediction tool using boosting techniques [27]. Boosting starts with creating a tree to determine the link between the input and output variables, and then more trees are created to minimize mistakes. According to GB, boosting is an optimization problem that aims to reduce error using a loss function.

The Gradient Boosting method created by Friedman et al. in 2000 has been updated to become Extreme Gradient Boosting [28]. The fundamental idea behind the GB technique is to sequentially integrate weak (i.e., high error) essential learning trees into a more robust learning model tree. The regularization component, which is introduced to the loss function in the XGB model, is used to assess the complexity of the model in order to enhance the performance of the GB model. It is feasible to harmonize the learning model's parameters and avoid overfitting by incorporating a regularization component. This algorithm's primary challenge is to maximize the value of the goal function. Additionally, it uses a gradient-enhanced framework to construct ML algorithms. Consequently, XGB parallel boost trees can swiftly and accurately address a variety of data science applications.

3.2. Performance evaluation of machine learning model

Many well-known and widely-used statistical metrics, including *RMSE* and R^2, are employed to assess the prediction ability of the ML models. According to Souza et al. [29], using two performance indices including R^2 and *RMSE* are effective enough to evaluate the performance of ML models in the regression problem. R^2 values range from 0 to 1, which is a crucial criterion in regression analysis. It is calculated as the square of the correlation between the expected and actual outcomes. As a result, a high R^2 value indicates a strong connection between the predicted and actual values. *RMSE* is an error assessment of the mean squared difference between a ML model's predicted and actual output. Lower values of *RMSE* demonstrate better prediction performance. These metrics' values are written as follows:

$$R^2 = 1 - \left[\frac{\sum_{j=1}^{N} \left(a_j - b_j\right)^2}{\sum_{i=1}^{N} \left(a_j\right)^2} \right] \tag{1}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{j=1}^{N} \left(a_j - b_j\right)^2} \tag{2}$$

where N is the number of samples in the database, a is the output's actual value, and b is the model's anticipated value.

4. Results and Discussion

4.1. Selection of machine learning model

Dividing 70% of the database for the training dataset and 30% for the test set leads to the randomization of the number of samples used in these two sets, according to Nguyen et al. [30]. The Monte Carlo simulation is proposed to validate the ML model performance. Monte Carlo simulation is frequently used to define the randomness of the model's uncertainty [31]. Based on previously created models, this approach chooses any combinations of input variables with a specific probability distribution. Using a Monte Carlo simulation, the uncertainty of the input variables is transferred to the outcomes. Following the aforementioned findings, 1000 Monte Carlo simulations are carried out (with 1000 different combinations of input variables for each simulation) to display the distributions of R^2 and *RMSE* on the training set and test set for the compressive strength of steel fiber concrete.

The results of the convergence analysis of the R^2 value for the testing dataset after 1000 simulations are shown in Fig. 2. It can be seen that except for the ANN model, the remaining five models, including KNN, CatB, RF, GB, and XGB, all have R^2 value of the testing dataset converges after about 150 simulations, the ANN model gives the converged R^2 value after about 700 simulations. Therefore, it is reasonable to choose to use 1000 simulations of Monte Carlo to validate the model performance.

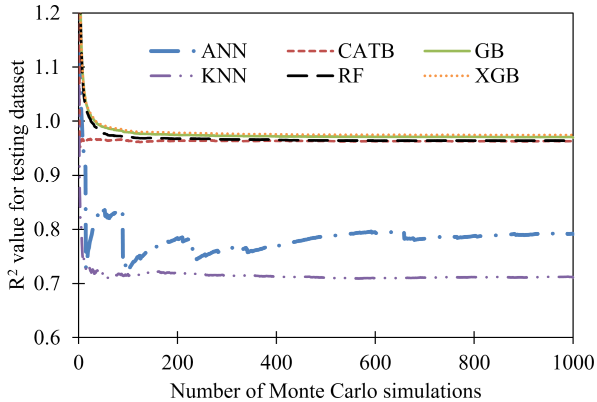


Figure 2. Convergence analysis of R^2 value for the testing dataset

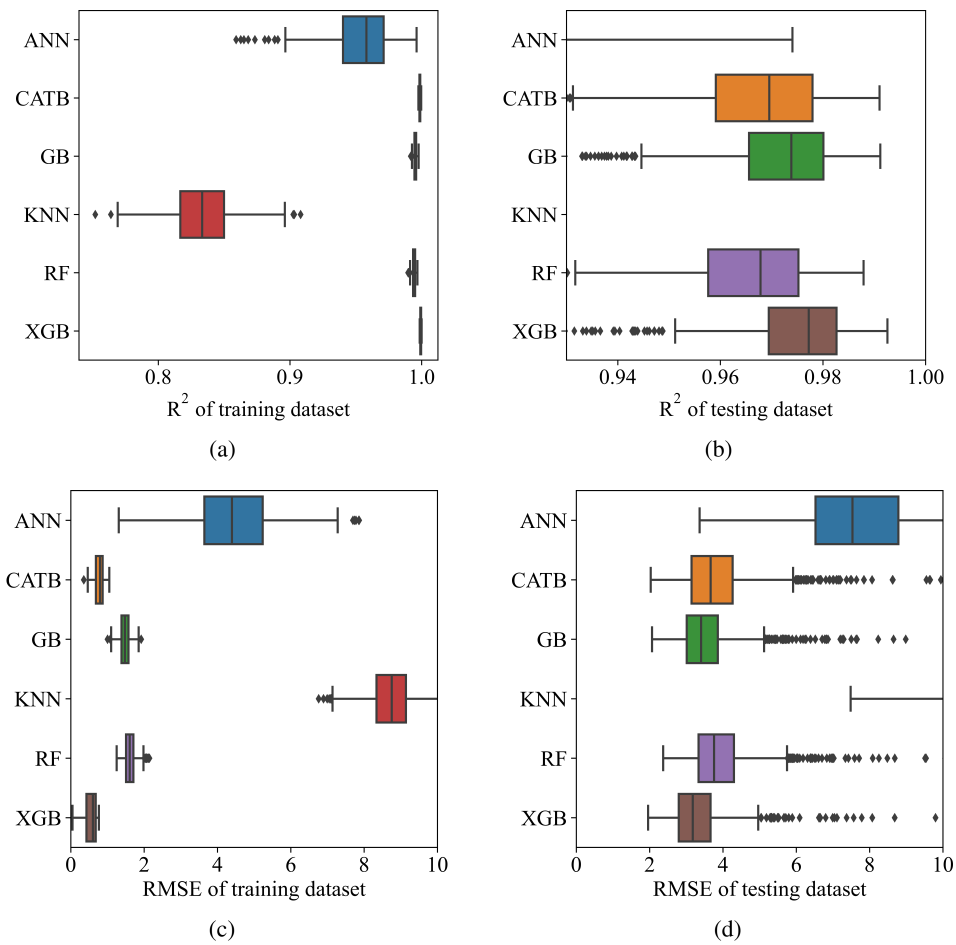


Figure 3. Boxplot of performance values for six ML models

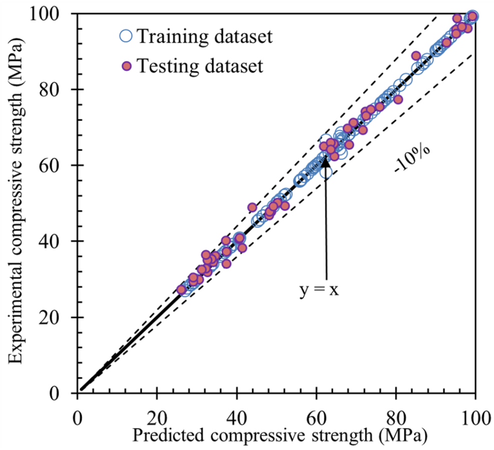
Fig. 3 shows the boxplot of 1000 R^2 and *RMSE* values for the training and testing datasets. Table 2 collects statistical values of 1000 R^2 and *RMSE* values for the training and testing datasets, including mean and StD (standard deviation) values. The results show that the ANN and KNN models have R^2 and RMSE values that vary widely for the training set, while the XGB, GB, RF, and CatB models for R^2 and *RMSE* values have high convergence through the StD value in Table 2. The dispersion of R^2 and *RMSE* values of ANN is 0.7911, 8.4727 MPa for the testing set and KNN is 0.0706 and 1.4293 MPa, respectively. The mean R^2 and *RMSE* of the two models ANN and KNN are 0.7911 and 8.4727 MPa, and 0.7115 and 11.2725 MPa for the testing dataset, respectively. ANN and KNN are the two models with the lowest performance among the six proposed ML models. Using the testing dataset, comparing two performance metrics, R^2 and *RMSE*, including mean and StD values of six ML models, the performance of the ML model can be sorted in descending order as follows XGB > GB > CatB > RF > ANN > KNN. In which the two highest performing models have the mean values of R^2 and *RMSE* after 1000 Monte Carlo simulations for the test data set as XGB with R^2 = 0.9737 on average and *RMSE* = 3.3325 MPa on average, GB with mean of R^2 = 0.9697 on average and *RMSE* = 3.5645 MPa on average. Therefore, two ML models, XGB and GB, will be selected to predict the compressive strength and quantify the influence of factors on the compressive strength of steel fiber concrete.

Table 2. Statistical values of two performance metrics (R^2 and *RMSE*)

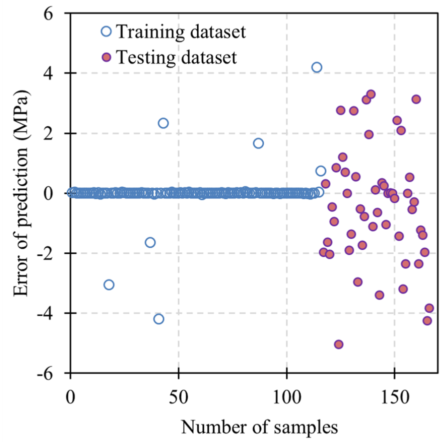
Model	Mean value of R^2		Mean value of *RMSE* (MPa)	
	Training dataset	Testing dataset	Training dataset	Testing dataset
XGB	0.9993	0.9734	0.5481	3.3325
GB	0.9952	0.9697	1.4776	3.5645
CatB	0.9987	0.9631	0.7751	3.9017
RF	0.9943	0.9629	1.6145	3.9573
ANN	0.9545	0.7911	4.4387	8.4727
KNN	0.8330	0.7115	8.7244	11.2725
	StD of R^2		StD of *RMSE* (MPa)	
XGB	0.0004	0.0159	0.1618	0.8286
GB	0.0009	0.0175	0.1380	0.8548
RF	0.0012	0.0209	0.1462	0.9384
CatB	0.0004	0.0256	0.1147	1.1850
KNN	0.0241	0.0706	0.5872	1.4293
ANN	0.0212	0.5575	1.0657	4.6208

4.2. Compressive strength prediction using the best machine learning models

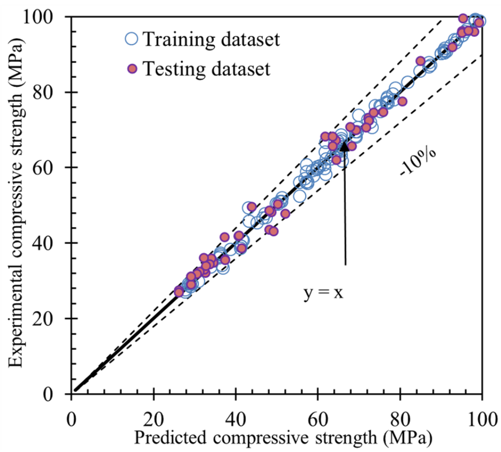
In this section, the accuracy of predicting the compressive strength of steel fiber concrete by two ML models, XGB and GB, will be described in Fig. 4 and Table 3. Figs. 4(a) and 4(c) compare the predicted and actual compressive strength values using the XGB and GB models. The prediction errors of the two models are also shown in Figs. 4(b) and 4(d). In Figures 4(a) and 4(c), the closer the data points are to the $y = x$ line, the better the predictive power of the model. It can be seen that the XGB model gives outstanding compressive strength prediction results when the data points of the testing dataset are mostly on the $y = x$ line and in the ±10% error ($y = 0.9x$ and $y = 1.1x$) limit domain, the corresponding error of compressive strength prediction for the test set is also in the region ±4 MPa, the error of prediction for the training set is deficient. Meanwhile, the GB model also predicts very well the compressive strength, but it is still lower than the XGB model when there are some data points outside the ±10% domain and the out-of-domain error ±4 MPa for the testing dataset.



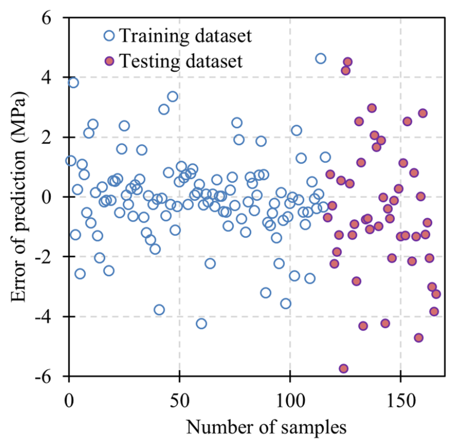
(a) XGB model



(b) Error prediction of XGB model



(c) GB model



(d) Error prediction of GB model

Figure 4. Comparison between experimental and predicted compressive strength of steel fiber concrete, XGB model and GB model

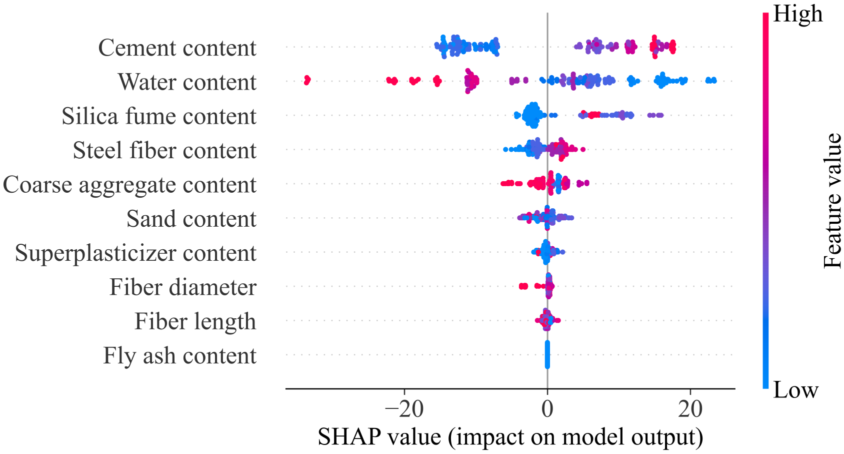
Table 3. Performance value of XGB and GB model in predicting compressive strength of steel fiber concrete

Dataset	Model	R^2	*RMSE* (MPa)
Training	XGB	0.9988	0.6956
	GB	0.9943	1.5234
Testing	XGB	0.9926	1.9814
	GB	0.9874	2.5763

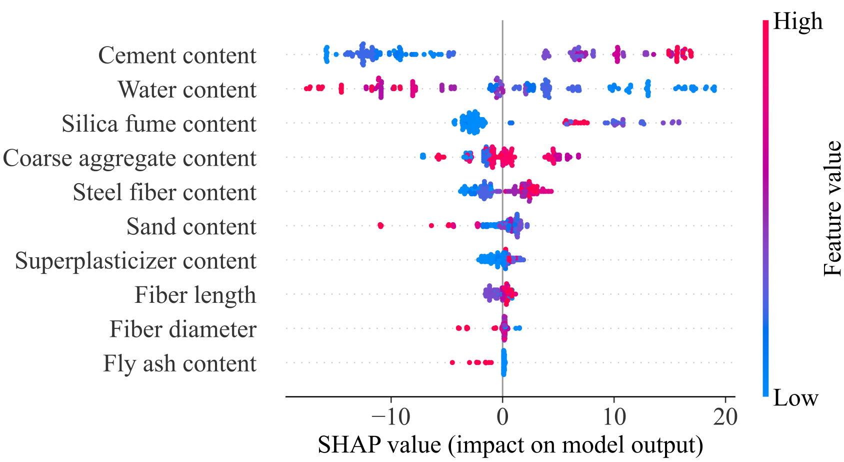
The performance of predicting compressive strength of the XGB model compared with the GB model is specified and quantified through Table 3, describing the R^2 and *RMSE* values of XGB for the test set as 0.9926 and 1.9814 MPa, respectively. The GB model has R^2 = 0.9874 and $RMSE$ = 2.5763 MPa values for the testing dataset.

4.3. Investigating of factor effect on compressive strength of concrete

Regarding the two highest-performing ML models, XGB and GB, the ten inputs were quantitatively analyzed using the SHAP algorithm. The global SHAP values depicted in Fig. 5 help to relatively quantify the influence of each factor on the accuracy of the ML model as well as the magnitude of the predicted compressive strength. The simultaneous use of two ML models to quantify influencing factors helps confirm the correctness of the results.



(a) XGB model



(b) GB model

Figure 5. Global SHAP value for interpreting input effect on the predicted value consisting of prediction capability and the magnitude value

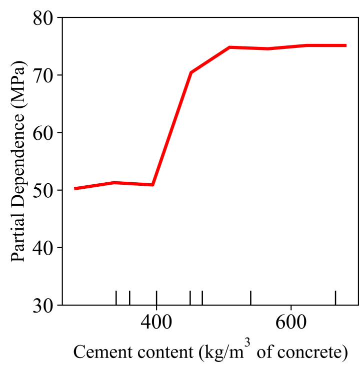
In Fig. 5, the results of the global SHAP value based on XGB and GB models confirm that the five factors that have the least influence on the performance of the two ML models, as well as the compressive strength are Fly ash content < Fiber length < Fiber diameter < Superplasticizer content < Sand content. The importance of Fiber length and Fiber diameter can be interchanged, but the difference is insignificant, as shown in Fig. 5(b). On the other hand, the five factors ranked in order Cement content > Water content > Silica fume content > Steel fiber content > Coarse aggregate content have a great influence on the prediction accuracy of compressive strength of the XGB and GB models, as well as the value compressive strength of steel fiber concrete.

Therefore, the order of influence of factors on compressive strength can be arranged in descending order Cement content > Water content > Silica fume content > Steel fiber content > Coarse aggregate content > Sand content > Superplasticizer content > Fiber diameter > Fiber length > Fly ash content.

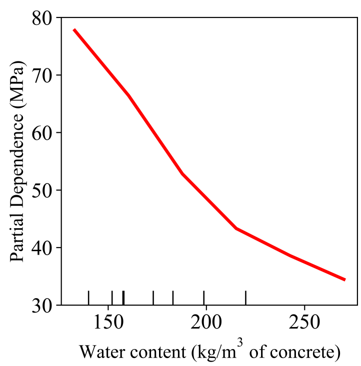
The four most important factors including Cement content > Water content > Silica fume content > Steel fiber content, strongly influence on the compressive strength of steel fiber concrete. In the remaining six factors, although Coarse aggregate content and sand content have an influence in the range of 8 MPa on the compressive strength, however, these two factors tend to have an insignificant influence when the SHAP values of these two factors are identical at different values of these two factors, as well as the same value of the two inputs, multiple SHAP values can be determined. This may be due to the input data distribution domain of Coarse aggregate content and Sand content. Especially, coarse aggregate content, it varies between 0 and 1170 kg/m^3, the median value is 1050.5 kg/m^3, and $Q_{75\%}$ = 1100 kg/m^3. It can be seen that the value of coarse aggregate content is mainly concentrated at 1000 kg/m^3. The uneven distribution of data makes the assessment of the trend of the influence of the input variable on the compressive strength unclear.

The first four factors tend to have an evident influence on compressive strength, these four factors will be quantified regarding their influence on compressive strength using the 1D Partial Dependence Plot technique, shown in Fig. 6. For instance, increasing Cement content increases compressive strength, but its effect remains the same when it reaches about 500 kg/m^3, as indicated in Fig. 6(a). The domain of influence of cement on compressive strength can be up to 25 MPa.

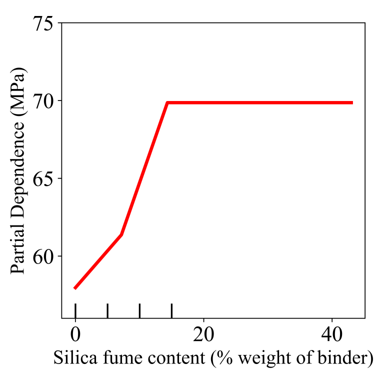
Water content varying from about 133 kg/m^3 to 270 kg/m^3 can cause the intensity to decrease from a value of 44 MPa (79 MPa to 35 MPa), a relatively linear decrease trend (Fig. 6(b)). The



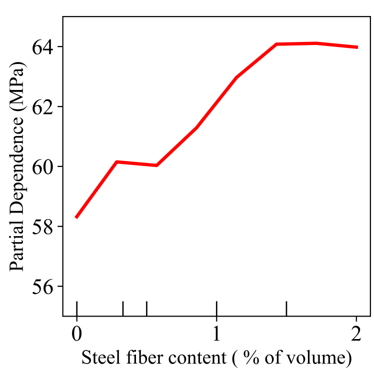
(a)



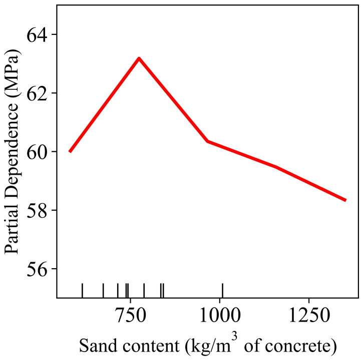
(b)



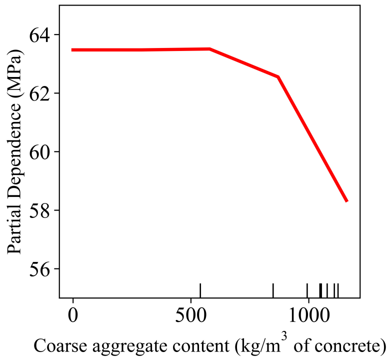
(c)



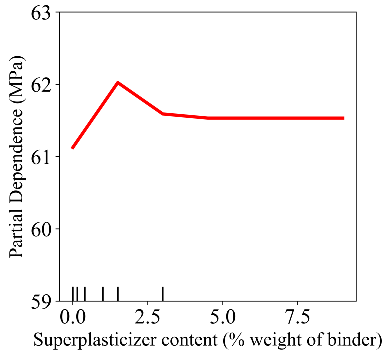
(d)



(e)



(f)



(g)

Figure 6. Partial dependence plot in analyzing the most important factors on compressive strength of steel fiber concrete

increase in compressive strength shows a significant variation in the range of 12 MPa when changing the amount of silica fume used from 0 to 15%. After this interval, the compressive strength did not increase despite an increase in the silica fume content to 40% weight of binder (Fig. 6(c)). Steel fiber has a positive effect on compressive strength. Fig. 6(d) shows that as the steel fiber content increases from 0 to 1.5% of volume, the compressive strength can increase by 6 MPa. However, the compressive strength does not continue to increase even though the steel fiber is greater than 1.5% to 2% of volume. This effect of steel fiber content on compressive strength is consistent with the results found by Köksal et al. [21] and Nili et al. [32].

It can be seen that in addition to the above four factors, Sand content, Superplasticizer content, and Coarse aggregate content have a relatively small influence on the variation of compressive strength. In particular, Superplasticizer does not seem to affect the compressive strength of steel fiber concrete, as shown in Fig. 6(g). The variation value of compressive strength caused by Sand content and Coarse aggregate content is about 5 MPa, as shown in Figs. 6(e) and 6(f). Therefore, increasing Sand content and Coarse aggregate content to a certain value will reduce the compressive strength of steel fiber concrete.

5. Conclusions and Perspectives

A data-driven approach is proposed to study the compressive strength of steel fiber concrete. In this study, six popular ML models were evaluated to select the ML models with good performance for predicting and studying the factors affecting the compressive strength of steel fiber concrete. Based on a database of 166 samples and ten input variables, including Cement content, Water content, Silica fume content, Steel fiber content, Coarse aggregate content, Sand content, Superplasticizer content, Fiber diameter, Fiber length, Fly ash content, six ML models including ANN, KNN, RF, CatB, GB, and XGB are evaluated against performance metrics including R^2 and *RMSE* and validated by 1000 Monte Carlo simulations.

Based on the average values of R^2 and *RMSE* for the testing dataset, the compressive strength prediction performance of the ML models is arranged in descending order as follows XGB > GB > CatB > RF > ANN > KNN, especially $R^2 = 0.9734$ on average and $RMSE = 3.3325$ MPa on average for XGB model. The two best models can predict the compressive strength of steel fiber concrete to be GB ($R^2 = 0.9874$ and $RMSE = 2.5763$ MPa) and XGB ($R^2 = 0.9926$ and $RMSE = 1.9814$ MPa), respectively.

Nonlinear analysis techniques, including SHAP and Partial Dependence Plot 1D, based on two high-performance models, GB and XGB, were used to analyze and quantify the influence of ten inputs on the precision value of the compressive strength of steel fiber concrete. The influence of the ten variables can be arranged in descending order as follows: Cement content > Water content > Silica fume content > Steel fiber content > Coarse aggregate content > Sand content > Superplasticizer content > Fiber diameter > Fiber length > Fly ash content. Cement content, Silica fume content, Steel fiber content, and Water content have an evident influence on the compressive strength of concrete. The three first factors positively improve the compressive strength of concrete, while the increased last factor causes the compressive strength to decrease. The geometric parameters of steel fiber (i.e., diameter and length) have a minimal influence on the compressive strength of steel fiber concrete. The effect of input variables on compressive strength was quantified. For instance, the steel fiber content used should be less than 1.5% of concrete volume to improve the efficiency of steel fiber in concrete.

This study helps engineers have an initial assessment when designing the strength of steel fiber concrete. In addition, the model selected could be used for the input value domain mentioned in the present study.

References

[1] ACI 544.1 R-96 (2009). *State-of-the-Art Report on Fiber Reinforced Concrete*. American Concrete Institute, Farmington Hills, MI, USA.

[2] Nguyen, N. T., Bui, T.-T., Bui, Q.-B. (2022). Fiber reinforced concrete for slabs without steel rebar reinforcement: Assessing the feasibility for 3D-printed individual houses. *Case Studies in Construction Materials*, 16:e00950.

[3] Thao, N. T. T., Tung, T. P. S., Nhan, N. D., Tan, N. N. (2022). Assessing the shear behavior of steel fiber reinforced concrete beams corroded under chloride attacks. *Journal of Science and Technology in Civil Engineering (STCE) - HUCE*, 16(3):97–110.

[4] Tan, N. N., Thao, N. T. T., Van, N. T., Hieu, D. D. (2022). Assessing the shear behavior of corroded steel fiber reinforced concrete beams without shear reinforcement using nonlinear finite element analysis. *Journal of Science and Technology in Civil Engineering (STCE) - HUCE*, 16(3):152–165.

[5] Song, P. S., Hwang, S. (2004). Mechanical properties of high-strength steel fiber-reinforced concrete. *Construction and Building Materials*, 18(9):669–673.

[6] Quan, V., Do, H. Q. (2021). Prediction of California Bearing Ratio (CBR) of Stabilized Expansive Soils with Agricultural and Industrial Waste Using Light Gradient Boosting Machine. *Journal of Science and Transport Technology*, 1–9.

[7] Tran, V. Q., Nguyen, L. Q. (2022). Using machine learning technique for designing reinforced lightweight soil. *Journal of Intelligent & Fuzzy Systems*, 43(1):1633–1650.

[8] Tran, V. Q. (2021). Compressive Strength Prediction of Stabilized Dredged Sediments Using Artificial Neural Network. *Advances in Civil Engineering*, 2021:1–8.

[9] Tran, V. Q. (2022). Hybrid gradient boosting with meta-heuristic algorithms prediction of unconfined compressive strength of stabilized soil based on initial soil properties, mix design and effective compaction. *Journal of Cleaner Production*, 355:131683.

[10] Tran, V. Q., Dang, V. Q., Ho, L. S. (2022). Evaluating compressive strength of concrete made with recycled concrete aggregates using machine learning approach. *Construction and Building Materials*, 323:126578.

[11] Tran, V. Q. (2022). Machine learning approach for investigating chloride diffusion coefficient of concrete containing supplementary cementitious materials. *Construction and Building Materials*, 328:127103.

[12] Tran, V. Q. (2023). Using Artificial Intelligence Approach for Investigating and Predicting Yield Stress of Cemented Paste Backfill. *Sustainability*, 15(4):2892.

[13] Tran, V. Q. (2022). Predicting and Investigating the Permeability Coefficient of Soil with Aided Single Machine Learning Algorithm. *Complexity*, 2022:1–18.

[14] Li, Y., Zhang, Q., Kamiński, P., Deifalla, A. F., Sufian, M., Dyczko, A., Kahla, N. B., Atig, M. (2022). Compressive Strength of Steel Fiber-Reinforced Concrete Employing Supervised Machine Learning Techniques. *Materials*, 15(12):4209.

[15] Khan, K., Ahmad, W., Amin, M. N., Ahmad, A., Nazar, S., Alabdullah, A. A. (2022). Compressive Strength Estimation of Steel-Fiber-Reinforced Concrete and Raw Material Interactions Using Advanced Algorithms. *Polymers*, 14(15):3065.

[16] Abbass, W., Khan, M. I., Mourad, S. (2018). Evaluation of mechanical properties of steel fiber reinforced concrete with different strengths of concrete. *Construction and Building Materials*, 168:556–569.

[17] Kim, J. J., Kim, D. J., Kang, S. T., Lee, J. H. (2012). Influence of sand to coarse aggregate ratio on the interfacial bond strength of steel fibers in concrete for nuclear power plant. *Nuclear Engineering and Design*, 252:1–10.

[18] Aruntaş, H. Y., Cemalgil, S., Şimşek, O., Durmuş, G., Erdal, M. (2008). Effects of super plasticizer and curing conditions on properties of concrete with and without fiber. *Materials Letters*, 62(19):3441–3443.

[19] Nili, M., Afroughsabet, V. (2012). Property assessment of steel–fibre reinforced concrete made with silica fume. *Construction and Building Materials*, 28(1):664–669.

[20] Challoob, M. A., Srivastava, V. (2013). Effect of fly ash and steel fibre on Portland pozzolana cement concrete. *International Journal of Engineering Trends and Technology*, 5(3):144–147.

[21] Köksal, F., Altun, F., Yiğit, İ., Şahin, Y. (2008). Combined effect of silica fume and steel fiber on the mechanical properties of high strength concretes. *Construction and Building Materials*, 22(8):1874–1880.

[22] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, 12:2825–2830.

[23] Tamura, S., Tateishi, M. (1997). Capabilities of a four-layered feedforward neural network: four layers versus three. *IEEE Transactions on Neural Networks*, 8(2):251–255.

[24] Altman, N. S. (1992). An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. *The American Statistician*, 46(3):175–185.

[25] Dorogush, A. V., Ershov, V., Gulin, A. (2018). CatBoost: gradient boosting with categorical features support. *arXiv preprint arXiv:1810.11363*.

[26] Breiman, L. (2001). *Machine Learning*, 45(1):5–32.

[27] Ayyadevara, V. K. (2018). Gradient Boosting Machine. In *Pro Machine Learning Algorithms*, Apress, 117–134.

[28] Chen, T., Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM.

[29] de Souza, C. M. P., Silva, L. A. P., Veloso, G. V., Leite, M. E., Filho, E. I. F. (2022). Environmental fragility by machine learning algorithms. *Mercator*, 21(2):1–16.

[30] Nguyen, Q. H., Ly, H.-B., Ho, L. S., Al-Ansari, N., Le, H. V., Tran, V. Q., Prakash, I., Pham, B. T. (2021). Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil. *Mathematical Problems in Engineering*, 2021:1–15.

[31] Soize, C. (2000). A nonparametric model of random uncertainties for reduced matrix models in structural dynamics. *Probabilistic Engineering Mechanics*, 15(3):277–294.

[32] Nili, M., Azarioon, A., Danesh, A., Deihimi, A. (2016). Experimental study and modeling of fiber volume effects on frost resistance of fiber reinforced concrete. *International Journal of Civil Engineering*, 16(3): 263–272.