Comparative analysis of various machine learning algorithms to predict 28-day compressive strength of Self-compacting concrete

Waleed Bin Inqiad [a],[*], Muhammad Shahid Siddique [a],[**], Saad S. Alarifi [b],
Muhammad Jamal Butt [c], Taoufik Najeh [d], Yaser Gamil [e]

[a] *Military College of Engineering (MCE), National University of Science and Technology (NUST), Islamabad 44000, Pakistan*
[b] *Department of Geology and Geophysics, College of Science, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia*
[c] *COMSATS University Islamabad, Abbottabad Campus, Pakistan*
[d] *Operation and Maintenance, Operation, Maintenance and Acoustics, Department of Civil, Environmental and Natural Resources Engineering, Lulea University of Technology, Sweden*
[e] *Department of Civil Engineering, School of Engineering, Monash University Malaysia, Jalan Lagoon Selatan, 47500 Bandar Sunway, Selangor, Malaysia*

ARTICLE INFO

ABSTRACT

Construction industry is indirectly the largest source of CO_2 emissions in the atmosphere, due to the use of cement in concrete. These emissions can be reduced by using industrial waste materials in place of cement. Self-Compacting Concrete (SCC) is a promising material to enhance the use of industrial wastes in concrete. However, there are very few methods available for accurate prediction of its strength, therefore, reliable models for estimating 28-day Compressive Strength (C–S) of SCC are developed in current study by using three Machine Learning (ML) algorithms including Multi Expression Programming (MEP), Extreme Gradient Boosting (XGB), and Random Forest (RF). The ML models were meticulously developed using a dataset of 231 points collected from internationally published literature considering seven most influential parameters including cement content, quantities of fly ash and silica fume, water content, coarse aggregate, fine aggregate, and superplasticizer dosage to predict C–S. The developed models were evaluated using different statistical errors including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), coefficient of determination (R^2) etc. The results showed that the XGB model outperformed the MEP and RF model in terms of accuracy with a correlation $R^2 = 0.998$ compared to 0.923 for MEP and 0.986 for RF. Similar trend was observed for other error metrices. Thus, XGB is the most accurate model for estimating C–S of SCC. However, it is pertinent to mention here that it does not give its output in the form of an empirical equation like MEP model. The construction of these empirical models will help to efficiently estimate C–S of SCC for practical purposes.

1. Introduction

The surge in greenhouse gas emissions such as CO_2 pose a significant environmental threat to the mankind. Most of the emissions

are due to human activities. Construction industry is a substantial emitter of CO_2 contributing almost 50 % to the global emissions [1]. Concrete is the intensely used building material due to its adaptability and ease of usage. The annual production of concrete is more than 25 billion tons [2]. Traditional concrete uses Portland cement as the main binding agent and it is the leading contributor of CO_2 emissions from concrete [3]. The manufacturing of cement accounts for 7 % of the global CO_2 emissions [4]. The yearly production of cement is about 4000 million tons, and it is set to increase to 6000 million tons in 2060 [5]. The emission of these harmful gases into the atmosphere brings about climate change. Thus, the invention and production of green construction materials is desirable to lower the annual carbon emissions and to conserve precious natural resources.

The concrete industry has undergone significant advancements in recent years. One notable development is the production of green concrete by using industrial waste materials in concrete. The process of cement manufacture utilizes large amount of energy and contributes significantly to greenhouse gas emissions. To reduce impact of cement production on environment, industrial wastes like limestone, fly ash, silica fume etc. have been used [6]. The use of these industrial wastes as Secondary Cementitious Materials (SCMs) in concrete resulted in a decrease in cement utilization across the globe and reduced the dangerous effects of cement production on the environment [7]. The use of SCMs has been seen by researchers as an effective way to reduce the emissions from the construction industry [8,9]. These waste materials also have other applications. For instance, metakaolin, fly ash, silica fume etc. are frequently used in soilcrete mixtures to modify their properties [10–12]. The use of SCMs in concrete not only reduces the carbon emissions from construction industry but also reduces the stress on depleting natural resources by meeting the growing demand for concrete in the modern world [13].

The increase in use of SCMs in concrete designing technology have resulted in the invention of SCC [14]. It is a concrete having the property to flow and compact itself under the influence of its own weight, without needing external compaction. This property of SCC makes it a preferable choice for use in complex forms or confined spaces, where traditional methods of vibration are not feasible or desirable, such as in the presence of delicate rebar structures. Additionally, SCC offers several other benefits, including improved surface finish, higher resistance to segregation and bleeding, and increased strength and durability as shown in Fig. 1. Different studies have explored the subject of using SCMs in SCC production [15–19]. The use of SCC containing industrial wastes results in cost savings, as it reduces the time required for concrete placement and finishing. Therefore, SCC has the ability to make the concrete industry more sustainable and environment friendly and improve the quality, efficiency, and safety of construction projects [20].

The mixture composition of SCC is designed to achieve the desired flowability and Self-compacting characteristics. One key component of SCC is to use high levels of fines, such as sand and other admixtures like fly ash, silica fume etc, which help to fill in the spaces between the coarse aggregate particles. These fines are typically mixed with a superplasticizer, which is a type of chemical admixture that helps to increase the flowability of the mixture. In general, SCC has a higher water-to-cement ratio and a lower coarse aggregate content, which helps to achieve the required flowability and Self-compacting characteristics [21]. SCC is widely utilized in construction practices, but the literature contains few reliable methods that can predict its compressive strength based on its mixture composition. It is because C–S of SCC exhibits a non-linear relation with the concrete mixture composition [22]. The SCC requires an appropriate mix design process to achieve its desirable properties. According to Boukendakdji [17], changes in cement or mineral admixtures, as well as changes in aggregate type, can result in significant variations in the properties of SCC. Another reason for the lack of reliable prediction methods is the lack of familiarity and expertise in the use of SCC. Thus, it is very important to have a robust quantitative method, which can accurately predict C–S of SCC. There are physical tests available to check C–S of concrete by casting

Fig. 1. Benefits of SCC.

cubic or cylindrical specimens, but these methods are less economical, ineffective and time consuming. Also, these methods contribute to the construction waste as the specimens are destroyed and thrown after conduction the test. Thus, it is inevitable to have a fast, convenient yet accurate method to estimate C–S of SCC.

Recently, the estimation of different properties of concrete having industrial wastes using machine learning (ML) algorithms have captured the attention of researchers as an efficient way to enhance the use of SCMs in concrete as an alternative of cement. ML techniques have an advantage that they are capable of capturing the underlying mechanisms, even in situations where there is a lack of information regarding specific parameters. This advantage stems from the ability of ML algorithms to learn patterns and associations from data, which can help to reveal the hidden relationships between variables [23]. By using this ability, ML techniques can effectively extract meaningful insights and make accurate predictions, even in complex situations. Due to these advantages, ML techniques have found extensive use in civil engineering to predict different concrete properties, soil properties, soil classification, slope stability and process control etc [24–34]. For instance, Asteris et al. [35], utilized artificial neural network (ANN) to predict different mechanical properties of sandcrete mixtures. Similarly, Sarrir et al. [36], utilized GEP and ANN to predict bearing capacity of steel tube columns filled with concrete.

Regarding the C–S prediction of SCC, Farooq et al. [37] used GEP alongside ANN and support vector machine (SVM) to predict C–S of SCC containing fly ash as an admixture. The best performing algorithm was selected having the least average error value. The results revealed that GEP is the most accurate algorithm having the least error value of 3.714 MPa compared to 5.428 MPa and 5.023 for ANN and SVM respectively. The study revealed an added advantage of GEP algorithm that it gives its output in the form of an empirical equation for C–S prediction unlike ANN and SVM. However, ANN gives the higher R value of 0.9588 compared to 0.9344 and 0.935 of SVM and GEP respectively. Similarly, Asteris et al. [38] leveraged ANN to estimate C–S of SCC containing a various admixture including limestone powder, fly ash, slag, and silica fume. The author used 169 data points and the results showed that ANN can be efficiently used to predict SCC strength. The coefficient of correlation for the developed ANN model was 0.9825 showing the excellent predicting capabilities of neural networks. Similarly, another study [39] utilized three algorithms: decision trees (DT), light gradient boosting machine (LGBM), and XGB to predict C–S of fibre-reinforced SCC having admixtures like fly ash, nano silica, marble powder etc. Fiber-reinforced SCC is a promising material to be used in place of normal concrete because of it offers many advantages such as crack resistance, cost savings and high strength etc. The authors revealed that XGB surpassed all other algorithms and proved to be the most accurate one with $R^2 = 0.992$ and mean absolute error of 1.438 MPa. Thus, the developed XGB model can be effectively used to foster the use of SCC reinforced with fibres in the construction industry.

There are several other studies attributed to predicting SCC strength containing recycled aggregates using different boosting algorithms. For example, a study [40] utilized random forest (RF), k-nearest neighbor (KNN), XGB, gradient boosting (GB), category boosting (CB), and extremely randomized trees (ERT), to predict the C–S of SCC having recycled aggregates. The study applied the algorithms on 515 data points from the literature and this study confirmed that RF technique is the most robust to predict C–S with testing $R^2 = 0.7128$. Furthermore, there are several studies that undertook the prediction of different SCC properties such as flexural

Table 1
Summary of SCC properties prediction by different researchers.

S. No	Machine learning method	Data set	Output	Year	Admixtures	References
1.	DT, XGB, LGBM	387	C–S	2023	Fly ash, Limestone, Basalt powder, Marble powder, Nano silica, Metakaolin	[39]
2.	ANN	111	C–S	2011	Fly ash, Bottom ash	[43]
3.	Ridge Regression, DT, Linear Regression, Lasso Regression, RF, Multi-layer Perceptron Regression	99	C–S	2022	Silica flour, Polypropylene Fiber	[44]
4.	CB, KNN, XGB, LGBM, Inverse Gaussian, ERT, Poisson Gaussian, RF, GB	515	C–S	2022	Fly ash	[40]
5.	M5′ algorithm, Multivariate adaptive regression splines (MARS)	114	C–S, V-funnel time, L-box ratio, Slump flow	2018	Fly ash	[45]
6.	SVR	327	C–S	2023	Fly ash	[46]
7.	GEP	26	C–S, Slump flow, J Ring	2009	Fly ash	[47]
8.	Multivariate Regression (MVR)	63	C–S, Modulus of Elasticity, Flexural Strength	2020	Silica fume with crumb rubber	[42]
9.	ANN	169	C–S	2016	Limestone powder, Fly ash, Slag, Silica fume, Rice husk ash	[38]
10.	ANN	114	C–S, Slump flow, V-funnel time L-box ratio,	2016	Fly ash	[48]
11.	SVR	115	Slump flow, L-box ratio, C–S, V-funnel time,	2019	Fly ash	[41]
12.	Extreme Learning Machine, long short-term memory (LSTM)	48	Slump flow, J Ring	2021	Slag, Fly ash, Limestone powder, Silica fume	[49]
13.	ANN, SVM	85	C–S	2022	Fly ash, Silica fume	[50]
14.	RF, Multilayer perceptron network (MLP), KNN	1030	C–S	2022	Fly ash, Slag	[51]
15.	SVM, ANN, GEP	300	C–S	2021	Fly ash	[37]
16.	SVR, Deep Learning (DL)	24	C–S, Splitting tensile strength	2021	Fly ash, Silica fume	[52]

strength, slump flow etc. Saha et al. [41] leveraged the support vector regression (SVR) to predict different properties of SCC. The study utilized 115 data points having fly ash as a mineral admixture to predict the L-box ratio, slump flow, V-funnel time, and C–S. The value of coefficient of correlation achieved for slump flow is 0.965 followed by 0.954 for L-box ratio, 0.979 for funnel time and 0.977 for strength indicating the robustness of these techniques to predict different properties of SCC. Apart from the mineral admixtures, several other industrial wastes such as crumb rubber etc have been used in SCC to impart special properties to SCC. Robert et al. [42] incorporated silica fume and rubber waste in SCC and used multivariate regression models to predict flexural strength, C–S, and modulus of elasticity. The study used crumb rubber of size ranging from 0.5 to 3.5 mm and replaced sand by crumb rubber by different percentages such as 5 %, 10 %, 20 % and 30 % of the aggregate volume. The model was created on a set of 63 instances and achieved R^2 value of 0.931 for C–S, 0.922 for modulus of elasticity and 0.937 for flexural strength. The summary of prediction of different properties of SCC using various ML algorithms is shown in Table 1.

The significance of this study is prediction of C–S of SCC incorporating silica fume and fly ash as mineral admixtures by using newest machine learning algorithms in a comparative manner. For this purpose, three machine learning algorithms including MEP, RF and extreme gradient XGB were used. Particularly, the novelty of this research is leveraging MEP to predict the strength of SCC. The accuracy of the developed models was compared using statistical evaluation and the accurate model was selected among the developed models.

2. Machine learning models

2.1. Multi Expression Programming

MEP was introduced by Oltean and Dumitrescu [53]. It involves building and evolving several mathematical expressions to find the
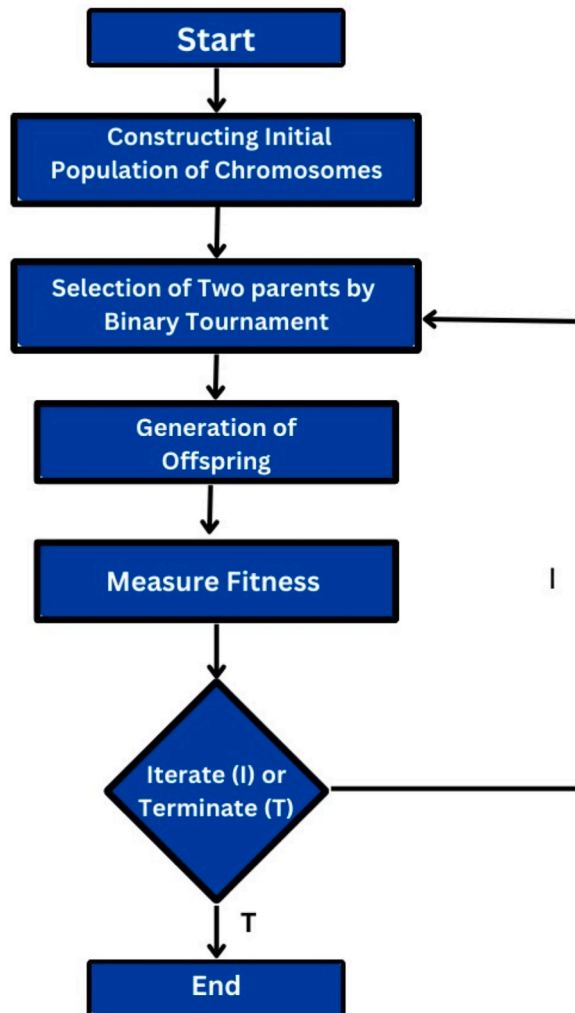

Fig. 2. Flow chart of MEP algorithm.

solution to a problem. MEP encode solutions using linear chromosomes, with a unique ability of encoding various solutions within one chromosome. On the basis of fitness of individuals, the optimal solution is selected to be the representation of the chromosome [54]. Compared to other GP techniques that allow to only encode one solution in one chromosome, MEP has the advantage that it does not rise the complexity of the equation [55]. The MEP techniques starts by generating a population of random individuals. To evolve the optimal solution over specific generations until a termination criterion is reached, MEP employs the following steps [53,55].

- The process of selection of two parents utilizing the binary tournament is done in first step, followed by recombination with a specified probability for crossover.
- Two offsprings are produced by recombining two parents.
- The obtained offspring is mutated, and the worst performing individual is replaced with the best offspring.

MEP is a versatile optimization technique that offers several advantages over other methods, particularly in its ability to handle complex optimization problems without relying on specific mathematical models or assumptions. Many studies suggest that MEP's iterative process of generating and evaluating mathematical expressions, along with its multi-expression chromosome, makes it a useful tool for solving optimization problems [56–59]. The flow chart of MEP algorithm [60] is shown in Fig. 2.

2.2. Random forest (RF)

Leo Breiman and Cutler Adele introduced the RF technique [61], which is utilized for classification and regression problems. The RFR algorithm has several advantages, including its efficiency in handling large data sets, its robustness against overfitting [62], and its lower complexity compared to other machine learning techniques. In geotechnical engineering, RFR is increasingly being employed [63,64]. Additionally, RFR is used in structural engineering to predict the properties of concretes with various admixtures.

RF is composed of two key components: Decision Tree (DT) and Bagging technique. Prior to implementing the algorithm, the feature space is further divided into a number of segments. This segmentation is done iteratively until the stop threshold is met. It results in the formation of three components: internal, external, and branches. The internal nodes are continually connected with functions to choose which node to contact next. DT nodes, which can no longer be split, are referred to as output nodes, terminals, or leaf nodes. The RFR algorithm has proven to be more reliable in various tasks compared to one decision tree [65]. The RF method is an ensemble learning algorithm that utilizes bagging to predict outputs [66]. This algorithm constructs a huge number of decision trees and summarizes their results on the training sample set. If the problem being solved is a problem of classification, the RF algorithm predicts the classification category based on decision trees. On the other hand, if the problem encountered is a regression one, the algorithm calculates the mean of the regression trees to obtain the result. Fig. 3 shows the diagram of the RF technique [67].
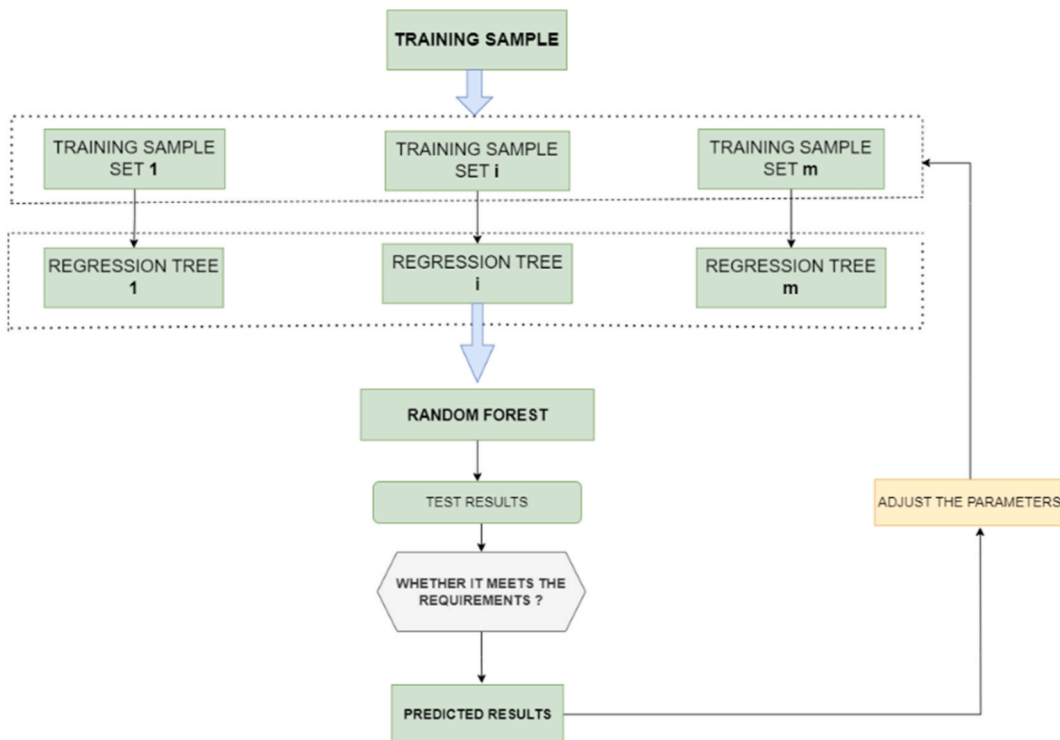

Fig. 3. Flow Chart of RF algorithm.

2.3. Extreme gradient boosting (XGB)

Chen and Guestrin [68] introduced extreme gradient boosting (XGB) in 2016. It is an ensemble ML method for boosting of trees. XGB is useful for both machine learning (ML) and data mining applications. This approach controls overfitting and improve performance, resulting in a decrease in model complexity and a significant reduction in overfitting [69]. By controlling overfitting, XGB can achieve superior results in comparison to other tree boosting algorithms.

It involves the use of a decision tree along with gradient boosting [70]. DT are known for their ability to provide insights into the importance of inputs in calculating the output. However, XGB takes this a step further by providing users with enhanced capabilities to learn the aforementioned insights [71]. It achieves this by iteratively training decision trees using gradient boosting, a technique that involves the optimization of a loss function by minimizing the errors of the previously trained tree. This iterative process enables the model to learn complex relations between inputs and the output, resulting in enhanced accuracy and interpretability. Thus, XGB offers an improved decision tree algorithm that not only provides accurate predictions but also offers a deeper understanding of the underlying data. Furthermore, it has been observed that extreme gradient boosting outperforms gradient boosting (GB) in terms of predictive accuracy, particularly when dealing with vast amounts of data. This can be attributed to XGB's implementation of a regularization and loss function to access the model's fitness [72].

3. Data acquisition and presentation

The most important step in the creation of a ML model is gathering reliable data. To achieve this goal, a thorough literature review
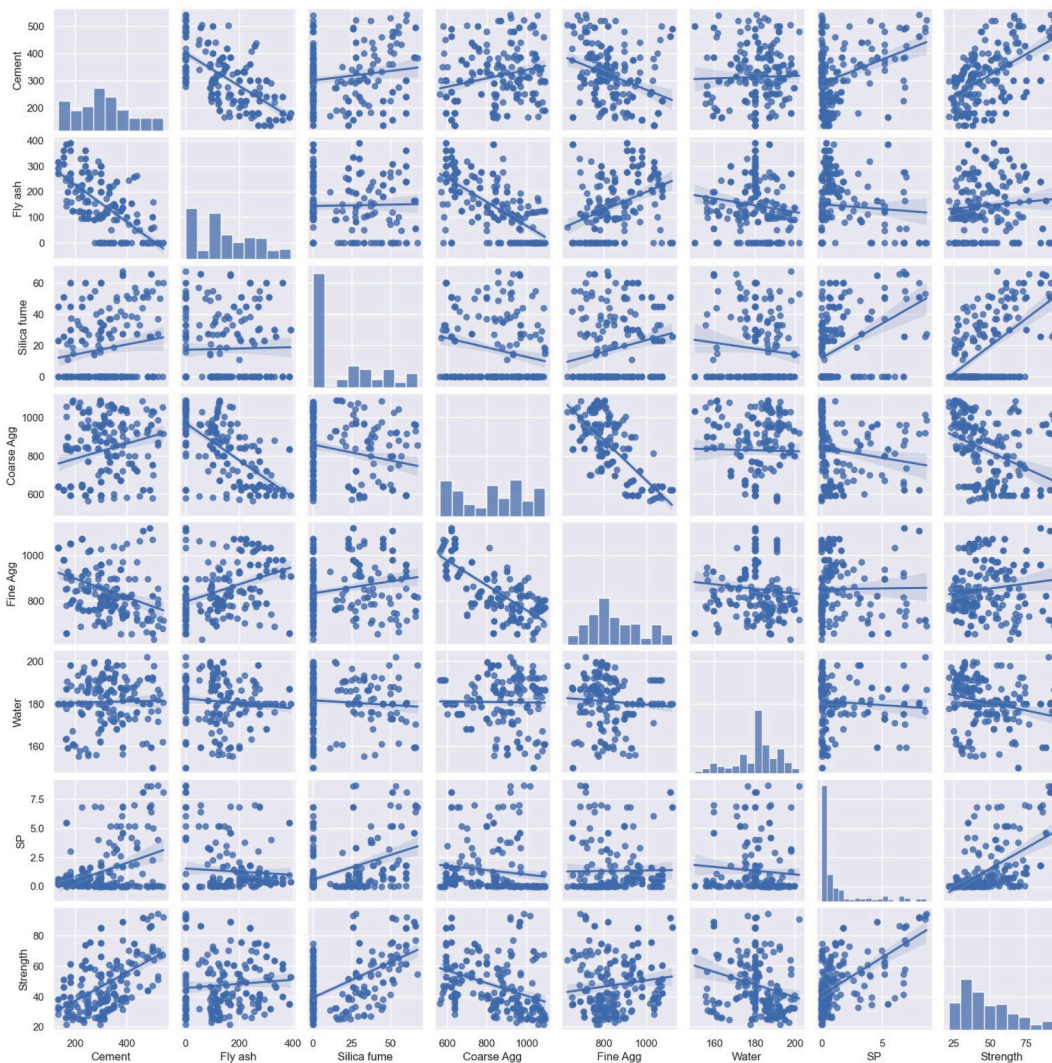

Fig. 4. Scatter matrix of inputs and output variables.

is performed, and a database was compiled from published research [19,21,73–83] containing 231 data points for prediction of C–S. To select the most influential parameters, a thorough evaluation was carried out, and the effectiveness and performance of numerous early trials were calculated. Thus, to forecast C–S of SCC, seven input factors were chosen: cement (kg/m^3), fly ash (kg/m^3), silica fume (kg/m^3), water (kg/m^3), coarse aggregate (kg/m^3), fine aggregate (kg/m^3), and superplasticizer (kg/m^3). These seven input parameters were used to build the prediction models for C–S (MPa).

The variance and covariance of the regressions between the variables can be better understood by using a descriptive statistical technique known as a scatter matrix. It is frequently used alongside other statistical analysis techniques to visualize the data and get useful insights about the regressions of the variables used in the model. It provides information on the fluctuation of each variable and the relationships between the inputs and the output. Fig. 4 demonstrate the scatter matrix of input and output variables. It is also called a pair plot because it shows pairwise relationships that exists in the dataset. The x-axis and y-axis of the pair plot are associated with rows and with columns respectively. Thus, it is a gird of axis that shows the associations between all the variables used in the development of the model [84]. It also shows the relation of C–S with the input variables. The diagonal of the scatter matrix presents a graphical representation of the input and output parameters distribution used in the creation of the models. Table 2 provides description of the data used to create the ML models. To obtain reliable predictions of C–S of SCC, it is advised to utilize the models within the specified range of variables presented in Table 2. These measures will contribute to the model's accuracy and overall validity. For the development of robust ML models, the dataset needs to be split into subsets. Thus, the 231 instances are randomly split into two data parts, named as training and validation data sets. The training set constitutes 70 % of the data, while the validation set comprises 30 % [85,86]. Out of the two data sets, the training data help in the creation of the models. On the other hand, the validation data is used to confirm the generalization capacity of the proposed model [87]. The reason behind this approach is to make sure that the model is not overfitted and can accurately predict outputs for unseen data.

4. Model development and performance assessment

4.1. MEP model development

It is necessary to mention that MEPX 2021.05.18.0 a computer software tool, is used to employ the MEP algorithm. Before modelling, numerous MEP fitting parameters must be established for constructing an efficient and generalised model. The inputs are chosen using recommendations from the literature and a trial-and-error method [88]. The optimized set parameters utilized to create the model are depicted in Table 3. The total number of programs present in the solution is specified by the size of subpopulation. Increase in subpopulation size increases the accuracy and complexity of the resulting equation but increasing it beyond a limit may cause overfitting.

4.2. RF model development

The random forest (RF) model is constructed using the Scikit-Learn module in Python. The algorithm's hyperparameters were carefully chosen through an iterative process, aiming to attain the highest accuracy. The number of estimators in the random forest regression (RFR) model, directly associated with the quantity of decision trees constructed, was predetermined prior to computing the maximum forecast averages. Increasing the number of trees enhances the model's effectiveness, but at the expense of increased computational requirements. The hyperparameter max depth signifies the depth each decision tree in the forest. The chosen hyperparameters are presented in Table 3.

4.3. XGBOOST model development

The XGB model was developed using Python programming language. An initial XGB model was constructed then the model was allowed to find out the most optimal hyperparameters with the lowest RMSE. Eventually, the model with optimal hyperparameters as shown in Table 3 was created.

Table 2
Statistical description of the data used in model development.

	Symbol	Average	Minimum	Maximum	Standard Deviation
Cement	x_0	311.98	135	542.10	102.45
Fly ash	x_1	145.45	0	390	112.54
Silica Fume	x_2	17.90	0	67.5	21.90
Coarse Aggregate	x_3	829.58	565	1091.4	162.27
Fine Aggregate	x_4	851.36	630	1120	116.80
Water	x_5	180.79	150	202.10	10.55
Superplasticizer	x_6	1.33	0	8.70	2.15
Strength	$f_c^{'}$	45.50	21.54	94.40	17.95

Table 3
Hyperparameters of the algorithms used for model development.

Parameters	Settings
MEP Parameters	
No. of Subpopulations	200
Size of Subpopulation	1000
Code Length	40
Number of Generations	1500
Runs	20
Functions	+, -, x, ÷, sqrt
RF Parameters	
n-estimators	100
Max-depth	40
Random State	10
XGB Parameters	
Max. Depth	3
n estimators	100
Initial Learning Rate	0.01

4.4. Performance assessment criteria of the models

Statistical evaluation is carried out to access the accuracy of the developed models. The statistical error metrics used to evaluate the model's performance are given below. In the expressions given below, x and y indicate experimental, and model predicted values respectively. Moreover, n indicates the total number of samples in a dataset and n20 shows the number of samples having ratio of actual and predicted values between 0.80 and 1.20.

$$\text{Mean Absolute Error (MAE)} = \frac{\sum |x - y|}{n}$$

$$\text{Root Mean Squared Error (RMSE)} = \sqrt{\frac{\sum (x - y)^2}{n}}$$

$$\text{Root Square Error} = \frac{\sum (x - y)^2}{\sum (x_{mean} - x)^2}$$

$$\text{Coefficient of Correlation}(R) = \frac{(n \sum xy - (\sum x)(\sum y))}{\sqrt{(n \sum x^2 - (\sum x)^2)(n \sum y^2 - (\sum y)^2)}}$$

$$\text{Coefficient of Determination} (R^2) = 1 - \frac{\sum (x - y)^2}{\sum (y - y_{mean})^2}$$

$$\text{Performance index}(\rho) = \frac{RRMSE}{1 + R}$$

$$\text{a20-index} = \frac{n20}{n}$$

The R^2 value is used to check accuracy of a model. A reliable and accurate model should have high R^2 value and minimum values of other error metrics. The R^2 value greater than 0.8 denotes a high relation between input and output variables [30]. However, it cannot be solely used to check a model's accuracy because it is not sensitive to multiplication or division of output with a constant [89]. Other statistical error metrics are also used alongside R^2. The average magnitude of mistakes is depicted by MAE and RMSE. A large RMSE value shows that the prediction with large errors is greater than the desired and should be reduced. The value of ρ ranges from 0 to infinity and a good model should have ρ value less than 0.2 [90]. It is worth mentioning that ρ considers RRMSE and R simultaneously and its value closer to zero indicates model's accuracy. The a20-index is also a very useful metric to assess the accuracy of ML models. It measures the proportion of predictions that deviate ±20 % from the predicted values [91]. It has been recommended in literature that for a perfect ML model, the a20-index value should be 1 [92].

The error metrics calculated for training and validation sets of the three algorithms are shown in Table 4. Notice that R^2 values of all the developed models are greater than 0.8 and surpassed 0.9 for both training and validation datasets. Also, according to the a20-index criteria, the developed models are accurate and robust. The a20-index value is 1 or close to 1 for both datasets of all the models. It indicates that the model predicted values are very close to the experimental values and none of the predicted values deviate more than

±20 % from the actual values. It is important to mention that the validation error metrics are less than the training ones which shows

excellent generalization capabilities of the models, and that the overfitting has been dealt effectively. Based on the error metrics, the XGB model outperformed both MEP and RF model and has the lowest values of RSE, RMSE, average error, and highest values of R^2. Thus, the most accurate model is XGB model, and MEP is the least accurate model, however it is important to mention that XGB does not generate an empirical equation like the MEP model.

5. Results

5.1. MEP result

In the development of a MEP model, the selection of most important parameters is the most important step. Thus, numerous trials were performed and the resulting equation for C–S is a function of the variables given in Equation (1).

$$f_c' = (x_0, x_1, x_2, x_5, x_6) \tag{1}$$

The output of MEP algorithm is decoded to get empirical equation for the C–S. The equation thus derived is given below:

$$f_c' = \left(\frac{x_6{}^2}{\sqrt{x_0} - x_1}\right)^3 + 2\left(\frac{x_1}{x_5}\right)^2 + \frac{x_2{}^2 + \left(\left((x_1 + x_0)\sqrt{x_0}\right) - x_0\right)}{x_5 + \sqrt{x_0}} + + \frac{34.6105 x_6}{x_0 - x_1 + 3.31565 - x_2} \tag{2}$$

Notice that Equation (2) is formulated using only the basic arithmetic operations and square root function. It is done to keep the equation simple and compact. The statistical error metrices calculated for MEP model are shown in Table 4. Note that the R^2 values of training and validation dataset are 0.924 and 0.923 respectively which shows the strong relation between the actual and predicted values. It is also evident from the scatter plots of MEP training and validation data shown in Fig. 5. The performance index values of training and validation are also closer to 0 which shows the accuracy of the algorithm.

5.2. RF result

The results of RF algorithm give a clear depiction of relation between the actual and predicted values of SCC compressive strength as shown in Fig. 6. The RF algorithm is more accurate than MEP algorithm with training $R^2 = 0.984$ and validation $R^2 = 0.986$. Also, the other error metrices like RMSE and average error are lower for RF algorithm than MEP as shown in Table 4. Similarly, the value of performance index of RF algorithm is closer to 0. Thus, the overall accuracy of RF algorithm is significantly greater than MEP, but it does not give an empirical equation like MEP.

5.3. XGB result

The excellent predictive capability of the XGB algorithm is evident from the scatter plots of actual versus predicted C–S values for both data sets. The scatter plots for the models are shown in Fig. 7. Also, the R^2 values of training and validation are 0.997 and 0.998 respectively which shows that XGB model is incredibly accurate, and it surpassed both the other algorithms. Moreover, the other error metrices such as MAE, RMSE, RSE and ρ are lowest for XGB algorithm.

5.4. Statistical comparison of the models

The excellent correlation that exists between the experimental and predicted C–S is evident from the scatter plots drawn for the three developed models. However, as discussed earlier, the accuracy of the algorithms is checked by calculating different statistical error metrics and the best model will be chosen having the highest correlation and minimum errors [30]. The comparison of R^2, average error and performance index of the three developed models is shown in Fig. 8. Notice from Fig. 8 (a) that the value of co-efficient of determination R^2 is the highest for XGB model. Both training and validation R^2 values are approaching 1 for XGB model. Also, the a20-index values are nearly equal to 1 for MEP and exactly equal to 1 for RF and XGB. It shows that the model predicted values

Table 4
Statistical evaluation of developed models.

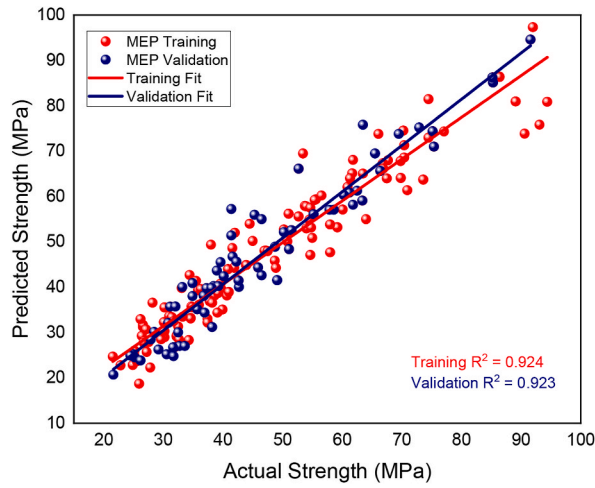|	MEP		RF		XGB	
	Training	Validation	Training	Validation	Training	Validation
MAE	3.45	3.58	1.45	1.35	0.527	0.45
RMSE	4.84	4.68	2.14	2.02	0.66	0.69
RSE	0.069	0.086	0.0145	0.0124	0.001	0.001
R	0.965	0.962	0.993	0.994	0.998	0.999
R^2	0.924	0.923	0.984	0.986	0.997	0.998
ρ	0.04	0.056	0.044	0.0426	0.006	0.007
a20-index	0.975	0.94	1	1	1	1

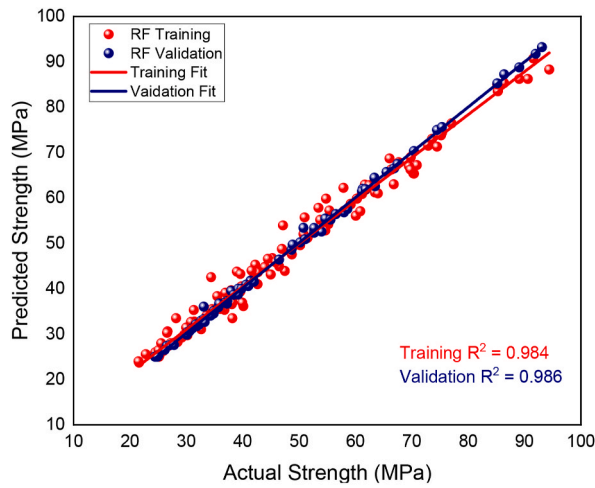**Fig. 5.** Scatter plots of MEP training and validation data.

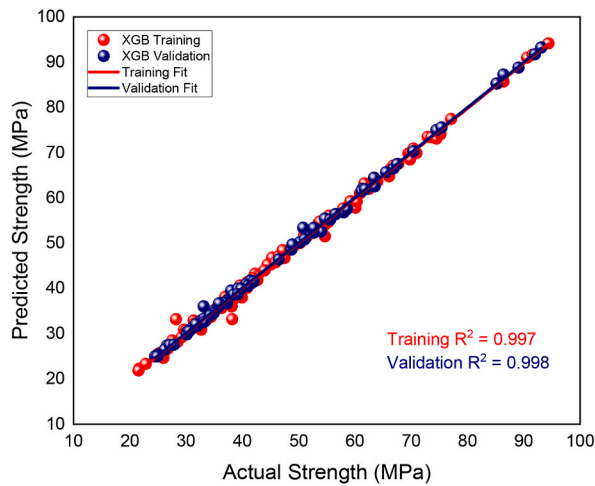**Fig. 6.** Scatter plots of RF training and validation data.

Fig. 7. Scatter plots for training and validation data sets.

lie close to the experimental values and doesn't deviate much from the experimental values. Similarly, it has the lowest average error out of the three developed models with training error 0.527 and validation error 0.47. Moreover, the value of performance index is also the lowest for XGB as shown in Fig. 8 (c). Also, notice from Fig. 8 (b) that MEP model has the highest absolute error for both datasets. The MEP validation error is 3.58 compared to 1.35 and 0.45 for RF and XGB respectively. Also, the R^2 values are the least for MEP model and highest for XGB. Thus, it can be concluded that XGB is the most accurate and robust model for predicting C–S of SCC having the least average error and highest correlation, followed by RF and MEP. However, MEP offers a significant advantage of presenting the result in the form of an empirical equation while RF and XGB do not yield an empirical equation.

Overfitting is a common problem that can occur when using ML algorithms. It means that a model performs well on the training data but cease to perform well on the validation or unseen data [93]. It is of sheer importance to check and avoid the overfitting of the ML models. Overfitting can be checked by comparing the training and validation error metrices for the model. If the validation errors are significantly less than the training error, means the model is overfitted to the training data [94]. Notice from Table 4 that almost all the validation error metrices are closer to the training ones and no validation error is significantly lesser than the training error metrices. It can also be verified from average error and correlation coefficient comparison given in Fig. 8. Thus, the developed models have good generalization capacity as they performed equally good on the validation data are not overfitted to the training data.

5.5. Sensitivity analysis (SA)

XGB model is considered for SA because it performed well than the other two algorithms. It is conducted to find the relative
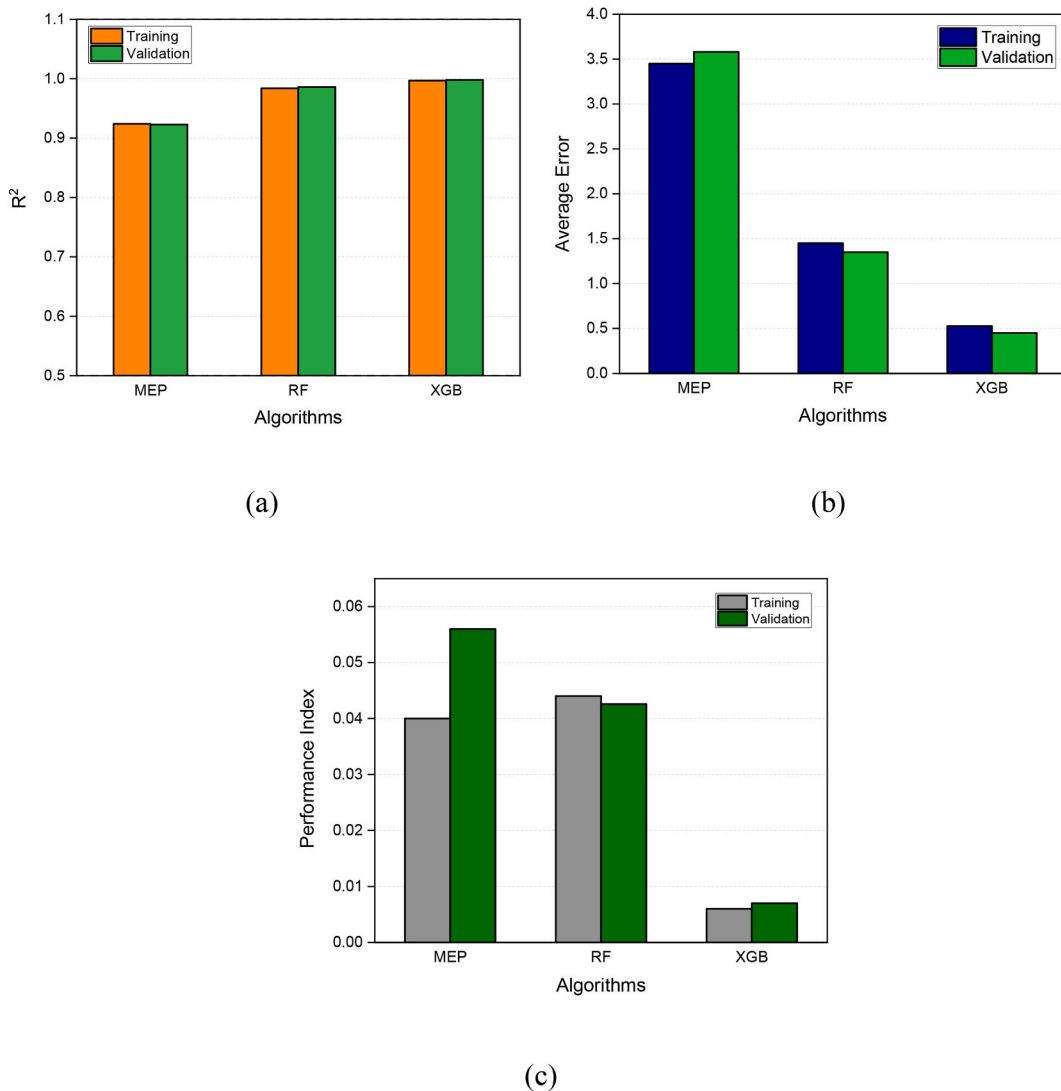

(a)   (b)

(c)

Fig. 8. Error comparison of the developed models; (a) R^2; (b) Average Error; (c) Performance Index.

contribution of inputs used in estimating fc' of SCC, using Equations (3) and (4). Sensitivity analysis tells that how much the output is dependent on the input variables. The results of SA are shown in Fig. 9.

$$N_i = f_{max}(x_i) - f_{min}(x_i)$$ (3)

$$SA = \frac{N_i}{\sum\limits_{n}^{j=1} N_j}$$ (4)

Where.

     $x_i = i\ th$ input variable keeping other variables constant
     $f_{max}(x_i)$ = maximum predicted output
     $f_{min}(x_i)$ = minimum predicted output

It is evident from Fig. 9 that every input plays a significant role in prediction of output. However, cement has a pivotal contribution in prediction of C–S contributing almost 35 % alone. The remaining parameters fly ash, silica fume, coarse aggregate, fine aggregate, water, and superplasticizer contributes 16.05 %, 12.64 %, 12.24 %,5.22 % and 16.75 % respectively. These conclusions are also in line with the findings of other studies [95]. A similar study attributed to finding C–S of SCC also concluded that cement is the most important factor in predicting the strength followed by admixtures like silica fume and fly ash [85]. The author concluded that small changes in the cement or mineral admixtures content can have a profound effect on C–S and other properties of SCC. Similarly, superplasticizer is also very important in achieving the strength and desired properties of SCC [96–98]. Also, the comparatively lesser role of fine aggregate and water content in the prediction of C–S has been explained in many studies [14,21,99].

6. Conclusions

This study employed MEP, RF and XGB to forecast C–S of SCC using a reliable dataset collected from literature. The models were evaluated using different statistical error metrices and the following conclusions are drawn from the developed models.

- The MEP predicted values showed a strong correlation with experimental values having validation R^2 equal to 0.923. The RF and XGB models also showed excellent accuracy with RF validation R^2 equal to 0.986 and XGB validation R^2 equal to 0.998. Thus, these algorithms are more accurate than MEP, but it is of paramount importance to mention that RF and XGB did not yield an empirical equation to predict strength like MEP.
- XGB had an average error of 0.45 compared to the 1.35 and 3.58 for RF and MEP respectively which shows that XGB is the most accurate of all the algorithms used. The same trend is observed for other error metrices.
- The scatter plots drawn between actual and predicted values show that the models are well-fitted across both datasets as evident from R^2 values for training and validation data. However, XGB model has the highest R^2 value, so it gives the most accurate predictions.
- The sensitivity analysis is done to study the influence of explanatory variables on the C–S and the results revealed that the trend of the relative importance of explanatory variables used in XGB followed the order: Cement (35.18 %) > Superplasticizer (16.75 %) > Fly ash (16.05 %) > Silica Fume (12.64 %) > Coarse aggregate (12.24 %) > fine aggregate (5.22 %) > water (1.3 %).

7. Limitations and future work

The aim of this study was to develop accurate prediction models for C–S of SCC to foster its use in the construction industry. While this study used a dataset of 231 points, it is recommended to use a larger dataset in subsequent research for the development of more comprehensive empirical models. The empirical models in the current study used seven explanatory variables, but it is crucial to consider other variables like age of concrete, different types of chemical and mineral admixtures etc. in future research endeavours. Also, the input variables used in the study had a limited range as given in Tables 2 and it is recommended to develop models with a broader range of input and output variables. Furthermore, predictive models for other mechanical properties of SCC like flexural strength, modulus of elasticity, tensile strength etc. should be developed to facilitate its use in the industry.

Data availability

Data will be made available on request.

CRediT authorship contribution statement

Waleed Bin Inqiad: Conceptualization, Data curation, Formal analysis, Software, Visualization, Writing - original draft. **Muhammad Shahid Siddique:** Validation, Supervision, Project administration. **Sana Shahab:** Project administration, Methodology. **Muhammad Jamal Butt:** Writing – review & editing. **Yaser Gamil:** Software, Funding acquisition, Methodology, Resources.
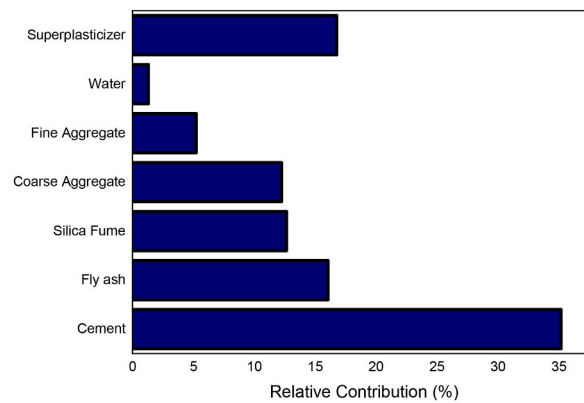
Fig. 9. Relative Contribution of input parameters (%).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e22036.

References

[1] A. Arrigoni, et al., Life cycle greenhouse gas emissions of concrete containing supplementary cementitious materials: cut-off vs. substitution, J. Clean. Prod. 263 (Aug) (2020), https://doi.org/10.1016/j.jclepro.2020.121465.
[2] N.C. Onat, M. Kucukvar, Carbon footprint of construction industry: a global review and supply chain analysis, Renew. Sustain. Energy Rev. 124 (May 2020), https://doi.org/10.1016/j.rser.2020.109783.
[3] R. Rehan, M. Nehdi, Carbon dioxide emissions and climate change: policy implications for the cement industry, Environ. Sci. Pol. 8 (2) (2005) 105–114, https://doi.org/10.1016/j.envsci.2004.12.006. Apr.
[4] E. Benhelal, G. Zahedi, E. Shamsaei, A. Bahadori, Global strategies and potentials to curb CO2 emissions in cement industry, Elsevier Ltd, J. Clean. Prod. 51 (2013) 142–161, https://doi.org/10.1016/j.jclepro.2012.10.049. Jul. 15.
[5] A.A. Shahmansouri, H. Akbarzadeh Bengar, S. Ghanbari, Compressive strength prediction of eco-efficient GGBS-based geopolymer concrete using GEP method, J. Build. Eng. 31 (Sep) (2020), https://doi.org/10.1016/j.jobe.2020.101326.
[6] H. Mikulčić, J.J. Klemeš, M. Vujanović, K. Urbaniec, N. Duić, Reducing greenhouse gasses emissions by fostering the deployment of alternative raw materials and energy sources in the cleaner cement manufacturing process, J. Clean. Prod. 136 (Nov. 2016) 119–132, https://doi.org/10.1016/j.jclepro.2016.04.145.
[7] K.M. Rahla, R. Mateus, L. Bragança, Comparative sustainability assessment of binary blended concretes using supplementary cementitious materials (SCMs) and ordinary Portland cement (OPC), J. Clean. Prod. 220 (May 2019) 445–459, https://doi.org/10.1016/j.jclepro.2019.02.010.
[8] P. Tang, W. Chen, D. Xuan, Y. Zuo, C.S. Poon, Investigation of cementitious properties of different constituents in municipal solid waste incineration bottom ash as supplementary cementitious materials, J. Clean. Prod. 258 (Jun) (2020), https://doi.org/10.1016/j.jclepro.2020.120675.
[9] R. Bajpai, K. Choudhary, A. Srivastava, K.S. Sangwan, M. Singh, Environmental impact assessment of fly ash and silica fume based geopolymer concrete, J. Clean. Prod. 254 (May 2020), https://doi.org/10.1016/j.jclepro.2020.120147.
[10] K.G. Kolovos, P.G. Asteris, D.M. Cotsovos, E. Badogiannis, S. Tsivilis, Mechanical properties of soilcrete mixtures modified with metakaolin, Construct. Build. Mater. 47 (2013) 1026–1036, https://doi.org/10.1016/j.conbuildmat.2013.06.008.
[11] K.G. Kolovos, P.G. Asteris, S. Tsivilis, Properties of sandcrete mixtures modified with metakaolin, European Journal of Environmental and Civil Engineering 20 (2016) s18–s37, https://doi.org/10.1080/19648189.2016.1246690. Nov.
[12] P.G. Asteris, K.G. Kolovos, Data on the physical and mechanical properties of soilcrete materials modified with metakaolin, Data Brief 13 (Aug. 2017) 487–497, https://doi.org/10.1016/j.dib.2017.06.014.
[13] V.W.Y. Tam, C.M. Tam, A review on the viable technology for construction waste recycling, Resour. Conserv. Recycl. 47 (3) (Jun. 2006) 209–221, https://doi.org/10.1016/j.resconrec.2005.12.002.
[14] H.J.H. Brouwers, H.J. Radix, Self-compacting concrete: theoretical and experimental study, Cem Concr Res 35 (11) (2005) 2116–2136, https://doi.org/10.1016/j.cemconres.2005.06.002.
[15] Geert De Schutter, Self-compacting Concrete, Whittles Pub., 2008.
[16] A. Bradu, N. Cazacu, N. Florea, and P. Mihai, "COMPRESSIVE STRENGTH OF SELF COMPACTING CONCRETE".
[17] O. Boukendakdji, S. Kenai, E.H. Kadri, F. Rouis, Effect of slag on the rheology of fresh self-compacted concrete, Construct. Build. Mater. 23 (7) (Jul. 2009) 2593–2598, https://doi.org/10.1016/j.conbuildmat.2009.02.029.

[18] O. Mohamed, Durability and compressive strength of high cement replacement ratio self-consolidating concrete, Buildings 8 (11) (2018), https://doi.org/10.3390/buildings8110153. Nov.

[19] Z. Grdic, I. Despotovic, G. Toplicic-Curcic, Properties of self-compacting concrete with different types of additives, Facta Univ. – Ser. Archit. Civ. Eng. 6 (2) (2008) 173–177, https://doi.org/10.2298/fuace0802173g.

[20] P.G. Asteris, K.G. Kolovos, Self-compacting concrete strength prediction using surrogate models, Neural Comput. Appl. 31 (Jan. 2019) 409–424, https://doi.org/10.1007/s00521-017-3007-7.

[21] M. Valcuende, E. Marco, C. Parra, P. Serna, Influence of limestone filler and viscosity-modifying admixture on the shrinkage of self-compacting concrete, Cem Concr Res 42 (4) (Apr. 2012) 583–592, https://doi.org/10.1016/j.cemconres.2012.01.001.

[22] P.G. Asteris, K.G. Kolovos, M.G. Douvika, K. Roinos, Prediction of self-compacting concrete strength using artificial neural networks, European Journal of Environmental and Civil Engineering 20 (2016) s102–s122, https://doi.org/10.1080/19648189.2016.1246693. Nov.

[23] I. Nunez, A. Marani, M.L. Nehdi, Mixture optimization of recycled aggregate concrete using hybrid machine learning model, Materials 13 (19) (2020) 1–24, https://doi.org/10.3390/ma13194331. Oct.

[24] C.W. Zobel, D.F. Cook, Evaluation of neural network variable influence measures for process control, Eng. Appl. Artif. Intell. 24 (5) (Aug. 2011) 803–812, https://doi.org/10.1016/j.engappai.2011.03.001.

[25] R.H. Faraj, A.A. Mohammed, K.M. Omer, Modeling the compressive strength of eco-friendly self-compacting concrete incorporating ground granulated blast furnace slag using soft computing techniques, Environ. Sci. Pollut. Control Ser. 29 (47) (Oct. 2022), 71338, https://doi.org/10.1007/s11356-022-20889-5. –71357.

[26] H. Tanyildizi, A. Çevik, Modeling mechanical performance of lightweight concrete containing silica fume exposed to high temperature using genetic programming, Construct. Build. Mater. 24 (12) (2010) 2612–2618, https://doi.org/10.1016/j.conbuildmat.2010.05.001.

[27] H.L. Wang, Z.Y. Yin, High performance prediction of soil compaction parameters using multi expression programming, Eng. Geol. 276 (Oct) (2020), https://doi.org/10.1016/j.enggeo.2020.105758.

[28] D. Mohammadzadeh, S.-F. Kazemi, A. Mosavi, E. Nasseralshariati, J.H.M. Tah, Prediction of Compression Index of Fine-Grained Soils Using a Gene Expression Programming Model, Infrastructures 4 (2) (2019) 26. https://doi.org/10.3390/infrastructures4020026.

[29] A. Demir, "New Computational Models for Better Predictions of the Soil-Compression Index".

[30] A.H. Alavi, A.H. Gandomi, M.G. Sahab, M. Gandomi, Multi expression programming: a new approach to formulation of soil classification, Eng. Comput. 26 (2) (Apr. 2010) 111–118, https://doi.org/10.1007/s00366-009-0140-7.

[31] M. Saridemir, Genetic programming approach for prediction of compressive strength of concretes containing rice husk ash, Construct. Build. Mater. 24 (10) (Oct. 2010) 1911–1919, https://doi.org/10.1016/j.conbuildmat.2010.04.011.

[32] V.L. Tran, S.E. Kim, Efficiency of three advanced data-driven models for predicting axial compression capacity of CFDST columns, Thin-Walled Struct. 152 (Jul) (2020), https://doi.org/10.1016/j.tws.2020.106744.

[33] M. Sarveghadi, A.H. Gandomi, H. Bolandi, A.H. Alavi, Development of prediction models for shear strength of SFRCB using a machine learning approach, Springer London, Neural Comput. Appl. 31 (7) (2019) 2085–2094, https://doi.org/10.1007/s00521-015-1997-6. Jul. 01.

[34] H.H. Chu, et al., Sustainable use of fly-ash: use of gene-expression programming (GEP) and multi-expression programming (MEP) for forecasting the compressive strength geopolymer concrete, Ain Shams Eng. J. 12 (4) (Dec. 2021) 3603–3617, https://doi.org/10.1016/j.asej.2021.03.018.

[35] P.G. Asteris, P.C. Roussis, M.G. Douvika, Feed-forward neural network prediction of the mechanical properties of sandcrete materials, Sensors 17 (6) (2017), https://doi.org/10.3390/s17061344. Jun.

[36] P. Sarir, J. Chen, P.G. Asteris, D.J. Armaghani, M.M. Tahir, Developing GEP tree-based, neuro-swarm, and whale optimization models for evaluation of bearing capacity of concrete-filled steel tube columns, Eng. Comput. 37 (1) (Jan. 2021) 1–19, https://doi.org/10.1007/s00366-019-00808-y.

[37] F. Farooq, et al., A comparative study for the prediction of the compressive strength of self-compacting concrete modified with fly ash, Materials 14 (17) (2021), https://doi.org/10.3390/ma14174934. Sep.

[38] P.G. Asteris, K.G. Kolovos, M.G. Douvika, K. Roinos, Prediction of self-compacting concrete strength using artificial neural networks, European Journal of Environmental and Civil Engineering 20 (2016) s102–s122, https://doi.org/10.1080/19648189.2016.1246693. Nov.

[39] H.V.T. Mai, M.H. Nguyen, H.B. Ly, Development of machine learning methods to predict the compressive strength of fiber-reinforced self-compacting concrete and sensitivity analysis, Construct. Build. Mater. 367 (2023), https://doi.org/10.1016/j.conbuildmat.2023.130339. Feb.

[40] J. de-Prado-Gil, C. Palencia, N. Silva-Monteiro, R. Martínez-García, To predict the compressive strength of self compacting concrete with recycled aggregates utilizing ensemble machine learning models, Case Stud. Constr. Mater. 16 (Jun) (2022), https://doi.org/10.1016/j.cscm.2022.e01046.

[41] P. Saha, P. Debnath, P. Thomas, Prediction of fresh and hardened properties of self-compacting concrete using support vector regression approach, Neural Comput. Appl. 32 (12) (Jun. 2020) 7995–8010, https://doi.org/10.1007/s00521-019-04267-w.

[42] R. Bušić, M. Benšić, I. Miličević, K. Strukar, Prediction models for the mechanical properties of self-compacting concrete with recycled rubber and silica fume, Materials 13 (8) (2020), https://doi.org/10.3390/MA13081821. Apr.

[43] R. Siddique, P. Aggarwal, Y. Aggarwal, Prediction of compressive strength of self-compacting concrete containing bottom ash using artificial neural networks, Adv. Eng. Software 42 (10) (2011) 780–786, https://doi.org/10.1016/j.advengsoft.2011.05.016.

[44] S.A. Rajakarunakaran, et al., Prediction of strength and analysis in self-compacting concrete using machine learning based regression techniques, Adv. Eng. Software 173 (2022), https://doi.org/10.1016/j.advengsoft.2022.103267. Nov.

[45] A. Kaveh, T. Bakhshpoori, S.M. Hamze-Ziabari, M5' and mars based prediction models for properties of selfcompacting concrete containing fly ash, Period. Polytech. Civ. Eng. 62 (2) (Mar. 2018) 281–294, https://doi.org/10.3311/PPci.10799.

[46] J. Web, L. Chen, and W. Jiang, "Advanced in Engineering and Intelligence Systems Estimation of the Compressive Strength of Self-Compacting Concrete (SCC) by a Machine Learning Technique Coupling with Novel Optimization Algorithms".

[47] M. Sonebi, A. Cevik, Genetic programming based formulation for fresh and hardened properties of self-compacting concrete containing pulverised fuel ash, Construct. Build. Mater. 23 (7) (Jul. 2009) 2614–2622, https://doi.org/10.1016/j.conbuildmat.2009.02.012.

[48] O. Belalia Douma, B. Boukhatem, M. Ghrici, A. Tagnit-Hamou, Prediction of properties of self-compacting concrete containing fly ash using artificial neural network, Neural Comput. Appl. 28 (Dec. 2017) 707–718, https://doi.org/10.1007/s00521-016-2368-7.

[49] C. Kina, K. Turk, E. Atalay, I. Donmez, H. Tanyildizi, Comparison of extreme learning machine and deep learning model in the estimation of the fresh properties of hybrid fiber-reinforced SCC, Neural Comput. Appl. 33 (18) (Sep. 2021) 11641–11659, https://doi.org/10.1007/s00521-021-05836-8.

[50] N. Abunassar, M. Alas, S.I.A. Ali, Prediction of compressive strength in self-compacting concrete containing fly ash and silica fume using ANN and SVM, Arabian J. Sci. Eng. 48 (4) (Apr. 2023) 5171–5184, https://doi.org/10.1007/s13369-022-07359-3.

[51] D. Ghunimat, A.E. Alzoubi, A. Alzboon, S. Hanandeh, Prediction of concrete compressive strength with GGBFS and fly ash using multilayer perceptron algorithm, random forest regression and k-nearest neighbor regression, Asian Journal of Civil Engineering 24 (1) (Jan. 2023) 169–177, https://doi.org/10.1007/s42107-022-00495-z.

[52] C. Kina, K. Turk, H. Tanyildizi, Estimation of strengths of hybrid FR-SCC by using deep-learning and support vector regression models, Struct. Concr. 23 (5) (Oct. 2022) 3313–3330, https://doi.org/10.1002/suco.202100622.

[53] M. Oltean, Multi Expression Programming for solving classification problems Fruit recognition from images using deep learning View project Optical Computing View project Mihai Oltean Multi Expression Programming for solving classification problems (2022), https://doi.org/10.21203/rs.3.rs-1458572/v1.

[54] Q. Zhang, X. Meng, B. Yang, W. Liu, MREP: multi-reference expression programming, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer Verlag, 2016, pp. 26–38, https://doi.org/10.1007/978-3-319-42294-7_3.

[55] M.O. Crina, G. Gros¸an, A Comparison of Several Linear GP Techniques A Comparison of Several Linear Genetic Programming Techniques, 2003 [Online].

[56] W. Deng, P. He, Z. Huang, Multi-expression based gene expression programming, in: Lecture Notes in Electrical Engineering, 2013, pp. 439–448, https://doi.org/10.1007/978-3-642-38466-0_49.

[57] C. Chisari, C. Bedon, Multi-objective optimization of FRP jackets for improving the seismic response of reinforced concrete frames, Am. J. Eng. Appl. Sci. 9 (3) (Sep. 2016) 669–679, https://doi.org/10.3844/ajeassp.2016.669.679.

[58] Oltean, Mihai, and D. Dumitrescu. "Multi expression programming." *Journal of Genetic Programming and Evolvable Machines* (2002).Oltean M., Multi Expression Programming (2006) [Online].

[59] A.H. Gandomi, A. Faramarzifar, P.G. Rezaee, A. Asghari, S. Talatahari, New design equations for elastic modulus of concrete using multi expression programming, J. Civ. Eng. Manag. 21 (6) (Aug. 2015) 761–774, https://doi.org/10.3846/13923730.2014.893910.

[60] F.E. Jalal, et al., Indirect estimation of swelling pressure of expansive soil: GEP versus MEP modelling, Adv. Mater. Sci. Eng. 2023 (Jan. 2023) 1–25, https://doi.org/10.1155/2023/1827117.

[61] L. Breiman, Random Forests, 2001.

[62] X. liang Jin, et al., Estimation of wheat agronomic parameters using new spectral indices, PLoS One 8 (8) (2013), https://doi.org/10.1371/journal.pone.0072736. Aug.

[63] C. Qi, A. Fourie, X. Du, X. Tang, Prediction of open stope hangingwall stability using random forests, Nat. Hazards 92 (2) (2018) 1179–1197, https://doi.org/10.1007/s11069-018-3246-7. Jun.

[64] K. Zhang, X. Wu, R. Niu, K. Yang, L. Zhao, The assessment of landslide susceptibility mapping using random forest and decision tree methods in the Three Gorges Reservoir area, China, Environ. Earth Sci. 76 (11) (2017), https://doi.org/10.1007/s12665-017-6731-5. Jun.

[65] C. Qi, X. Tang, Slope stability prediction using integrated metaheuristic and machine learning approaches: a comparative study, Comput. Ind. Eng. 118 (Apr. 2018) 112–122, https://doi.org/10.1016/j.cie.2018.02.028.

[66] C. Brokamp, R. Jandarov, M.B. Rao, G. LeMasters, P. Ryan, Exposure assessment models for elemental components of particulate matter in an urban environment: a comparison of regression and random forest approaches, Atmos. Environ. 151 (Feb. 2017) 1–11, https://doi.org/10.1016/j.atmosenv.2016.11.066.

[67] X. Wu, F. Zhu, M. Zhou, M.M.S. Sabri, J. Huang, Intelligent design of construction materials: a comparative study of ai approaches for predicting the strength of concrete with blast furnace slag, Materials 15 (13) (2022), https://doi.org/10.3390/ma15134582. Jul.

[68] T. Chen, C. Guestrin, XGBoost: a scalable tree boosting system, in: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, Aug. 2016, pp. 785–794, https://doi.org/10.1145/2939672.2939785.

[69] L. Guelman, Gradient boosting trees for auto insurance loss cost modeling and prediction, Expert Syst. Appl. 39 (3) (Feb. 2012) 3659–3667, https://doi.org/10.1016/j.eswa.2011.09.058.

[70] L. Cui, P. Chen, L. Wang, J. Li, H. Ling, Application of extreme gradient boosting based on grey relation analysis for prediction of compressive strength of concrete, Adv. Civ. Eng. 2021 (2021), https://doi.org/10.1155/2021/8878396.

[71] C. Wang, S. Xu, J. Yang, Adaboost algorithm in artificial intelligence for optimizing the IRI prediction accuracy of asphalt concrete pavement, Sensors 21 (17) (2021), https://doi.org/10.3390/s21175682. Sep.

[72] A. Marani, A. Jamali, M.L. Nehdi, Predicting ultra-high-performance concrete compressive strength using tabular generative adversarial networks, Materials 13 (21) (2020) 1–24, https://doi.org/10.3390/ma13214757. Nov.

[73] M. Sonebi, Medium strength self-compacting concrete containing fly ash: modelling using factorial experimental plans, Cem Concr Res 34 (7) (Jul. 2004) 1199–1208, https://doi.org/10.1016/j.cemconres.2003.12.022.

[74] R. Siddique, Properties of self-compacting concrete containing class F fly ash, Mater. Des. 32 (3) (Mar. 2011) 1501–1507, https://doi.org/10.1016/j.matdes.2010.08.043.

[75] P. Ramanathan, I. Baskar, P. Muthupriya, R. Venkatasubramani, Performance of self-compacting concrete containing different mineral admixtures, KSCE J. Civ. Eng. 17 (2) (2013) 465–472, https://doi.org/10.1007/s12205-013-1882-8. Mar.

[76] H. Yazici, The effect of silica fume and high-volume Class C fly ash on mechanical properties, chloride penetration and freeze-thaw resistance of self-compacting concrete, Construct. Build. Mater. 22 (4) (Apr. 2008) 456–462, https://doi.org/10.1016/j.conbuildmat.2007.01.002.

[77] O.M. Ofuyatan, A.G. Adeniyi, J.O. Ighalo, Evaluation of fresh and hardened properties of blended silica fume self-compacting concrete (SCC), Research on Engineering Structures and Materials 7 (2) (Jun. 2021) 211–223, https://doi.org/10.17515/resm2020.228ma1023.

[78] Z. Yang, S. Liu, L. Yu, L. Xu, A comprehensive study on the hardening features and performance of self-compacting concrete with high-volume fly ash and slag, Materials 14 (15) (2021), https://doi.org/10.3390/ma14154286. Aug.

[79] Z. Guo, T. Jiang, J. Zhang, X. Kong, C. Chen, D.E. Lehman, Mechanical and durability properties of sustainable self-compacting concrete with recycled concrete aggregate and fly ash, slag and silica fume, Construct. Build. Mater. 231 (Jan) (2020), https://doi.org/10.1016/j.conbuildmat.2019.117115.

[80] B. Felekoğlu, S. Türkel, B. Baradan, Effect of water/cement ratio on the fresh and hardened properties of self-compacting concrete, Build. Environ. 42 (4) (Apr. 2007) 1795–1802, https://doi.org/10.1016/j.buildenv.2006.01.012.

[81] M. Gesoğlu, E. Özbay, Effects of mineral admixtures on fresh and hardened properties of self-compacting concretes: binary, ternary and quaternary systems, Materials and Structures/Materiaux et Constructions 40 (9) (2007) 923–937, https://doi.org/10.1617/s11527-007-9242-0. Nov.

[82] M. Gesoğlu, E. Güneyisi, E. Özbay, Properties of self-compacting concretes made with binary, ternary, and quaternary cementitious blends of fly ash, blast furnace slag, and silica fume, Construct. Build. Mater. 23 (5) (May 2009) 1847–1854, https://doi.org/10.1016/j.conbuildmat.2008.09.015.

[83] O. Boukendakdji, E.H. Kadri, S. Kenai, Effects of granulated blast furnace slag and superplasticizer type on the fresh properties and compressive strength of self-compacting concrete, Cem. Concr. Compos. 34 (4) (Apr. 2012) 583–590, https://doi.org/10.1016/j.cemconcomp.2011.08.013.

[84] R.A. Groeneveld, G. Meeden, Measuring Skewness and Kurtosis, 1984 [Online]. Available: https://about.jstor.org/terms.

[85] J. De-Prado-gil, C. Palencia, P. Jagadesh, R. Martínez-García, A comparison of machine learning tools that model the splitting tensile strength of self-compacting recycled aggregate concrete, Materials 15 (12) (2022), https://doi.org/10.3390/ma15124164. Jun.

[86] F. Farooq, et al., A comparative study for the prediction of the compressive strength of self-compacting concrete modified with fly ash, Materials 14 (17) (2021), https://doi.org/10.3390/ma14174934. Sep.

[87] A. Gholampour, A.H. Gandomi, T. Ozbakkaloglu, New formulations for mechanical properties of recycled aggregate concrete using gene expression programming, Construct. Build. Mater. 130 (Jan. 2017) 122–145, https://doi.org/10.1016/j.conbuildmat.2016.10.114.

[88] S.M. Mousavi, A.H. Alavi, A.H. Gandomi, M.A. Esmaeili, M. Gandomi, A Data Mining Approach to Compressive Strength of CFRP-Confined Concrete Cylinders, 2010.

[89] F.E. Jalal, Y. Xu, M. Iqbal, B. Jamhiri, M.F. Javed, Predicting the compaction characteristics of expansive soils using two genetic programming-based algorithms, Transportation Geotechnics 30 (2021), https://doi.org/10.1016/j.trgeo.2021.100608. Sep.

[90] A.H. Gandomi, D.A. Roke, Assessment of artificial neural network and genetic programming as predictive tools, Adv. Eng. Software 88 (Jun. 2015) 63–72, https://doi.org/10.1016/j.advengsoft.2015.05.007.

[91] Y. Li, et al., The effects of rock index tests on prediction of tensile strength of granitic samples: a neuro-fuzzy intelligent system, Sustainability 13 (19) (2021), https://doi.org/10.3390/su131910541. Oct.

[92] P.G. Asteris, M. Koopialipoor, D.J. Armaghani, E.A. Kotsonis, P.B. Lourenço, Prediction of cement-based mortars compressive strength using machine learning techniques, Neural Comput. Appl. 33 (19) (Oct. 2021) 13089–13121, https://doi.org/10.1007/s00521-021-06004-8.

[93] D.J. Armaghani, P.G. Asteris, A comparative study of ANN and ANFIS models for the prediction of cement-based mortar materials compressive strength, Neural Comput. Appl. 33 (9) (May 2021) 4501–4532, https://doi.org/10.1007/s00521-020-05244-4.

[94] S. Whiteson, B. Tanner, M. E. Taylor, and P. Stone, "Protecting against Evaluation Overfitting in Empirical Reinforcement Learning".

[95] F. Farooq, et al., A comparative study for the prediction of the compressive strength of self-compacting concrete modified with fly ash, Materials 14 (17) (2021), https://doi.org/10.3390/ma14174934. Sep.

[96] P. Jagadesh, J. de Prado-Gil, N. Silva-Monteiro, R. Martínez-García, Assessing the compressive strength of self-compacting concrete with recycled aggregates from mix ratio using machine learning approach, J. Mater. Res. Technol. 24 (May 2023) 1483–1498, https://doi.org/10.1016/j.jmrt.2023.03.037.

[97] J. Sobhani, M. Najimi, A.R. Pourkhorshidi, T. Parhizkar, Prediction of the compressive strength of no-slump concrete: a comparative study of regression, neural network and ANFIS models, Construct. Build. Mater. 24 (5) (May 2010) 709–718, https://doi.org/10.1016/j.conbuildmat.2009.10.037.

[98] O. Boukendakdji, E.H. Kadri, S. Kenai, Effects of granulated blast furnace slag and superplasticizer type on the fresh properties and compressive strength of self-compacting concrete, Cem. Concr. Compos. 34 (4) (Apr. 2012) 583–590, https://doi.org/10.1016/j.cemconcomp.2011.08.013.

[99] S.A. Rajakarunakaran, et al., Prediction of strength and analysis in self-compacting concrete using machine learning based regression techniques, Adv. Eng. Software 173 (2022), https://doi.org/10.1016/j.advengsoft.2022.103267. Nov.