



## OPEN Predicting 28-day compressive strength of fibre-reinforced self-compacting concrete (FR-SCC) using MEP and GEP

Waleed Bin Inqiad<sup>1</sup>, Muhammad Shahid Siddique<sup>1</sup>, Mujahid Ali<sup>2</sup> & Taoufik Najeh<sup>3</sup>✉

The utilization of Self-compacting Concrete (SCC) has escalated worldwide due to its superior properties in comparison to normal concrete such as compaction without vibration, increased flowability and segregation resistance. Various other desirable properties like ductile behaviour, increased strain capacity and tensile strength etc. can be imparted to SCC by incorporation of fibres. Thus, this study presents a novel approach to predict 28-day compressive strength (C–S) of FR-SCC using Gene Expression Programming (GEP) and Multi Expression Programming (MEP) for fostering its widespread use in the industry. For this purpose, a dataset had been compiled from internationally published literature having six input parameters including water-to-cement ratio, silica fume, fine aggregate, coarse aggregate, fibre, and superplasticizer. The predictive abilities of developed algorithms were assessed using error metrics like mean absolute error (MAE), a20-index, and objective function (OF) etc. The comparison of MEP and GEP models indicated that GEP gave a simple equation having lesser errors than MEP. The OF value of GEP was 0.029 compared to 0.031 of MEP. Thus, sensitivity analysis was performed on GEP model. The models were also checked using some external validation checks which also verified that MEP and GEP equations can be used to forecast the strength of FR-SCC for practical uses.

**Keywords** Self-compacting concrete, Genetic Programming, Fiber-reinforced self-compacting concrete, Multi expression programming, Gene expression programming

The manufacture of cement prompts CO<sub>2</sub> emission in atmosphere and depletion of natural resources like limestone etc. The annually manufactured quantity of cement is around 4000 million tons<sup>1</sup> and it is the most emission intensive substance from the construction industry which contributes 7% to the global carbon emissions alone<sup>2</sup>. The release of these harmful greenhouse gases in atmosphere poses a serious threat to humanity and result in rising global temperatures, melting of glaciers etc. To reduce the carbon emissions associated with production of cement, new materials are being introduced with the aim to replace cement in concrete<sup>3</sup>. These products include a variety of materials like marble powder, fly ash, silica fume, slag etc. and are known as Secondary Cementitious Materials (SCMs). The utilization of these SCMs in concrete can help to lessen the ill effects of concrete industry on the environment for a sustainable future<sup>4</sup>. The recent advancements in low-carbon concrete composites have led to the invention of a new form of concrete named as SCC. It was invented in Japan in 1980s<sup>5</sup>. Soon after its development, it found many uses in buildings, bridges, and precast concrete components around the world<sup>6</sup> due to its advantages such as high structural quality, increased productivity, and durability<sup>7</sup>. SCC has many properties that deviate from that of normal concrete. The most important being its ability to flow and compact itself under its own weight. It can flow and fill the spaces in members that have congested arrangement of reinforcement or where conventional vibrating methods are not applicable<sup>8</sup>. SCC also offers many other advantages such as segregation resistance, improved surface finish, and increased durability. It also leads to improvement in the site conditions by eliminating the need of mechanical compaction methods such as vibrators that cause fossil fuels burning and CO<sub>2</sub> emission<sup>9</sup>.

<sup>1</sup>Military College of Engineering (MCE), National University of Science and Technology (NUST), Islamabad 44000, Pakistan. <sup>2</sup>Department of Transport Systems, Traffic Engineering and Logistics, Faculty of Transport and Aviation Engineering, Silesian University of Technology, Krasińskiego 8 Street, 40-019 Katowice, Poland. <sup>3</sup>Operation and Maintenance, Operation, Maintenance and Acoustics, Department of Civil, Environmental and Natural Resources Engineering, Lulea University of Technology, Lulea, Sweden. ✉email: taoufik.najeh@ltu.se

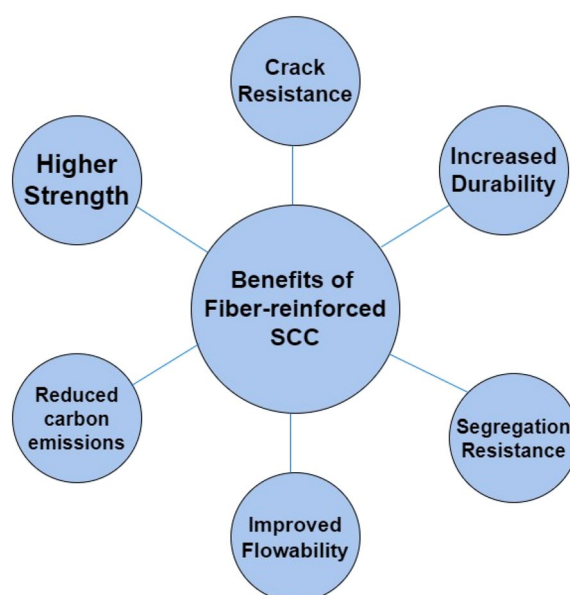
### Fibre-reinforced SCC

The SCC mix design is at the heart of achieving the desirable characteristics. Generally, SCC requires higher water-to-cement ratio and the use of chemical admixtures to impart the flowability characteristics to SCC. Also, the mix design calls for the use of greater quantity of fine materials to fill the spaces between coarse aggregate<sup>10</sup>. Since the use of huge quantities of cement to increase the level of fines can be costly, researchers have suggested using several mineral admixtures as a replacement of cement in SCC<sup>11–14</sup>. The use of these admixtures results in cost reduction and improve strength and durability characteristics<sup>15–20</sup>. Despite the benefits of SCC as an innovative material, it has the drawback of being a brittle material having low tensile strength and strain capacity like normal concrete<sup>21,22</sup>. Thus, the incorporation of fibres is desirable to render a ductile behaviour to SCC and improve its crack resistance, tensile strength, and strain capacity<sup>23</sup>. Fibre-reinforced concrete is created by adding short and discontinuous fibres that are dispersed throughout the concrete matrix<sup>24</sup>. The various types of fibres used in industry include steel fibres, glass fibres, polypropylene fibres etc. It has been reported that using a mix of fibres significantly improves the concrete properties<sup>25</sup>. Thus, this study used a dataset collected from literature having a mix of polypropylene and glass fibres. The incorporation of fibres in SCC results in the development of F-SCC which results in improvement of post-cracking behaviour by preventing the propagation of small and newly developed cracks<sup>23</sup>. Thus, F-SCC offers many added advantages compared to non-reinforced SCC as shown in Fig. 1. However, for effective utilization of F-SCC in construction industry, we must be able to accurately predict its 28-day compressive strength because it is an indicator of overall concrete quality and durability. However, to effectively utilize a revolutionary product like FR-SCC we must be able to accurately estimate its different properties out of which 28-day compressive strength (C-S) is the most important one. The 28-day compressive strength test is useful to check the overall quality and performance of concrete<sup>26–29</sup>. But it has been observed that a comprehensive experimental investigation to check the mechanical properties of new concrete composites is time consuming as a large number of loading conditions, exposure types and mixture compositions must be studied to get an accurate estimate of the properties<sup>30</sup>. Also, it is not an easy task to conduct these experiments in identical and controlled environment due to the inherent complexities and limitations of the conventional testing procedures<sup>31</sup>. Moreover, most of the standard tests are destructive in nature means that the specimen is destroyed during the test, so this also contributes to the resource wastage and increases construction waste. Thus, there are many limitations to the testing of concrete composites using conventional testing methods.

To overcome the limitations stated above, researchers have come up with the idea that mathematical models should be used to predict the different properties of composite materials<sup>32</sup>. These machine learning models reduce the time, effort, and cost required for experimental investigation and accurately predict the behaviour of different materials under various conditions without conducting experiments<sup>33</sup>. The literature regarding the utilization of several famous machine learning techniques to predict different concrete properties is described in next section. In Section “[Research significance](#)”, the particular research novelty and significance has been highlighted followed by detailed explanation of the employed algorithms in Section “[Research methodology](#)”. Section “[Research methodology](#)” also sheds light on data collection which was used to build the predictive models and statistical analysis performed on the data while Section “[Results and discussions](#)” deals with the results of the algorithms.

### Overview of machine learning in civil engineering

During the last few years, civil engineering industry is making a shift towards the sustainable materials and data-driven decision-making tools like every other industry. In the field of civil engineering, AI is mainly used to



**Figure 1.** Benefits of fibre-reinforced SCC.

optimize material utilization, foster sustainable construction, and improve the overall efficiency and accuracy of construction processes by using various machine learning and deep learning techniques. Machine learning (ML) is a subset of artificial intelligence (AI) which refers to the process in which machines learn from vast amounts of data without any human intervention. ML techniques learn patterns from the data and use that to make informed decisions for future data. Deep learning is in turn a subset of ML which refers to the use of neural networks to make predictions on unseen data. Recently, the prediction of various properties of new concrete composites using machine learning (ML) models has captured the interest of researchers. This is due to the simplicity, accuracy, cost, and time effectiveness of the ML techniques<sup>32</sup>. ML algorithms are basically computational methods designed to solve complex problems like humans. It has been reported that ML algorithms can discover the hidden patterns in the data and make predictions based on the learning from the data<sup>34</sup>. ML techniques have been successfully applied to solve problems related to finance, biology, engineering etc. In the domain of civil engineering, ML techniques have found their use in predicting various properties of different concrete and cement composites<sup>35–40</sup>, composite concrete columns and tubes<sup>41–45</sup>, soil classification and compaction factors etc.<sup>46,47</sup>. ML techniques have also found their uses in geotechnical and mining engineering to determine slope stability, landslide susceptibility, and rock fracture prediction etc.<sup>48–60</sup>. Moreover, ML algorithms have been used to determine service life of concrete structures in marine environments<sup>61</sup> and health monitoring of asphalt pavements<sup>62</sup>.

## Literature review

The prediction of SCC properties using various ML algorithms like Support Vector Machine (SVM), Artificial Neural Networks (ANN), etc. has increased in the last few years due to the immense potential of SCC to reduce the concrete associated carbon emissions. In 2021, Farooq et al.<sup>63</sup> utilized different ML algorithms including ANN, SVM and GEP to predict CS of fly ash containing SCC. The study concluded that GEP algorithm is the accurate one having average error of only 3.71 followed by 5.03 and 5.42 for ANN and SVM respectively. Similarly in 2022, Abunassar et al.<sup>64</sup> utilized ANN and SVM to predict CS of SCC having silica fume and fly ash as admixtures. The authors utilized 85 data points to develop ML models that have seven input and one output variable. The comparison was drawn between the two algorithms based on the  $R^2$  value when models were used to predict unseen test data. The results showed that both techniques are reliable and accurate to predict CS of SCC, but SVM achieved  $R^2 = 0.97$  compared to 0.96 of ANN thus it is more accurate than ANN. Also, Asteris et al.<sup>10</sup> utilized ANN to forecast strength of SCC having a wide variety of admixtures. These admixtures included slag, fly ash, limestone powder, silica fume, rice husk ash, etc. The employed a dataset of 169 points and the accuracy of resulting model was measured using coefficient of determination. The authors concluded that ANN showcased a great ability to predict strength of SCC blended with various admixtures having  $R^2$  value equal to 0.98. In the same way, the strength estimation of lightweight concrete made by using oil palm by-product and concrete made with pozzolanic admixtures was done by<sup>65</sup>. Moreover, to foster the utilization of recycled construction waste in concrete, Ashrafian et al. utilized neuro-fuzzy algorithm to provide accurate prediction model for compressive strength of environmentally friendly concrete containing recycled construction wastes<sup>66</sup>. The authors gathered an extensive database of more than 300 concrete mixture proportions having construction waste and used a neuro-fuzzy algorithm with a Horse Herd Optimization Algorithm (HOA) for the purpose of strength estimation and reported that this approach performed well than other standalone algorithms and optimization approaches for prediction of strength.

As far as the subject of prediction of F-SCC properties is concerned, there are very few studies available. The important ones include the study conducted by Rahman et al.<sup>33</sup> in which the authors predicted strength of basalt fibre-reinforced geopolymer SCC. The study consisted of both experimental investigation and modelling using ANN for structural assessment of concrete beams. The developed model predicted the ultimate moment capacity of beams with 99% accuracy. Also, Saha et al.<sup>67</sup> utilized ANN and multivariate regression analysis (MRA) algorithms to predict CS of SCC reinforced with different fibres including glass and polypropylene fibres. The author assessed the accuracy of the algorithms based on correlation coefficient and concluded that ANN can predict CS with a correlation of 0.9919 between actual and predicted values and ANN is more accurate having low RMSE than RMA. Also, Pakzad et al.<sup>21</sup> predicted CS of SCC modified with steel fibres using a data set of 176 instances. The study used several ML algorithms including ANN, convolution neural networks (CNN), support vector regression (SVR) etc. The models were developed having 9 input variables and only one output variable. The study concluded that CNN is the most robust algorithm having  $R^2 = 0.928$  on the test data set. Similarly, another study<sup>68</sup> incorporated crumb rubber having size between 0.5 and 3.5 mm in place of sand in SCC containing silica fume blended with cement. The study used 63 data points and predicted CS, flexural strength, and elastic modulus by using multivariate regression algorithm. The analysis revealed the excellent performance of the ML algorithm in predicting various F-SCC properties having  $R^2$  for CS equal to 0.931 and  $R^2$  equal to 0.937 and 0.992 for flexural strength and modulus of elasticity respectively. This shows the potential of ML techniques to accurately model different SCC properties. The summary of the related literature is represented in Table 1.

## Research significance

It is evident from Section “Literature review” that ML models like ANN, SVM, RF, and XGB etc. have been frequently utilized to predict different properties of SCC. However, the subject of utilization of ML algorithms like MEP and GEP for C–S prediction of FR-SCC is comparatively unexplored. Thus, this study presents a novel approach to estimate C–S of silica fume-based FR-SCC using MEP and GEP algorithms. The motive behind using MEP and GEP in this study instead of other algorithms like XGB and ANN etc. stems from the fact that these algorithms are regarded as grey-box models in comparison to black-box models like ANN and XGB etc. Although ANN and other algorithms are very famous for their accuracy and robustness, these techniques require the specification of a best model architecture prior to the prediction process and require a large memory<sup>82</sup>. Also,

| S. no. | Algorithm used                                          | Year | Waste materials used                           | Output(s)                                    | Reference |
|--------|---------------------------------------------------------|------|------------------------------------------------|----------------------------------------------|-----------|
| 1      | ANN                                                     | 2016 | Fly ash, Slag, Silica fume, RHA, Limestone     | CS                                           | 10        |
| 2      | DT, XGB, Light Gradient Boosting                        | 2023 | Nano silica, Limestone, Fly ash, Marble powder | CS                                           | 69        |
| 3      | ANN                                                     | 2019 | Fly ash, Slag, Silica fume                     | CS                                           | 9         |
| 4      | ANN                                                     | 2017 | Fly ash                                        | CS                                           | 70        |
| 5      | ANN                                                     | 2011 | Fly ash                                        | CS                                           | 71        |
| 6      | Intelligent rule-based enhanced multiclass, SVM         | 2019 | Fly ash                                        | CS                                           | 72        |
| 7      | SVR, Deep Learning                                      | 2021 | Fly ash                                        | CS, Splitting tensile strength               | 73        |
| 8      | SVM                                                     | 2020 | Fly ash                                        | L-box test, Slump test, V-funnel test, CS    | 74        |
| 9      | Multivariate adaptive regression spline                 | 2018 | Fly ash                                        | Slump test, V-funnel test, L-box test, CS    | 75        |
| 10     | ANN                                                     | 2017 | Fly ash                                        | CS, Slump flow                               | 76        |
| 11     | ANN                                                     | 2011 | Fly ash                                        | CS                                           | 71        |
| 12     | SVR                                                     | 2023 | Fly ash                                        | CS                                           | 77        |
| 13     | GEP                                                     | 2009 | Fly ash                                        | Slump flow, CS, J Ring                       | 78        |
| 14     | Multivariate Regression (MVR)                           | 2020 | Silica fume, Crumb rubber                      | Flexural Strength, CS, Modulus of Elasticity | 79        |
| 15     | Extreme Learning Machine, long short-term memory (LSTM) | 2021 | Slag, Fly ash, Silica fume                     | Slump flow, J Ring                           | 80        |
| 16     | Multilayer perceptron network (MLP), KNN                | 2022 | Fly ash, Slag                                  | CS                                           | 81        |

**Table 1.** Summary of previous related literature.

the large number of hidden layers used in these algorithms (as in case of ANN) makes it almost impossible to develop an empirical relationship between input and output variables. Thus, these algorithms are called “black box” models due to the following reasons<sup>83</sup>:

- Lack of transparency.
- Inability to accurately describe the underlying prediction process.
- Inability to create empirical equations relating input and output parameters.

In contrast, MEP and GEP are often referred to as grey-box models<sup>84</sup> due to the symbolic representation of the underlying prediction process and their ability to express the output as an empirical equation<sup>85</sup>. Also, these techniques do not require any pre-defined optimization of the model architecture, thus requiring less memory and computing power<sup>86</sup>. Therefore, this study is attributed to predicting C–S of FR-SCC using MEP and GEP.

Moreover, despite the fact that FR-SCC offers various advantages over normal concrete, there is a lack of work focusing on the prediction of FR-SCC strength particularly using the two subtypes of GP used in this study called GEP and MEP. Thus, this study is conducted to foster the use of FR-SCC in the industry by providing empirical equations for estimation of its C–S by using MEP and GEP in a comparative manner. The accuracy of models will be compared by using different error metrics and performance indices. The models will also be checked by some external validation criteria and their performance will be compared with regression models too.

## Research methodology

After the problem identification, dataset was compiled to be used for training of the algorithms. This section highlights the working mechanism of both MEP and GEP algorithms and also sheds light on the database collection which was used for training the algorithms. The results of several statistical analyses performed on the gathered database are also represented in Section “Data collection”.

## Prediction models

The two prediction models used in this study are the subtypes of Genetic Programming (GP). GP refers to a type of ML algorithms that use evolutionary rules to create and refine computer programs for solving a problem. Darwin’s principles of natural selection are the basis of genetic programming, and it uses an iterative process to improve and generate increasingly complex programs<sup>87</sup>. The process starts with the formation of a population of programs according to some predefined rules and parameters. These programs are then evaluated using a fitness functions chosen earlier. After the evaluation, the best performing programs are selected to develop new programs. This process repeats over several generations with the goal of reaching to a reliable, accurate and efficient program<sup>78</sup>.

One of the main advantages of GP is that it can discover solutions to problems that humans could not solve. It can find solutions at a much higher pace than humans. This ability makes it an attractive option for solving problems that need a lot of computing power. However, there are certain limitations to it as well. The selection of various fitting parameters of GP for a particular problem maybe challenging and the solution may always not converge. Moreover, it may require a large computing power to find solutions to some problems making it not the most efficient way to find solutions in some cases<sup>88</sup>. But despite these limitations, GP has proven to be a

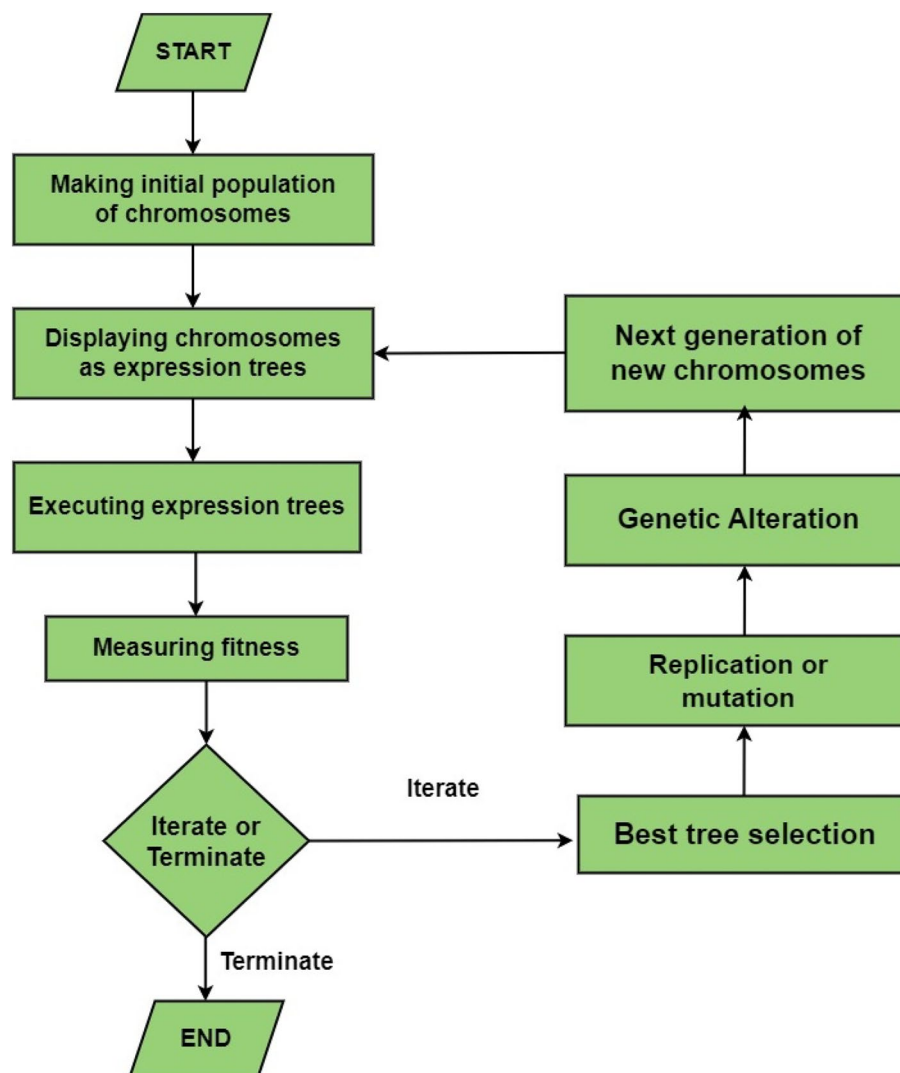
useful tool to solve a wide range of problems in fields like finance, engineering, biology etc. The explanation of two subtypes of GP used in this study is given below:

#### *Gene expression programming (GEP)*

GEP is a subtype of genetic programming used to solve problems by evolving a set of mathematical expressions. The central idea of GEP revolves around using a set of genes which represents a small piece of code or function. The genes are built using a set of function called primitive functions. These functions can be simple arithmetic operations like addition, subtraction etc. or more complex mathematical functions such as sine or cosine etc. These genes are combined using various evolutionary rules to create a complete program<sup>89</sup>. This process of creating the programs by combining different genes is similar to that in human beings. That's why it is called gene expression programming. The process of solving a problem using GEP begins by creating a random population of programs called chromosomes. Then these chromosomes are evaluated using a "fitness function". The good performing chromosomes are selected for the next generation while the worst performing ones are discarded<sup>90</sup>. This process repeats over several generations with the goal of having a population of chromosomes that gives the highest accuracy. The algorithm constantly modifies and improves the chromosomes while running eventually coming at a solution as close to the actual as possible<sup>91</sup>. The flowchart of GEP algorithm is shown in Fig. 2.

#### *Multi expression programming (MEP)*

MEP is a sub technique of GP developed by Oltean<sup>92</sup>. It is an optimization method that involves producing and evolving a set of mathematical expressions to find solution of a problem<sup>93</sup>. MEP offers the advantage of handling a wide variety of problems whether non-linear, multimodal or they have complex constraints. This ability of MEP stems from the fact that it does not need any previous assumptions about the problem on hand, rather it creates a population of mathematical expressions and chooses the best performing expressions to be the solution of the

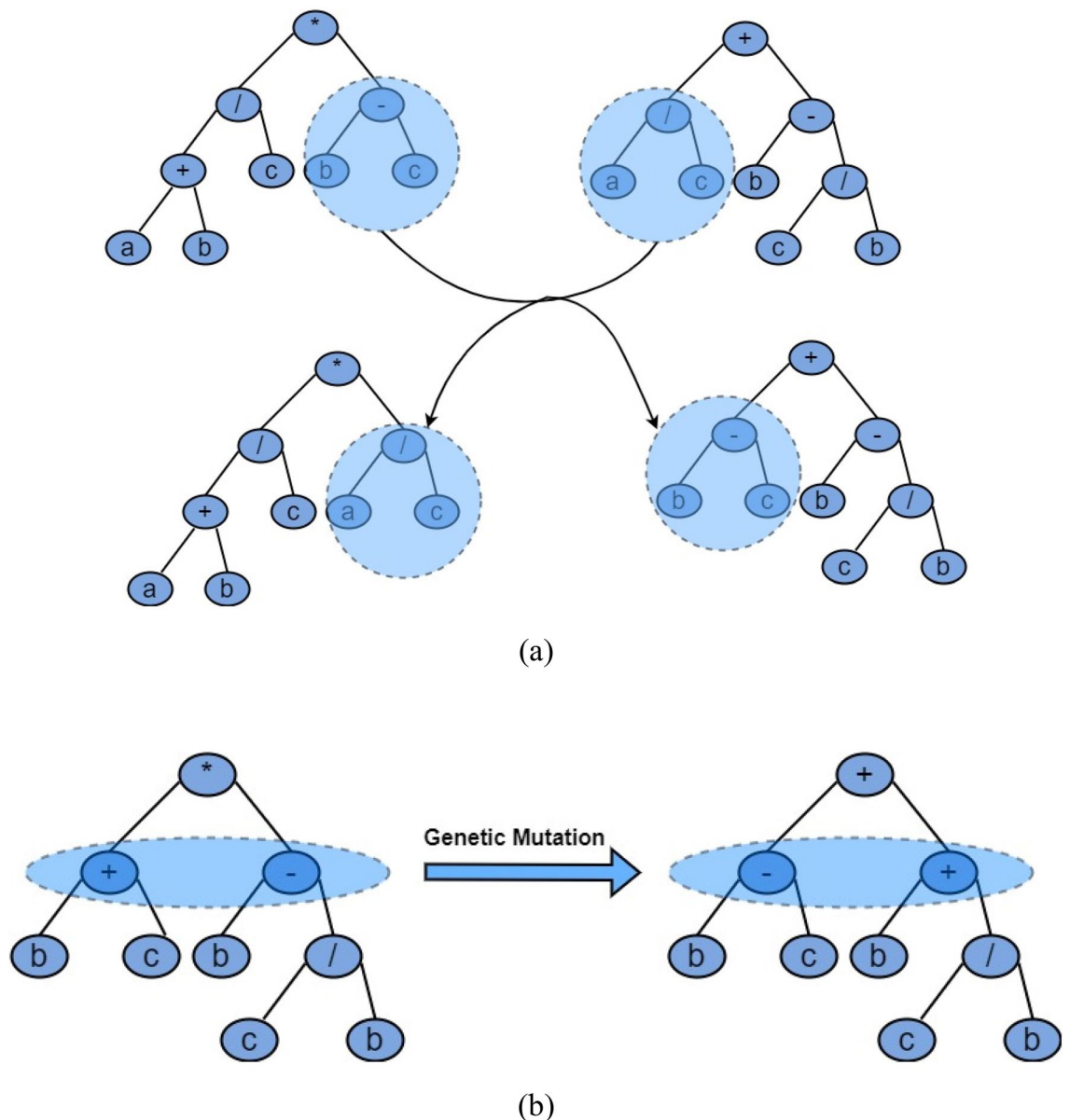


**Figure 2.** Flowchart of gene expression programming.

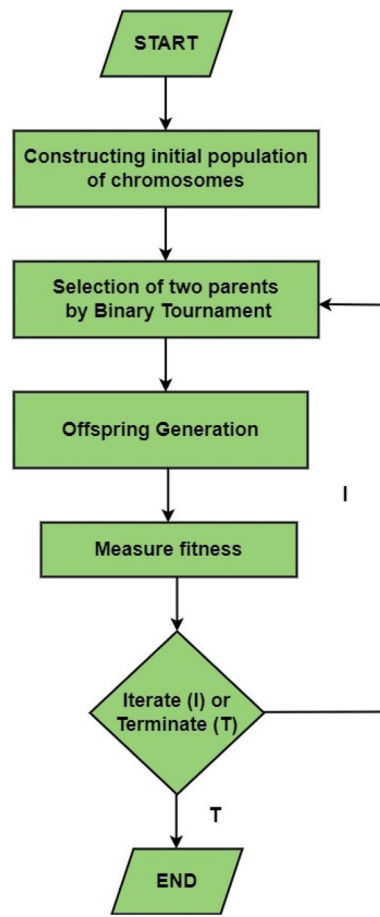
problem using various evolutionary rules<sup>94</sup>. The algorithm starts by creating a set of randomly generated initial expressions. These expressions are then evaluated according to their ability to solve the problem accurately. The expression which performed good are used to develop a new population of expressions. The new population of expressions is developed using evolutionary principles such as mutation and crossover on the initially developed expressions. Crossover means combining two expressions to create a new one while mutation involves altering an existing expression<sup>95</sup>. The expression tree representation of both mutation and crossover is given in Fig. 3. Thus, MEP explores new solutions to the problem and discover potentially better expressions for the representation of the optimization problem through these two processes. This whole process of generating expressions, evaluating them, and using them to create a new population is called an iteration. After the creation of a new generation of expressions, the process continues until the desire accuracy is reached or a predefined criteria is met. This criterion either be a certain fitness value or a specific number of iterations. The flowchart of MEP algorithm is shown in Fig. 4.

### Data collection

The collection of a reliable dataset is the central point of developing a machine learning model. Thus, to develop robust models, a reliable database of 78 points is collected from the internationally published literature given in Table A. Although this dataset is a good starting point to be used for developing predictive models, it is recommended that future studies should consider much larger datasets so that the algorithms can be trained and tested on a variety of data and the resulting models can be used on variable data configurations and distributions. For the selection of most influential parameters to predict C–S, several initial trials were performed, and



**Figure 3.** Representation of (a) genetic crossover, (b) genetic mutation.



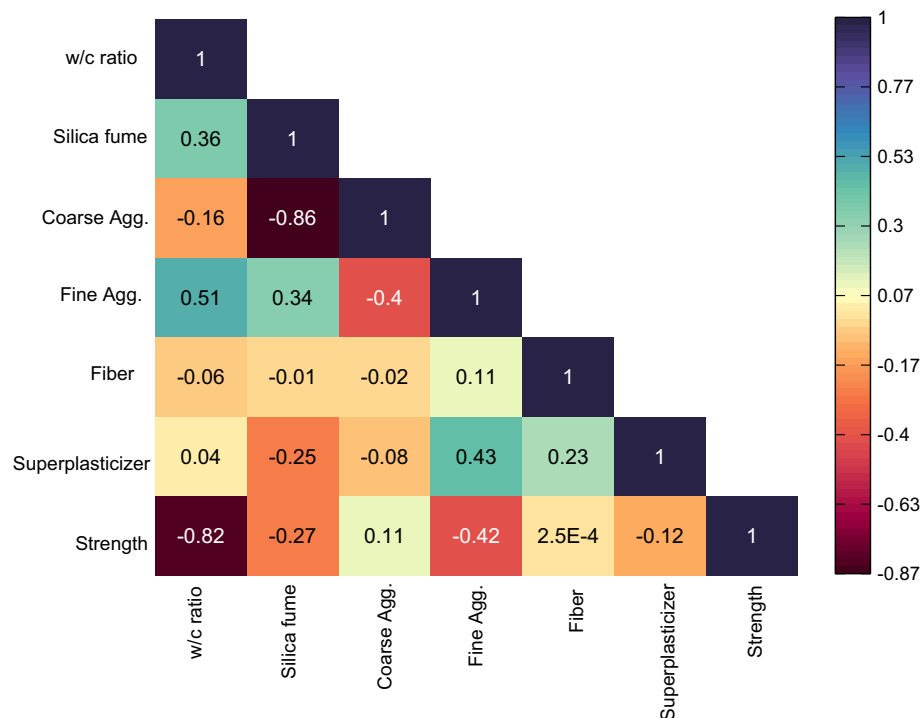
**Figure 4.** Flowchart of multi expression programming.

ultimately the following six input parameters were chosen: water-to-cement ratio, silica fume, coarse aggregate, fine aggregate, fibres, and superplasticizer. These six parameters will be used to predict the single output i.e., C-S. To develop a robust ML model, the data should be splitted in two or three sets<sup>96–98</sup>. Thus, the dataset gathered is split into two parts named training and validation data. Training data (70%) is used to train the model and validation set (30%) is used to test the model's accuracy. This splitting makes sure that the model is not overfitted to the training data and is equally accurate for predicting on the validation data<sup>99</sup>. The statistical description of the variables used to develop the models, along with their units and symbols is represented in Table 2. Notice from the minimum and maximum values of explanatory variables that the dataset is spread over a large range. The spread of data over a large range ensures a more robust and accurate model<sup>100</sup>.

The interdependence between various variables used to develop the models can be checked using two very useful statistical analysis techniques called correlation matrix and scatter matrix shown in Figs. 5 and 6 respectively. A correlation matrix is used to investigate the effect of explanatory variables on each other. It quantifies the relationship between different variables used in the study in the form of coefficient of correlation (R). By looking at the R values between two variables, we can tell the extent to which these variables are related and what effect does the change in one variable have on the other variable. The correlation between any two variables can

|                  | Units | Symbol | Maximum | Minimum | Average | Standard deviation |
|------------------|-------|--------|---------|---------|---------|--------------------|
| W/C ratio        | –     | $x_0$  | 0.739   | 0.39    | 0.52    | 0.125              |
| Silica fume      | kg    | $x_1$  | 978     | 840     | 913.7   | 44.62              |
| Coarse aggregate | kg    | $x_2$  | 870     | 707     | 811.29  | 35.97              |
| Fine aggregate   | kg    | $x_3$  | 190     | 70      | 120.35  | 40.74              |
| Fiber            | kg    | $x_4$  | 1.40    | 0       | 0.730   | 0.57               |
| Superplasticizer | kg    | $x_5$  | 14.7    | 7       | 9.47    | 1.87               |
| Strength         | MPa   | $f_c'$ | 68.5    | 28.24   | 48.10   | 12.9               |

**Table 2.** Description of the dataset used for model development.



**Figure 5.** Correlation matrix of variables.

be positive, negative or zero. Generally, a value above 0.8 indicates the presence of a good correlation between two variables<sup>101</sup>. The interdependence between multiple variables must be checked before developing a model. If most of the variables are strongly correlated to each other, then it might cause complications in the model. This is sometimes referred to as multi-collinearity<sup>102</sup>. Notice from Fig. 5 that strength is affected by all the explanatory variables used in the study. It is strongly correlated with cement and is having the weakest correlation with the fibre content. The correlations between other variables are less than 0.8 mostly so they will not give rise to the problem of multi-collinearity.

Scatter matrix is another tool used widely alongside the correlation matrix. Its main function is to help visualize the bivariate correlations between all the possible combinations of the variables used in the model. It offers useful insights and helps to visualize different combinations of regressions of the variables used<sup>103</sup>. Notice from Fig. 6 that the x-axis and y-axis of scatter matrix are associated with the variables used in the study. It's a grid of different scatter plots and each scatter plot helps to visualize and explore the relationship between any two variables. It can also be used as a useful tool for outlier detection<sup>104</sup>. The frequency distribution histograms of all variables used to build the models are also shown in the diagonal of the scatter matrix. It is recommended to have the distribution of variables as close to the normal distribution as possible since it helps in the development of a reliable model that covers a wide range of variables<sup>105</sup>. Thus, it is useful tool that depicts the regression between variables, distribution of variables and helps in outlier detection simultaneously.

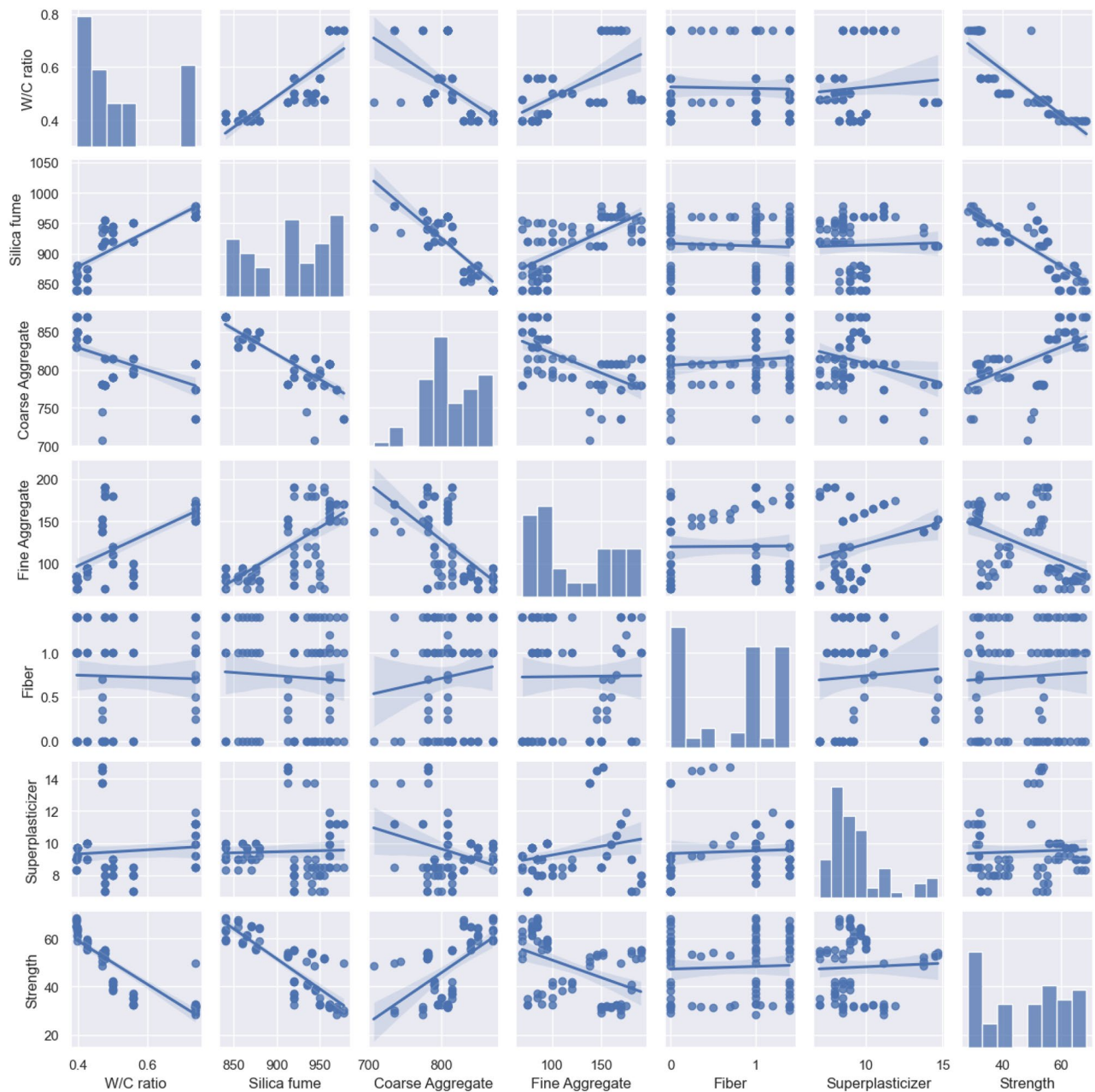
### Performance assessment

The performance of both models will be checked using some statistical error metrics suggested in the literature. These error metrics include mean absolute error (MAE), root mean square error (RMSE), coefficient of determination ( $R^2$ ) and performance index (p) etc. The mathematical expressions to calculate these metrics are given below:

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

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

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



**Figure 6.** Scatter matrix of variables and their frequency distribution.

$$a20 - \text{index} = \frac{N20}{N}$$

$$\text{Performance Index (p)} = \frac{RRMSE}{1 + R}$$

$$\text{Objective Function (OF)} = \left( \frac{N_{\text{Training}} - N_{\text{Validation}}}{N} \right) p_{\text{Training}} + 2 \left( \frac{N_{\text{Validation}}}{N} \right) p_{\text{Validation}}$$

where  $x$  and  $y$  represent actual and predicted C-S values while  $N$  and  $N20$  represents the total samples and number of samples for which the ratio of actual and predicted values lies between 0.80 and 1.20 respectively.

An accurate and reliable model should have a higher correlation between actual and predicted values. This correlation is quantified by the coefficient of determination ( $R^2$ ). The  $R^2$  value ranges from 0 to 1, and value above 0.8 shows an excellent correlation between the actual and predicted values<sup>106</sup>. However,  $R^2$  alone cannot be used as an indication of accuracy of the model. It is due to the inability of  $R^2$  to respond to the output being multiplied or divided with a constant<sup>107</sup>. Thus, the error metrics such as MAE, RMSE have also been identified as

crucial for evaluating the efficiency of any machine learning model<sup>108</sup>. MAE and RMSE both signify the value of average error. RMSE gives more weight to large errors, thus it is used for interpretation of large errors. According to Despotovic et al.<sup>109</sup>, a model is outstanding if RMSE value lies between 0 and 0.11. MAE on the other hand, associates less weight with larger errors and its value is always less than RMSE. It is used as an indication of the average error between the experimental and predicted values in the dataset. The newly developed a20-index is another useful metric to evaluate the performance of ML algorithms. It signifies the number of samples with deviation more than  $\pm 20\%$  from the experimental values<sup>38</sup>. The ideal value of a20-index is 1 and a model must have value of a20 close to 1 to be acceptable<sup>110</sup>. The metrics OF and p can have values between 0 and infinity but a good model must have p and OF value less than 0.2<sup>111</sup>. Performance index offers the advantage of covering both the values of R and relative root mean square error (RRMSE) at the same time. Similarly, OF takes into account the effect of R, RRMSE and the number of data points in training and validation sets. So, it is an indicator of the overall performance of the model. Thus, a model with lower OF indicates overall superior performance. The same criteria were used during trials of model development and the models represented in the study are those that gave the least OF value.

Results and discussions  
Formulation of strength using GEP

A specialized software called GeneXpro Tools version 5.0 was used to employ the GEP algorithm in current study. Before training, the database was split randomly into training and testing sets in 70:30 ratio as specified earlier using the data split option in the software. Also, there are several fitting parameters of GEP algorithm that need to be specified before the commencement of actual model training. This is because each algorithm calls for a specific set of hyperparameters which can only be determined by extensive testing using different set of values. For GEP algorithm, the important parameters include number of genes, head size, functions, number of chromosomes etc. The selection right set of values for these parameters is important since the model's performance is affected by them. For example, head size and number of chromosomes dictate the convergence of the model towards the solution. The GEP parameters used in this study were chosen after extensive testing and after consulting previous studies. The initial parameter values were selected using a previous literature recommendation<sup>1,112</sup> and these values were varied across a wide range of values. The set of values which yielded highest accuracy were used to build the GEP models for C–S prediction in current study are given in Table 3. Notice from Table 3 that the number of genes are selected as 4 which indicates that there will be 4 sub-expression trees generated for C–S prediction. These sub-expression trees will be linked by the specified linking function (addition in our case) to get the final equation for output prediction. Similarly, the head size and number of chromosomes were carefully chosen to not to overfit the model or make it computationally challenging because increasing these parameters beyond a certain point increases the run time of the algorithm and complexity of the model. Also, notice that the set of functions used to be included in the final equation for C–S prediction consist of simple arithmetic functions and square root function. It was deliberately done in order to keep the resultant equations simple and compact for fast computation and implementation.

The size of the population controls the running duration of the program. The model generation was initiated by selecting the number of chromosomes equal to 10. The number of chromosomes and other hyperparameters are varied until the equation with highest accuracy is reached with chromosome number at 20 and head size equal to 4 and 10 respectively. The resulting equation contains the explanatory variables mentioned in Eq. (1).

$$f_c' = (x_0, x_1, x_2, x_3, x_4, x_5) \tag{1}$$

| Parameters            | Settings         |
|-----------------------|------------------|
| GEP parameters        |                  |
| No. of chromosomes    | 20               |
| No. of genes          | 4                |
| Head size             | 10               |
| Linking function      | Addition         |
| Constants per gene    | 10               |
| Functions             | +, −, ×, ÷, sqrt |
| MEP parameters        |                  |
| No. of subpopulations | 200              |
| Subpopulation size    | 500              |
| Code length           | 40               |
| Crossover probability | 0.9              |
| Number of generations | 500              |
| Runs                  | 10               |
| Functions             | +, −, ×, ÷, sqrt |

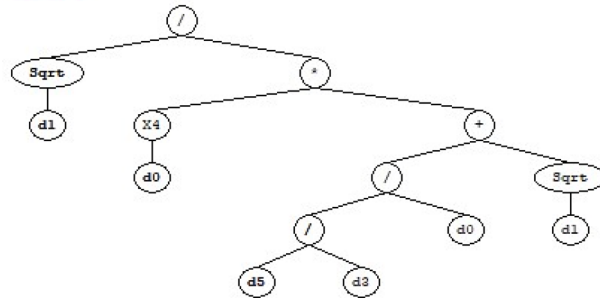
Table 3. Hyperparameters of GEP and MEP model.

The addition function is used as a linking function for GEP model with other arithmetic operation such as multiplication, subtraction, division etc. These functions are chosen to keep the resulting equation simple. The output of GEP algorithm given in the form of expression tree represented in Fig. 7 is decoded to get Eq. (2). The subexpressions from each expression tree are linked by the chosen linking function to get the final equation.

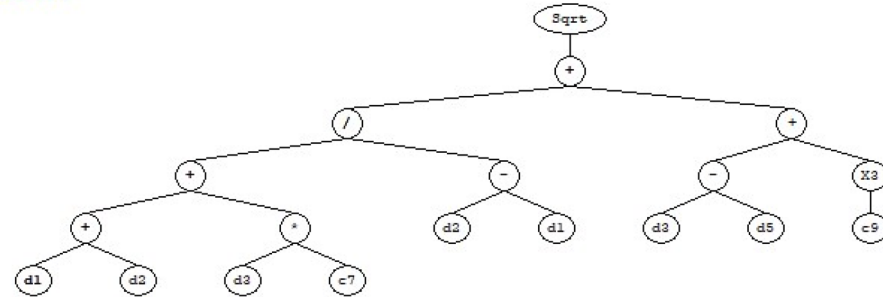
$$f_c' = A + B + C + D \quad (2)$$

where:

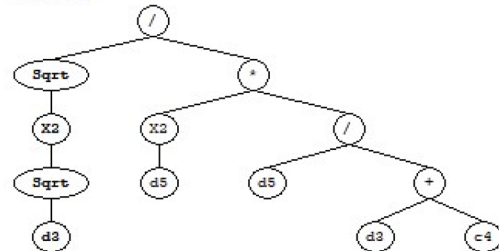
Sub-ET 1



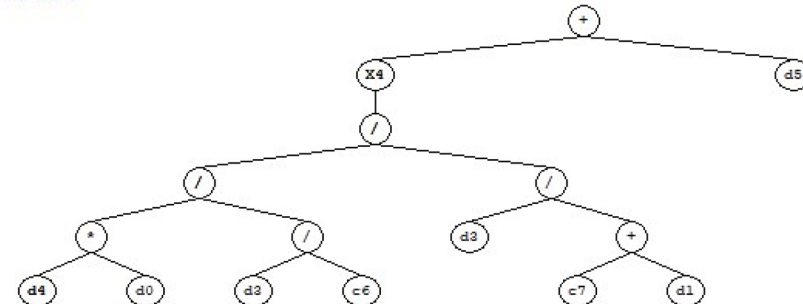
Sub-ET 2



Sub-ET 3



Sub-ET 4



**Figure 7.** Expression tree representation of GEP equation.

$$A = \frac{\sqrt{x_1}}{(x_0)^4 \left\{ \left( \frac{x_0 x_5}{x_3} \right) + \sqrt{x_1} \right\}}$$
$$B = \sqrt{\left( \frac{(x_1 + x_2) - 9.306x_3}{x_2 - x_1} \right) + ((x_3 - x_5) + 143.38)}$$
$$C = \frac{\sqrt{(\sqrt{x_3})^2}}{(x_5)^2 \left( \frac{x_5}{x_3 + 15.45} \right)}$$
$$D = \left( \frac{\frac{(x_0 x_4)}{-0.137}}{\frac{x_3}{x_1 + 827.31}} \right)^4 + x_5$$

Statistical assessment of GEP

The developed models are assessed by calculating the statistical metrics for both data sets. The summary of error metrics of both models is given in Table 4. Notice from the table that the R<sup>2</sup> value of both models is well above the threshold of 0.8. It can also be seen from the scatter plots given in Figs. 8 and 9. Also, the p values are significantly below 0.2 which means that there exists a strong correlation between actual and model predicted values and both models can efficiently predict strength of FR- SCC. Notice that the training MAE of GEP is less than MEP, which implies that GEP did a good job in optimizing to the training data than MEP. Also, the RMSE values are lower for GEP. This is because GEP has lesser predictions with large errors as compared to MEP. It is also indicated by higher a20-index values of GEP. The validation value of a20-index is 1 which means that GEP does not give a single prediction that deviates more than ± 20% from the experimental values. Moreover, the value of objective function which is a useful metric to assess the overall performance of the model is lower for GEP. Thus, it can be inferred that GEP despite having slightly lesser validation R<sup>2</sup> than MEP, proved to be more robust than MEP.

|                | GEP      |            | MEP      |            |
|----------------|----------|------------|----------|------------|
|                | Training | Validation | Training | Validation |
| MAE            | 1.97     | 2.11       | 3.02     | 2.75       |
| RMSE           | 3.06     | 2.69       | 4.74     | 3.58       |
| R <sup>2</sup> | 0.941    | 0.965      | 0.852    | 0.973      |
| p              | 0.032    | 0.028      | 0.049    | 0.040      |
| a20-index      | 0.981    | 1          | 0.945    | 0.956      |
| OF             | 0.0296   |            | 0.031    |            |

Table 4. Error metrics of both models.

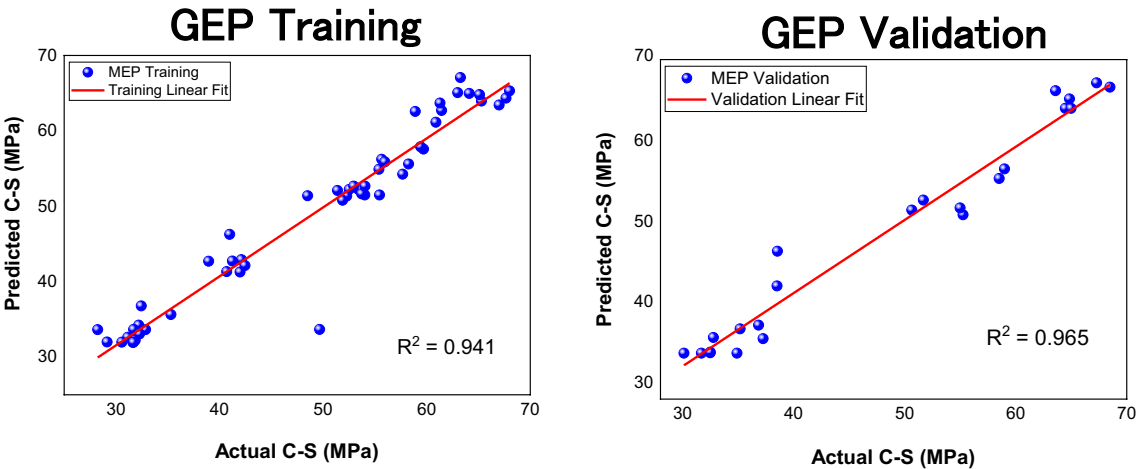
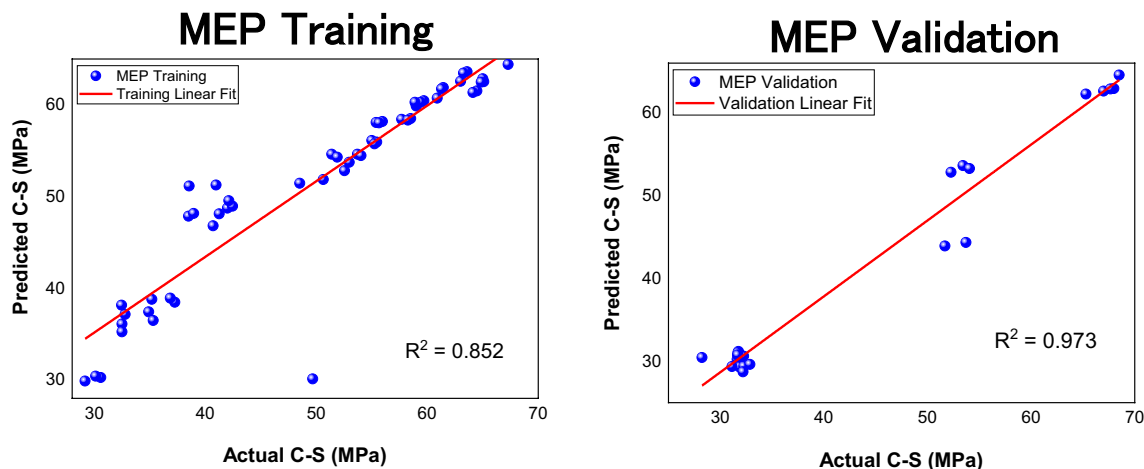


Figure 8. Scatter plot of GEP.



**Figure 9.** Scatter plot of MEP.

### Formulation of strength using MEP

The MEP model development was done using a software known as MEPX 2021.05.18.0. Same as for GEP algorithm, there are several parameter settings that must be done before actual development of a predictive model by MEP. The most important MEP parameters include subpopulation size, number of subpopulations, set of functions, code length etc. The initial values of these parameters were chosen using recommendations from prior studies<sup>113–115</sup>. After having some preliminary values, these parameters were varied across a wide range of values and a trial-and-error approach was followed to reach at the set of parameters which resulted in maximum accuracy of the predictive models. The set of these hyperparameter values which exhibited highest accuracy for C–S prediction are given in Table 3. The number of computer programs developed by the algorithm are dictated by the size of subpopulation. A predictive model having large size of subpopulations can be more accurate and reliable, however it can also be computationally challenging, complex, and time-consuming<sup>116</sup>. Similarly, number of generations are used to indicate the iterations done by the algorithm before it is terminated. Its value was selected as 500 in current study for development of both models so that the algorithm performs a large number of iterations of generating a suitable chromosome for representation of the solution. In the same way, the parameter code length is directly related with the length of the resulting MEP equation. An abnormally larger value of code length may result in the equation being too complex and computationally exhaustive while a smaller value may result in premature convergence of the algorithm and lack of accuracy. Therefore, its value was also chosen carefully by consulting previous studies and considering many possible values before reaching at the values given in Table 3<sup>117</sup>. Moreover, the set of functions to construct the MEP equations involve simple mathematical functions as an attempt to keep the equations simple.

The model development by MEP is an iterative and evolving process trying to optimize the accuracy of the model and complexity of the resulting equation. To find an optimal mode, several initial trials were performed until reaching at one final equation that incorporates the effect of all of the variables and gives the maximum accuracy. The expression given by MEP is a function of the following input parameters:

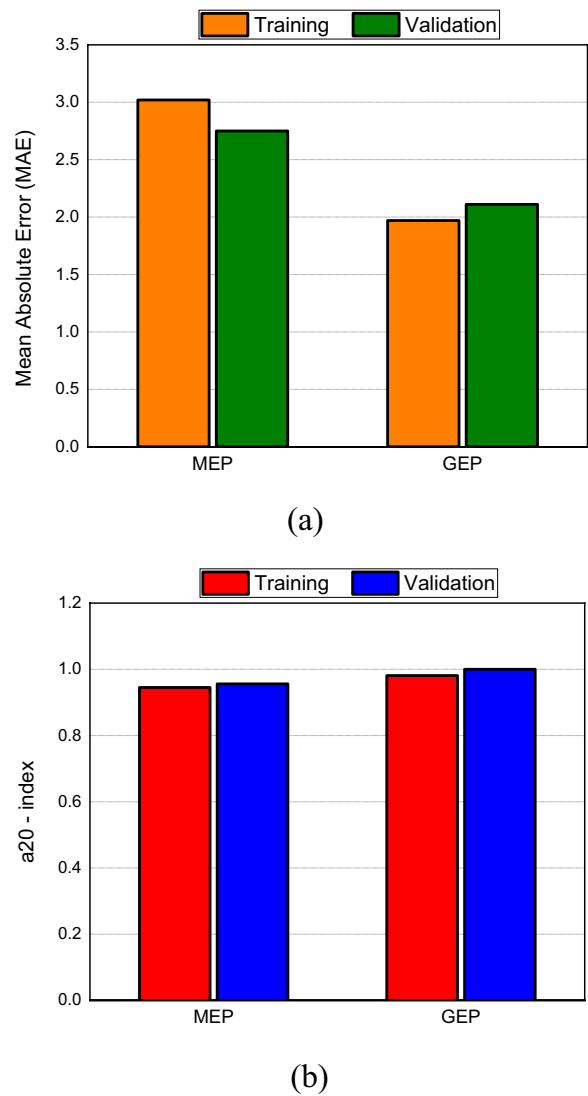
$$f_c' = (x_0, x_1, x_2, x_3, x_4, x_5) \quad (3)$$

The output of the MEP algorithm is given in the form a C++ code. It is decoded to get the expression given by Eq. (4). Notice that the equation is generated by using simple mathematical operations such as addition, subtraction, division, multiplication, and square root.

$$f_c' = \frac{\sqrt{x_2}}{x_0} - (x_0 x_4) - \left( \frac{2x_1 - x_2 - \frac{x_0}{x_5}}{x_3} \right) \quad (4)$$

### Statistical assessment of MEP

The error metrics for MEP are given in Table 4. A comparison has been made between training and validation error metrics of both models and results are shown in Fig. 10. Notice from Fig. 10 (a) that the training MAE value of GEP is less than MEP which means that there is less average deviation in GEP values from actual values. Also, the validation RMSE value of GEP is less than MEP which means MEP has more predictions with larger errors than GEP. It is also clear from training and validation a20 index of MEP. The validation performance index is closer to zero for GEP than MEP, and the value of objective function is also less for GEP. Thus, it is clear that MEP has more deviations in its predictions than GEP and it can be concluded that GEP did a good job in formulating an empirical equation based on the given data that relates all the input variables with the output with great accuracy. Although, the GEP equation is a bit complex and requires more computing effort than MEP equation, its performance overall is better than MEP equation as indicated by objective function and a20-index values. Thus, GEP equation will be considered for further analysis.



**Figure 10.** Errors comparison of both models; (a) MAE; (b) a20-index.

**External validation of MEP and GEP models**

Several external testing criteria are recommended by previous studies to test the developed models. The summary of these validation checks along with their recommended range in the literature is given in Table 4. The inclination of the regression lines passing over the origin is depicted by the value of  $s$  or  $s'$ . For a model to be accurate, both values must be close to 1. Similarly, the squared coefficient  $R_o^2$  and  $R_o'^2$  should also be close to 1 for an accurate model<sup>85</sup>. The values of MEP and GEP equations for these validation checks are also shown in Table 5. Notice from the table that both the models satisfy the external validation checks, and their values lie in

| Expression                                                                                              | Criteria           | MEP   | GEP  | Reference |
|---------------------------------------------------------------------------------------------------------|--------------------|-------|------|-----------|
| $s = \frac{\sum_{i=1}^n (e_i \times m_i)}{\sum_{i=1}^n (e_i^2)}$                                        | $0.85 < s < 1.15$  | 0.99  | 0.99 | 118       |
| $s' = \frac{\sum_{i=1}^n (e_i \times m_i)}{\sum_{i=1}^n (m_i^2)}$                                       | $0.85 < s' < 1.15$ | 0.99  | 0.99 | 118       |
| $R_m = R^2 \times (1 - \sqrt{ R^2 - R_o^2 })$<br>where                                                  | $R_m > 0.5$        | 0.515 | 0.55 | 119       |
| $R_o^2 = 1 - \frac{\sum_{i=1}^n (m_i - e_i^o)^2}{\sum_{i=1}^n (m_i - m_i^o)^2}, e_i^o = s \times m_i$   | $R_o^2 \approx 1$  | 0.98  | 0.99 | 120       |
| $R_o'^2 = 1 - \frac{\sum_{i=1}^n (e_i - m_i^o)^2}{\sum_{i=1}^n (m_i - m_i^o)^2}, m_i^o = s' \times e_i$ | $R_o'^2 \approx 1$ | 0.99  | 0.99 | 121       |

**Table 5.** External validation of MEP and GEP model.

the recommended range for both models. This means the equations developed by MEP and GEP will accurately predict C–S of FR-SCC.

### Comparison with linear and non-linear regression models

This study used six explanatory variables for strength prediction of FR-SCC. To date, no empirical equation has been devised using GEP or MEP for strength estimation of FR-SCC that considers the same set of input variables. So, it is necessary to compare the results of MEP and GEP with Linear Regression (LR) and simple Non-linear Regression (NLR) models developed with the same set of input variables using the same dataset. Thus, the LR and NLR equations for calculating C–S of FR-SCC are given by Eqs. (5) and (6). It is evident from the series plot between actual, NLR and LR predicted values given in Fig. 11 that LR and NLR does not perform well on the dataset. The average error of LR and NLR is 3.2 and 4.2 respectively compared to 2.11 of GEP and 2.75 of MEP. Similarly, the RMSE value of LR and NLR are greater than that of MEP and GEP. Also, NLR predicted values have a lesser correlation with actual values and it failed to accurately predict strength at several points and gave predictions with average error as large as 25. Also, LR and NLR have some other disadvantages as well. The simple regression models assume some pre-defined equations and normality of the residuals for making predictions<sup>122</sup>. Thus, simple LR and NLR are not suitable for modelling the relationship between strength and the input variables used in this study. This further reinforces the importance of using machine learning techniques like MEP and GEP to capture the complex multi non-linear relationships between input and output variables.

$$f_c' = 270.55 - 51.2x_0 - 0.2x_1 - 0.03x_2 + 0.07x_3 + 0.14x_4 + 0.23x_5 \quad (5)$$

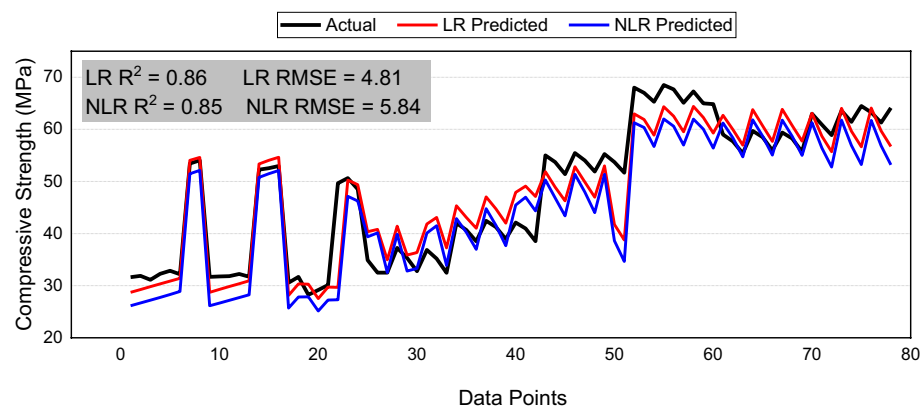
$$f_c' = 187.8 - 34x_0^2 - 0.00014x_1^2 - (3.2 \times 10^{-5})x_2^2 + 0.0003x_3^2 - 0.04x_4^2 + 0.005x_5^2 \quad (6)$$

### Best model selection

Based on the above discussion, LR and simple NLR displayed less accuracy as compared to MEP and GEP. Out of the two evolutionary techniques used, GEP performed better than MEP. The comparison between actual, algorithm prediction and regression predicted values can be well understood by a Taylor Diagram<sup>123</sup> as shown in Fig. 12. It offers the significant advantage of simultaneously using RMSE, coefficient of correlation and standard deviation of the predicted values to draw comparison between different models. The accuracy of various models is checked by computing their distance from the target point of experimental data. It can be seen from Taylor diagram that GEP performs well in terms of RMSE and correlation, and it is closer to the reference data line. In contrast, the MEP shows larger RMSE and lesser correlation and is placed away from reference line. Moreover, LR and simple NLR has lowest correlation and largest RMSE and are located farther away from actual data. Hence, GEP comes turns out to be the most accurate followed by MEP. Thus, the overall order of model accuracy is: GEP > MEP > LR > NLR. Therefore, the sensitivity analysis will be performed on GEP equation as it is more accurate and have lesser errors than MEP, LR and NLR.

### Overfitting of models

A common problem that can arise when dealing with mathematical modelling is overfitting. It happens when an algorithm fits well to the data it is trained on but cannot maintain its accuracy when exposed to new data<sup>124</sup>. It is important to assess a model against the possible issue of overfitting. A model can be checked for overfitting by drawing a comparison between the error metrics of its training and validation data<sup>125</sup>. If a model gives predictions with large errors on validation data, it implies that the model overfitted to the training data and is not suitable for unseen data and the model is not a generalized one. Regarding the models developed in the current study, it is evident from Table 4 and Fig. 9 that the validation error metrics for both models doesn't significantly vary from the training ones. It means that the model maintains its accuracy when it is used for making predictions



**Figure 11.** Series plot of actual, linear, and non-linear regression predictions.

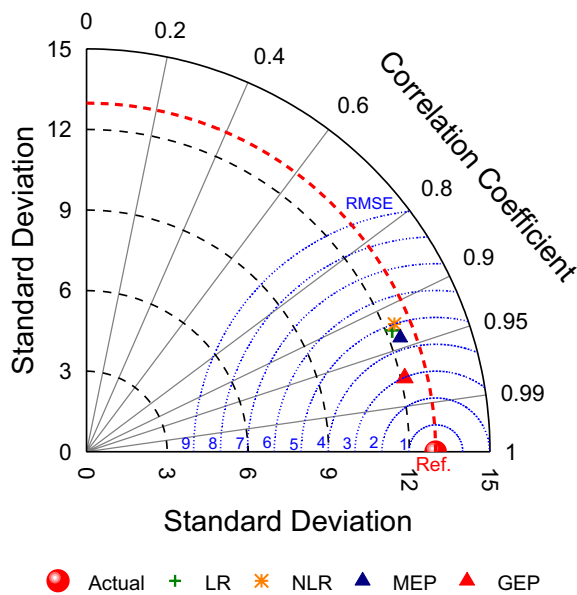


Figure 12. Taylor diagram to compare model and regression predicted values.

on new data. Therefore, the developed equations by MEP and GEP have good generalization capabilities and can be practically used to forecast C–S of FR-SCC.

Sensitivity analysis (SA)

Since the GEP model performed well than its counterpart, sensitivity analysis is performed on the GEP equation to check the impact of input variables on output. SA tells about the sensitivity of the output to the uncertainties in the input data. Higher SA value of a variable, higher the effect it has on the output<sup>96</sup>. It is done on the GEP mode using Eqs. (7) and (8). The results of sensitivity analysis are shown in Fig. 13.

$$N_i = f_{max}(x_i) - f_{min}(x_i) \tag{7}$$

$$SA = \frac{N_i}{\sum_{j=1}^n N_j} \tag{8}$$

where  $x_i = i$  th input variable with all other variables constant,  $f_{max}(x_i)$  = maximum predicted output,  $f_{min}(x_i)$  = minimum predicted output.

The results indicate that all input variables contribute to in determine the C–S of FR-SCC. However, water-to-cement ratio is the crucial factor to predict strength (60%), followed by superplasticizer (12%) and silica fume (10%). Other variables like fine aggregate (8.27%), fibres (6.77%) and coarse aggregate (2.26%) have

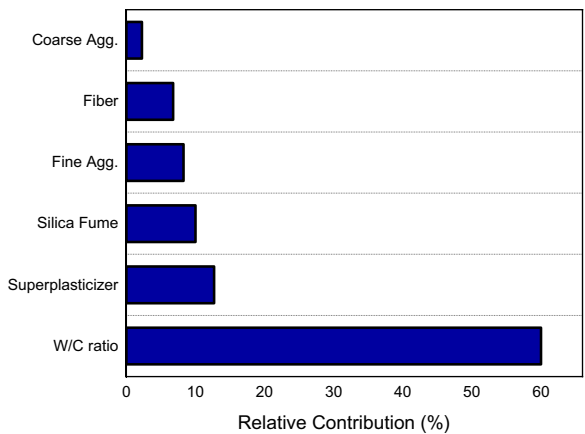


Figure 13. Relative contribution of input variables.

comparatively less contribution in predicting the strength. The result of SA depends upon several factors such as data points used to construct the models and the number of input variables<sup>126</sup>. In relation to this, several studies reported that water-to-cement is most important in prediction of SCC strength<sup>127,128</sup>. It is also evident from the strong correlation between water-to-cement ratio and strength given in Fig. 4. These studies also highlight the relative importance of superplasticizer in the output prediction. Ahmed et al.<sup>129</sup> reported the less effective role of coarse aggregate towards predicting the strength. Moreover, the lesser contribution of fibres and fine aggregate also aligns with the findings of previous research<sup>41,96,127</sup>.

## Conclusions

This study aimed to provide empirical equations for estimating C–S of FR-SCC using MEP and GEP using a dataset collected from internationally published literature to foster the use of FR-SCC in construction industry. The following conclusions were drawn from this research:

- Both MEP and GEP algorithms expressed their output as an empirical equation for computing C–S and the error evaluation revealed that the performance of both algorithms is satisfactory as per the criteria suggested in the literature.
- The GEP algorithm outperformed MEP having training  $R^2 = 0.941$  compared to 0.85 of MEP. Also, GEP yielded lesser predictions with larger errors as depicted by RMSE value of 2.69 compared to 3.58 of MEP.
- The accuracy of the models was also checked by employing external validation checks and the results indicated that both models are accurate. The comparison of MEP and GEP with LR and NLR models also suggests the same.
- The GEP equation was considered for sensitivity analysis and the results showed that water-to-cement ratio, superplasticizer, and silica fume are more important parameters to predict strength having contributions 60%, 12% and 10% respectively. However, fine aggregate (8.27%), fibres (6.77%) and coarse aggregate (2.26%) have relatively less contribution.

## Data availability

The data used to develop the models is provided in the article.

Received: 4 December 2023; Accepted: 25 June 2024

Published online: 27 July 2024

## References

1. Shahmansouri, A. A., Akbarzadeh Bengar, H. & Ghanbari, S. Compressive strength prediction of eco-efficient GGBS-based geopolymer concrete using GEP method. *J. Build. Eng.* <https://doi.org/10.1016/j.jobbe.2020.101326> (2020).
2. Benhelal, E., Zahedi, G., Shamsaei, E. & Bahadori, A. Global strategies and potentials to curb CO<sub>2</sub> emissions in cement industry. *J. Clean. Prod.* **51**, 142–161. <https://doi.org/10.1016/j.jclepro.2012.10.049> (2013).
3. Mikulčić, H., Klemenš, J. J., Vujanović, M., Urbanec, K. & Duić, N. Reducing greenhouse gases emissions by fostering the deployment of alternative raw materials and energy sources in the cleaner cement manufacturing process. *J. Clean. Prod.* **136**, 119–132. <https://doi.org/10.1016/j.jclepro.2016.04.145> (2016).
4. Rahla, K. M., Mateus, R. & Bragança, L. Comparative sustainability assessment of binary blended concretes using supplementary cementitious materials (SCMs) and ordinary portland cement (OPC). *J. Clean. Prod.* **220**, 445–459. <https://doi.org/10.1016/j.jclepro.2019.02.010> (2019).
5. Okamura, H. & Ouchi, M. *Self-Compacting Concrete* (Springer, 2003).
6. Brouwers, H. J. H. & Radix, H. J. Self-compacting concrete: Theoretical and experimental study. *Cem. Concr. Res.* **35**(11), 2116–2136. <https://doi.org/10.1016/j.cemconres.2005.06.002> (2005).
7. Kamal, M. M., Safan, M. A., Bashandy, A. A. & Khalil, A. M. Experimental investigation on the behavior of normal strength and high strength self-curing self-compacting concrete. *J. Build. Eng.* **16**, 79–93. <https://doi.org/10.1016/j.jobbe.2017.12.012> (2018).
8. De Schutter, G. *Self-Compacting Concrete* (Whittles Pub, 2008).
9. Asteris, P. G. & Kolovos, K. G. Self-compacting concrete strength prediction using surrogate models. *Neural Comput. Appl.* **31**, 409–424. <https://doi.org/10.1007/s00521-017-3007-7> (2019).
10. Asteris, P. G., Kolovos, K. G., Douvika, M. G. & Roinos, K. Prediction of self-compacting concrete strength using artificial neural networks. *Eur. J. Environ. Civ. Eng.* **20**, s102–s122. <https://doi.org/10.1080/19648189.2016.1246693> (2016).
11. Ofuyatan, O. M., Adeniyi, A. G. & Ighalo, J. O. Evaluation of fresh and hardened properties of blended silica fume self-compacting concrete (SCC). *Res. Eng. Struct. Mater.* **7**(2), 211–223. <https://doi.org/10.17515/resm2020.228ma1023> (2021).
12. Yang, Z., Liu, S., Yu, L. & Xu, L. A comprehensive study on the hardening features and performance of self-compacting concrete with high-volume fly ash and slag. *Materials* **14**, 15. <https://doi.org/10.3390/ma14154286> (2021).
13. Sun, G., Kong, G., Liu, H. & Amenuvor, A. C. Vibration velocity of X-section cast-in-place concrete (XCC) pile–raft foundation model for a ballastless track. *Can. Geotech. J.* **54**(9), 1340–1345. <https://doi.org/10.1139/CGJ-2015-0623> (2017).
14. Jiang, Y. et al. Mechanical properties and acoustic emission characteristics of soft rock with different water contents under dynamic disturbance. *Int. J. Coal Sci. Technol.* **11**(1), 1–14. <https://doi.org/10.1007/S40789-024-00682-0> (2024).
15. Bani Ardalan, R., Joshaghani, A. & Hooton, R. D. Workability retention and compressive strength of self-compacting concrete incorporating pumice powder and silica fume. *Constr. Build. Mater.* **134**, 116–122. <https://doi.org/10.1016/j.conbuildmat.2016.12.090> (2017).
16. Choudhary, R., Gupta, R. & Nagar, R. Impact on fresh, mechanical, and microstructural properties of high strength self-compacting concrete by marble cutting slurry waste, fly ash, and silica fume. *Constr. Build. Mater.* <https://doi.org/10.1016/j.conbuildmat.2019.117888> (2020).
17. Choudhary, R., Gupta, R., Nagar, R. & Jain, A. Mechanical and abrasion resistance performance of silica fume, marble slurry powder, and fly ash amalgamated high strength self-consolidating concrete. *Constr. Build. Mater.* <https://doi.org/10.1016/j.conbuildmat.2020.121282> (2021).
18. Hu, D. et al. Experiment and application of NATM tunnel deformation monitoring based on 3D laser scanning. *Struct. Control Health Monit.* <https://doi.org/10.1155/2023/3341788> (2023).

19. Sun, L. *et al.* Experimental investigation on the bond performance of sea sand coral concrete with FRP bar reinforcement for marine environments. *Adv. Struct. Eng.* **26**(3), 533–546. <https://doi.org/10.1177/13694332221131153> (2023).
20. Wang, L. *et al.* Effect of long reaction distance on gas composition from organic-rich shale pyrolysis under high-temperature steam environment. *Int. J. Coal Sci. Technol.* **11**(1), 1–18. <https://doi.org/10.1007/S40789-024-00689-7/TABLES/2> (2024).
21. Pakzad, S. S., Roshan, N. & Ghalehnovi, M. Comparison of various machine learning algorithms used for compressive strength prediction of steel fiber-reinforced concrete. *Sci. Rep.* **13**, 1. <https://doi.org/10.1038/s41598-023-30606-y> (2023).
22. Ramkumar, K. B., Kannan Rajkumar, P. R. & Gunasekaran, K. Performance of hybrid steel fiber-reinforced self-compacting concrete RC beam under flexure. *Eng. Sci. Technol. Int. J.* <https://doi.org/10.1016/j.jestch.2023.101432> (2023).
23. Ramkumar, K. B., Kannan Rajkumar, P. R., Noor Ahmmad, S. & Jegan, M. A review on performance of self-compacting concrete: use of mineral admixtures and steel fibres with artificial neural network application. *Constr. Build. Mater.* <https://doi.org/10.1016/J.CONBUILDMAT.2020.120215> (2020).
24. Behbahani, H. P., Nematollahi, B. & Farasatpour, M. *Steel Fiber Reinforced Concrete: A Review* (Springer, 2023).
25. Wu, Z., Henri Khayat, K. & Shi, C. *How Do Fiber Shape and Matrix Composition Affect Fiber Pullout Behavior and Flexural Properties of UHPC?* (Springer, 2018).
26. Sobhani, J., Najimi, M., Pourkhorshidi, A. R. & Parhizkar, T. Prediction of the compressive strength of no-slump concrete: A comparative study of regression, neural network and ANFIS models. *Constr. Build. Mater.* **24**(5), 709–718. <https://doi.org/10.1016/j.conbuildmat.2009.10.037> (2010).
27. Liu, C. *et al.* The role of TBM asymmetric tail-grouting on surface settlement in coarse-grained soils of urban area: Field tests and FEA modelling. *Tunnel. Undergr. Space Technol.* <https://doi.org/10.1016/J.TUST.2021.103857> (2021).
28. Li, Z. *et al.* Ternary cementless composite based on red mud, ultra-fine fly ash, and GGBS: Synergistic utilization and geopolymerization mechanism. *Case Stud. Constr. Mater.* <https://doi.org/10.1016/J.CSCM.2023.E02410> (2023).
29. Saberi, F. & Hosseini-Barzi, M. Effect of thermal maturation and organic matter content on oil shale fracturing. *Int. J. Coal Sci. Technol.* **11**(1), 1–19. <https://doi.org/10.1007/S40789-024-00666-0/FIGURES/16> (2024).
30. Wasim, M., Duc Ngo, T. & Law, D. Durability performance of reinforced waste-based geopolymer foam concrete under exposure to various corrosive environments. *Case Stud. Constr. Mater.* <https://doi.org/10.1016/j.cscm.2021.e00703> (2021).
31. Dahou, Z., Mehdi Sbartaï, Z., Castel, A. & Ghomari, F. Artificial neural network model for steel-concrete bond prediction. *Eng. Struct.* **31**(8), 1724–1733. <https://doi.org/10.1016/j.engstruct.2009.02.010> (2009).
32. Rahman, S. K. & Al-Ameri, R. Experimental investigation and artificial neural network based prediction of bond strength in self-compacting geopolymer concrete reinforced with basalt FRP bars. *Appl. Sci.* **11**, 11. <https://doi.org/10.3390/app11114889> (2021).
33. Rahman, S. K. & Al-Ameri, R. Structural assessment of Basalt FRP reinforced self-compacting geopolymer concrete using artificial neural network (ANN) modelling. *Constr. Build. Mater.* <https://doi.org/10.1016/j.conbuildmat.2023.132464> (2023).
34. Nunez, I., Marani, A. & Nehdi, M. L. Mixture optimization of recycled aggregate concrete using hybrid machine learning model. *Materials* **13**(19), 1–24. <https://doi.org/10.3390/ma13194331> (2020).
35. Alyami, M. *et al.* Application of metaheuristic optimization algorithms in predicting the compressive strength of 3D-printed fiber-reinforced concrete. *Dev. Built Environ.* **17**, 100307. <https://doi.org/10.1016/j.dibe.2023.100307> (2024).
36. Cao, J. *et al.* Crack detection in ultrahigh-performance concrete using robust principal component analysis and characteristic evaluation in the frequency domain. *Struct. Health Monit.* **23**(2), 1013–1024. <https://doi.org/10.1177/14759217231178457> (2024).
37. Cui, D. *et al.* Dynamic splitting performance and energy dissipation of fiber-reinforced concrete under impact loading. *Materials* **17**, 2. <https://doi.org/10.3390/ma17020421> (2024).
38. Asteris, P. G., Koopialipoor, M., Armaghani, D. J., Kotsonis, E. A. & Lourenço, P. B. Prediction of cement-based mortars compressive strength using machine learning techniques. *Neural Comput. Appl.* **33**(19), 13089–13121. <https://doi.org/10.1007/S00521-021-06004-8> (2021).
39. Emad, W., Mohammed, A. & Kurda, R. Structures, and undefined 2022. Prediction of concrete materials compressive strength using surrogate models. <https://www.sciencedirect.com/science/article/pii/S2352012422010414>. Accessed 12 May 2024.
40. Long, X., Mao, M. H., Su, T. X., Su, Y. T. & Tian, M. K. Machine learning method to predict dynamic compressive response of concrete-like material at high strain rates. *Defence Technol.* **23**, 100–111. <https://doi.org/10.1016/J.DT.2022.02.003> (2023).
41. Huang, H., Huang, M., Zhang, W., Pospisil, S. & Wu, T. Experimental investigation on rehabilitation of corroded RC columns with BSP and HPFL under combined loadings. *J. Struct. Eng.* **146**, 8. [https://doi.org/10.1061/\(ASCE\)ST.1943-541X.0002725](https://doi.org/10.1061/(ASCE)ST.1943-541X.0002725) (2020).
42. Sarir, P., Chen, J., Asteris, P. G., Armaghani, D. J. & Tahir, M. M. 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), 1–19. <https://doi.org/10.1007/s00366-019-00808-y> (2021).
43. Memarzadeh, A., Shahmansouri, A. A. & Poologanathan, K. A novel prediction model for post-fire elastic modulus of circular recycled aggregate concrete-filled steel tubular stub columns. *Steel Compos. Struct.* **44**(3), 309–324. <https://doi.org/10.12989/SCS.2022.44.3.309> (2022).
44. Ly, H. B. *et al.* Estimation of axial load-carrying capacity of concrete-filled steel tubes using surrogate models. *Neural Comput. Appl.* **33**(8), 3437–3458. <https://doi.org/10.1007/S00521-020-05214-W> (2021).
45. Le, T. T., Asteris, P. G. & Lemonis, M. E. Prediction of axial load capacity of rectangular concrete-filled steel tube columns using machine learning techniques. *Eng. Comput.* **38**, 3283–3316. <https://doi.org/10.1007/S00366-021-01461-0> (2022).
46. Asteris, P. G., Douvika, M. G. & Karamani, C. Prediction of mechanical characteristics of soilcrete materials using artificial neural networks. (2017). <https://www.researchgate.net/publication/324685061>.
47. Wang, H. L. & Yin, Z. Y. High performance prediction of soil compaction parameters using multi expression programming. *Eng. Geol.* <https://doi.org/10.1016/j.enggeo.2020.105758> (2020).
48. Soleimani, F., Si, G., Roshan, H. & Zhang, J. Numerical modelling of gas outburst from coal: A review from control parameters to the initiation process. *Int. J. Coal Sci. Technol.* **10**, 1. <https://doi.org/10.1007/S40789-023-00657-7> (2023).
49. Zhang, C., Wang, P., Wang, E., Chen, D. & Li, C. Characteristics of coal resources in China and statistical analysis and preventive measures for coal mine accidents. *Int. J. Coal Sci. Technol.* **10**(1), 1–13. <https://doi.org/10.1007/S40789-023-00582-9/FIGURES/13> (2023).
50. Ali, Z., Karakus, M., Nguyen, G. D. & Amrouch, K. Effect of loading rate and time delay on the tangent modulus method (TMM) in coal and coal measured rocks. *Int. J. Coal Sci. Technol.* **9**(1), 1–13. <https://doi.org/10.1007/S40789-022-00552-7/FIGURES/16> (2022).
51. Wang, M. *et al.* Sulfate diffusion in coal pillar: experimental data and prediction model. *Int. J. Coal Sci. Technol.* **10**(1), 1–12. <https://doi.org/10.1007/S40789-023-00575-8/FIGURES/12> (2023).
52. Wang, G. *et al.* Research and practice of intelligent coal mine technology systems in China. *Int. J. Coal Sci. Technol.* **9**(1), 1–17. <https://doi.org/10.1007/S40789-022-00491-3/FIGURES/13> (2022).
53. Cai, Y. *et al.* A review of monitoring, calculation, and simulation methods for ground subsidence induced by coal mining. *Int. J. Coal Sci. Technol.* **10**(1), 1–23. <https://doi.org/10.1007/S40789-023-00595-4> (2023).
54. Qi, Q., Yue, X., Duo, X., Xu, Z. & Li, Z. Spatial prediction of soil organic carbon in coal mining subsidence areas based on RBF neural network. *Int. J. Coal Sci. Technol.* **10**(1), 1–13. <https://doi.org/10.1007/S40789-023-00588-3/TABLES/4> (2023).

55. Ma, D., Duan, H., Zhang, J. & Bai, H. A state-of-the-art review on rock seepage mechanism of water inrush disaster in coal mines. *Int. J. Coal Sci. Technol.* **9**(1), 1–28. <https://doi.org/10.1007/S40789-022-00525-W> (2022).
56. Wu, H. *et al.* Stability analysis of rib pillars in highwall mining under dynamic and static loads in open-pit coal mine. *Int. J. Coal Sci. Technol.* <https://doi.org/10.1007/S40789-022-00504-1> (2022).
57. Huang, F. *et al.* Slope stability prediction based on a long short-term memory neural network: Comparisons with convolutional neural networks, support vector machines and random forest models. *Int. J. Coal Sci. Technol.* **10**(1), 1–14. <https://doi.org/10.1007/S40789-023-00579-4/FIGURES/5> (2023).
58. Wang, S., Guo, J., Yu, Y., Shi, P. & Zhang, H. Quality evaluation of land reclamation in mining area based on remote sensing. *Int. J. Coal Sci. Technol.* **10**(1), 1–10. <https://doi.org/10.1007/S40789-023-00601-9/TABLES/6> (2023).
59. Qi, C. & Tang, X. Slope stability prediction using integrated metaheuristic and machine learning approaches: A comparative study. *Comput. Ind. Eng.* **118**, 112–122. <https://doi.org/10.1016/j.cie.2018.02.028> (2018).
60. Zhang, K., Wu, X., Niu, R., Yang, K. & Zhao, L. 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. <https://doi.org/10.1007/s12665-017-6731-5> (2017).
61. Ashrafiyan, A., Panahi, E. & Salehi, M. S. On the implementation of the interpretable data-intelligence model for designing service life of structural concrete in a marine environment. <https://www.sciencedirect.com/science/article/pii/S0029801822008927>. Accessed 11 May 2024.
62. Ma, X., Dong, Z., Quan, W., Dong, Y. & Tan, Y. Real-time assessment of asphalt pavement moduli and traffic loads using monitoring data from built-in sensors: Optimal sensor placement and identification algorithm. *Mech. Syst. Signal Process.* **187**, 109930. <https://doi.org/10.1016/j.ymssp.2022.109930> (2023).
63. Farooq, F. *et al.* A comparative study for the prediction of the compressive strength of self-compacting concrete modified with fly ash. *Materials* **14**, 17. <https://doi.org/10.3390/ma14174934> (2021).
64. Abunassar, N., Alas, M. & Ali, S. I. A. Prediction of compressive strength in self-compacting concrete containing fly ash and silica fume using ANN and SVM. *Arab. J. Sci. Eng.* **48**(4), 5171–5184. <https://doi.org/10.1007/s13369-022-07359-3> (2023).
65. Ashrafiyan, A., Panahi, E. & Salehi, S. Mapping the strength of agro-ecological lightweight concrete containing oil palm by-product using artificial intelligence techniques. <https://www.sciencedirect.com/science/article/pii/S2352012422012966>. Accessed 12 May 2024.
66. Ashrafiyan, A., Hamzehkolaei, N. S., Dwijendra, N. K. A. & Yazdani, M. An evolutionary neuro-fuzzy-based approach to estimate the compressive strength of eco-friendly concrete containing recycled construction wastes. *Buildings* **12**, 8. <https://doi.org/10.3390/buildings12081280> (2022).
67. Saha, P., Prasad, M. L. V. & RathishKumar, P. Predicting strength of SCC using artificial neural network and multivariable regression analysis. *Comput. Concr.* **20**(1), 31–38. <https://doi.org/10.12989/cac.2017.20.1.031> (2017).
68. Bušić, R., Benšić, M., Miličević, I. & Strukar, K. Prediction models for the mechanical properties of self-compacting concrete with recycled rubber and silica fume. *Materials* **13**, 8. <https://doi.org/10.3390/MA13081821> (2020).
69. Mai, H. V. T., Nguyen, M. H. & Ly, H. B. Development of machine learning methods to predict the compressive strength of fiber-reinforced self-compacting concrete and sensitivity analysis. *Constr. Build. Mater.* <https://doi.org/10.1016/j.conbuildmat.2023.130339> (2023).
70. Belalia Douma, O., Boukhatem, B., Ghrici, M. & Tagnit-Hamou, A. Prediction of properties of self-compacting concrete containing fly ash using artificial neural network. *Neural Comput. Appl.* **28**, 707–718. <https://doi.org/10.1007/s00521-016-2368-7> (2017).
71. Siddique, R., Aggarwal, P. & Aggarwal, Y. Prediction of compressive strength of self-compacting concrete containing bottom ash using artificial neural networks. *Adv. Eng. Softw.* **42**(10), 780–786. <https://doi.org/10.1016/j.advengsoft.2011.05.016> (2011).
72. Selvaraj, S. & Sivaraman, S. Prediction model for optimized self-compacting concrete with fly ash using response surface method based on fuzzy classification. *Neural Comput. Appl.* **31**(5), 1365–1373. <https://doi.org/10.1007/s00521-018-3575-1> (2019).
73. Kina, C., Turk, K. & Tanyildizi, H. Estimation of strengths of hybrid FR-SCC by using deep-learning and support vector regression models. *Struct. Concr.* **23**(5), 3313–3330. <https://doi.org/10.1002/suco.202100622> (2022).
74. Saha, P., Debnath, P. & Thomas, P. Prediction of fresh and hardened properties of self-compacting concrete using support vector regression approach. *Neural Comput. Appl.* **32**(12), 7995–8010. <https://doi.org/10.1007/s00521-019-04267-w> (2020).
75. Kaveh, A., Bakhshpoori, T. & Hamze-Ziabari, S. M. M5' and mars based prediction models for properties of selfcompacting concrete containing fly ash. *Period. Polytech. Civ. Eng.* **62**(2), 281–294. <https://doi.org/10.3311/PPci.10799> (2018).
76. Abu Yaman, M., Abd Elaty, M. & Taman, M. Predicting the ingredients of self compacting concrete using artificial neural network. *Alexandria Eng. J.* **56**(4), 523–532. <https://doi.org/10.1016/j.aej.2017.04.007> (2017).
77. Web, J., Chen, L. & Jiang, W. *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.*
78. Sonebi, M. & Cevik, A. Genetic programming based formulation for fresh and hardened properties of self-compacting concrete containing pulverised fuel ash. *Constr. Build. Mater.* **23**(7), 2614–2622. <https://doi.org/10.1016/j.conbuildmat.2009.02.012> (2009).
79. Yao, X. *et al.* AI-based performance prediction for 3D-printed concrete considering anisotropy and steam curing condition. *Constr. Build. Mater.* <https://doi.org/10.1016/J.CONBUILDMAT.2023.130898> (2023).
80. Kina, C., Turk, K., Atalay, E., Donmez, I. & Tanyildizi, H. 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), 11641–11659. <https://doi.org/10.1007/s00521-021-05836-8> (2021).
81. Li, H., Yang, Y., Wang, X. & Tang, H. Effects of the position and chloride-induced corrosion of strand on bonding behavior between the steel strand and concrete. *Structures* <https://doi.org/10.1016/J.ISTRUC.2023.105500> (2023).
82. Iqbal, M. F. *et al.* Prediction of mechanical properties of green concrete incorporating waste foundry sand based on gene expression programming. *J. Hazard. Mater.* <https://doi.org/10.1016/j.jhazmat.2019.121322> (2020).
83. Shahin, M., Jaksa, M. B. & Maier, H. R. Physical modeling of rolling dynamic compaction view project artificial neural networks-pile capacity prediction view project (2008). <https://www.researchgate.net/publication/228364758>.
84. Zare Naghadehi, M., Samaei, M., Ranjbarnia, M. & Nourani, V. State-of-the-art predictive modeling of TBM performance in changing geological conditions through gene expression programming. *Measurement* **126**, 46–57. <https://doi.org/10.1016/j.measurement.2018.05.049> (2018).
85. Chu, H. H. *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), 3603–3617. <https://doi.org/10.1016/j.asej.2021.03.018> (2021).
86. Xiao, C. *et al.* Additive manufacturing of high solid content lunar regolith simulant paste based on vat photopolymerization and the effect of water addition on paste retention properties. *Addit. Manuf.* <https://doi.org/10.1016/J.ADDMA.2023.103607> (2023).
87. Crina, M. O. & Grosan, G. A Comparison of Several Linear GP Techniques A Comparison of Several Linear Genetic Programming Techniques (2003). [www.mep.cs.ubbcluj.ro](http://www.mep.cs.ubbcluj.ro).
88. He, H. *et al.* Deciphering size-induced influence of carbon dots on mechanical performance of cement composites. *Constr. Build. Mater.* <https://doi.org/10.1016/J.CONBUILDMAT.2024.136030> (2024).
89. Koza, J. R. & Jacks Hall, M. *Survey of Genetic Algorithms And Genetic Programming.* <http://www-cs-faculty.stanford.edu/~koza/>.

90. Gholampour, A., Gandomi, A. H. & Ozbakkaloglu, T. New formulations for mechanical properties of recycled aggregate concrete using gene expression programming. *Constr. Build. Mater.* **130**, 122–145. <https://doi.org/10.1016/j.conbuildmat.2016.10.114> (2017).
91. Huang, H., Li, M., Yuan, Y. & Bai, H. Theoretical analysis on the lateral drift of precast concrete frame with replaceable artificial controllable plastic hinges. *J. Build. Eng.* <https://doi.org/10.1016/j.jobe.2022.105386> (2022).
92. Oltean, M. *Multi Expression Programming* (2006). [www.cs.ubbcluj.ro/~molteanwww.mep.cs.ubbcluj.ro](http://www.cs.ubbcluj.ro/~molteanwww.mep.cs.ubbcluj.ro).
93. Gandomi, A. H., Faramarzi, A., Rezaee, P. G., Asghari, A. & Talatahari, S. New design equations for elastic modulus of concrete using multi expression programming. *J. Civil Eng. Manag.* **21**(6), 761–774. <https://doi.org/10.3846/13923730.2014.893910> (2015).
94. Zhang, Q., Meng, X., Yang, B. & Liu, W. MREP: Multi-reference expression programming. in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 26–38 (Springer, 2016). [https://doi.org/10.1007/978-3-319-42294-7\\_3](https://doi.org/10.1007/978-3-319-42294-7_3).
95. Liu, T. Q. *et al.* Flexural performance of curved-pultruded GFRP arch beams subjected to varying boundary conditions. *Eng. Struct.* <https://doi.org/10.1016/j.engstruct.2024.117962> (2024).
96. De-Prado-gil, J., Palencia, C., Jagadesh, P. & Martínez-García, R. A Comparison of machine learning tools that model the splitting tensile strength of self-compacting recycled aggregate concrete. *Materials* **15**, 12. <https://doi.org/10.3390/ma15124164> (2022).
97. Huang, H., Li, M., Zhang, W. & Yuan, Y. Seismic behavior of a friction-type artificial plastic hinge for the precast beam-column connection. *Arch. Civ. Mech. Eng.* **22**, 4. <https://doi.org/10.1007/S43452-022-00526-1> (2022).
98. Jiang, Y., Liu, L., Yan, J. & Wu, Z. Room-to-low temperature thermo-mechanical behavior and corresponding constitutive model of liquid oxygen compatible epoxy composites. *Compos. Sci. Technol.* <https://doi.org/10.1016/J.COMPSCITECH.2023.110357> (2024).
99. Gandomi, A. H., Alavi, A. H., Mousavi, M. & Tabatabaei, S. M. A hybrid computational approach to derive new ground-motion prediction equations. *Eng. Appl. Artif. Intell.* **24**(4), 717–732. <https://doi.org/10.1016/j.engappai.2011.01.005> (2011).
100. Wei, J. *et al.* Seismic performance of concrete-filled steel tubular composite columns with ultra high performance concrete plates. *Eng. Struct.* <https://doi.org/10.1016/J.ENGSTRUCT.2022.115500> (2023).
101. Sarveghadi, M., Gandomi, A. H., Bolandi, H. & Alavi, A. H. Development of prediction models for shear strength of SFRCB using a machine learning approach. *Neural Comput. Appl.* **31**(7), 2085–2094. <https://doi.org/10.1007/s00521-015-1997-6> (2019).
102. Iftikhar, B. *et al.* Predictive modeling of compressive strength of sustainable rice husk ash concrete: Ensemble learner optimization and comparison. *J. Clean Prod.* <https://doi.org/10.1016/j.jclepro.2022.131285> (2022).
103. Ware, C. Interacting with visualizations. *Inf. Vis.* **4**, 359–392. <https://doi.org/10.1016/B978-0-12-812875-6.00010-4> (2021).
104. Kononenko, I. & Kukar, M. Data preprocessing. *Mach. Learn. Data Min.* **1**, 181–211. <https://doi.org/10.1533/9780857099440.181> (2007).
105. Groeneveld, R. A. & Meeden, G. *Measuring Skewness and Kurtosis* (1984). <https://about.jstor.org/terms>.
106. Alavi, A. H., Gandomi, A. H., Sahab, M. G. & Gandomi, M. Multi expression programming: A new approach to formulation of soil classification. *Eng. Comput.* **26**(2), 111–118. <https://doi.org/10.1007/s00366-009-0140-7> (2010).
107. Jalal, F. E., Xu, Y., Iqbal, M., Jamhiri, B. & Javed, M. F. Predicting the compaction characteristics of expansive soils using two genetic programming-based algorithms. *Transp. Geotech.* <https://doi.org/10.1016/j.tgeo.2021.100608> (2021).
108. Rostami, A. *et al.* Rigorous framework determining residual gas saturations during spontaneous and forced imbibition using gene expression programming. *J. Nat. Gas Sci. Eng.* <https://doi.org/10.1016/J.JNGSE.2020.103644> (2020).
109. Despotovic, M., Nedic, V., Despotovic, D. & Cvetanovic, S. Evaluation of empirical models for predicting monthly mean horizontal diffuse solar radiation. *Renew. Sustain. Energy Rev.* **56**, 246–260. <https://doi.org/10.1016/J.RSER.2015.11.058> (2016).
110. Li, Y. *et al.* The effects of rock index tests on prediction of tensile strength of granitic samples: A neuro-fuzzy intelligent system. *Sustainability* **13**, 19. <https://doi.org/10.3390/su131910541> (2021).
111. Gandomi, A. H. & Roke, D. A. Assessment of artificial neural network and genetic programming as predictive tools. *Adv. Eng. Softw.* **88**, 63–72. <https://doi.org/10.1016/j.advengsoft.2015.05.007> (2015).
112. Lu, D. *et al.* A dynamic elastoplastic model of concrete based on a modeling method with environmental factors as constitutive variables. *J. Eng. Mech.* **149**, 12. <https://doi.org/10.1061/JENMDT.EMENG-7206> (2023).
113. Xia, X. Optimizing and hyper-tuning machine learning models for the water absorption of eggshell and glass-based cementitious composite. *PLoS ONE* **19**, 1. <https://doi.org/10.1371/journal.pone.0296494> (2024).
114. Bin Inqiad, W. *et al.* Comparative analysis of various machine learning algorithms to predict 28-day compressive strength of self-compacting concrete. *Heliyon* **9**(11), e22036. <https://doi.org/10.1016/j.heliyon.2023.e22036> (2023).
115. Lu, D., Zhou, X., Du, X. & Wang, G. A 3D fractional elastoplastic constitutive model for concrete material. *Int. J. Solids Struct.* **165**, 160–175. <https://doi.org/10.1016/J.IJSSOLSTR.2019.02.004> (2019).
116. Khan, M., Ali, M., Najeh, T. & Gamil, Y. Computational prediction of workability and mechanical properties of bentonite plastic concrete using multi-expression programming. *Sci. Rep.* **14**, 1. <https://doi.org/10.1038/s41598-024-56088-0> (2024).
117. Ismael Jaf, D. K. *et al.* Hybrid nonlinear regression model versus MARS, MEP, and ANN to evaluate the effect of the size and content of waste tire rubber on the compressive strength of concrete. *Heliyon* **10**, 4. <https://doi.org/10.1016/j.heliyon.2024.e25997> (2024).
118. Roy, P. P. & Roy, K. On some aspects of variable selection for partial least squares regression models. *QSAR Comb. Sci.* **27**(3), 302–313. <https://doi.org/10.1002/QSAR.200710043> (2008).
119. Khan, M. A., Zafar, A., Akbar, A., Javed, M. F. & Mosavi, A. Application of gene expression programming (GEP) for the prediction of compressive strength of geopolymer concrete. *Materials* **14**(5), 1–23. <https://doi.org/10.3390/ma14051106> (2021).
120. Huang, H., Yuan, Y., Zhang, W. & Zhu, L. Property assessment of high-performance concrete containing three types of fibers. *Int. J. Concr. Struct. Mater.* **15**, 1. <https://doi.org/10.1186/s40069-021-00476-7> (2021).
121. Lu, D., Wang, G., Du, X. & Wang, Y. A nonlinear dynamic uniaxial strength criterion that considers the ultimate dynamic strength of concrete. *Int. J. Impact Eng.* **103**, 124–137. <https://doi.org/10.1016/J.IJIMPENG.2017.01.011> (2017).
122. Gandomi, A. H., Alavi, A. H., Mirzahosseini, M. R. & Nejad, F. M. Nonlinear genetic-based models for prediction of flow number of asphalt mixtures. *J. Mater. Civil Eng.* **23**(3), 248–263. [https://doi.org/10.1061/\(asce\)mt.1943-5533.0000154](https://doi.org/10.1061/(asce)mt.1943-5533.0000154) (2011).
123. Taylor, K. E. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res. Atmos.* **106**(D7), 7183–7192. <https://doi.org/10.1029/2000JD900719> (2001).
124. Armaghani, D. J. & Asteris, P. G. A comparative study of ANN and ANFIS models for the prediction of cement-based mortar materials compressive strength. *Neural Comput. Appl.* **33**(9), 4501–4532. <https://doi.org/10.1007/s00521-020-05244-4> (2021).
125. Whiteson, S., Tanner, B., Taylor, M. E. & Stone, P. *Protecting Against Evaluation Overfitting in Empirical Reinforcement Learning* (Springer, 2023).
126. Alaskar, A. *et al.* Comparative study of genetic programming-based algorithms for predicting the compressive strength of concrete at elevated temperature. *Case Stud. Constr. Mater.* <https://doi.org/10.1016/j.cscm.2023.e02199> (2023).
127. Shang, M. *et al.* Predicting the mechanical properties of RCA-based concrete using supervised machine learning algorithms. *Materials* **15**, 2. <https://doi.org/10.3390/ma15020647> (2022).
128. Zeng, H. *et al.* Performance evolution of low heat cement under thermal cycling fatigue: A comparative study with moderate heat cement and ordinary Portland cement. *Constr. Build Mater.* <https://doi.org/10.1016/J.CONBUILDMAT.2024.134863> (2024).

129. Ahmad, A. *et al.* Comparative study of supervised machine learning algorithms for predicting the compressive strength of concrete at high temperature. *Materials* **14**, 15. <https://doi.org/10.3390/ma14154222> (2021).

### Author contributions

W.B.I: formal analysis, methodology, data curation, writing original draft; M.S.S: supervision, writing-reviewing and editing; Mujahid Ali: funding acquisition, project administration; Taoufik Najeh: funding acquisition, project administration, reviewing and editing. All authors have read and approved the final manuscript.

### Funding

Open access funding provided by Lulea University of Technology.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-65905-5>.

**Correspondence** and requests for materials should be addressed to T.N.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024