

Research Article

Qiong Tian, Yijun Lu, Ji Zhou*, Shutong Song, Liming Yang, Tao Cheng*, and Jiandong Huang*

Compressive strength of waste-derived cementitious composites using machine learning

<https://doi.org/10.1515/rams-2024-0008>

received December 19, 2023; accepted March 11, 2024

Keywords: marble cement, gene expression programming, multi-expression programming

Abstract: Marble cement (MC) is a new binding material for concrete, and the strength assessment of the resulting materials is the subject of this investigation. MC was tested in combination with rice husk ash (RHA) and fly ash (FA) to uncover its full potential. Machine learning (ML) algorithms can help with the formulation of better MC-based concrete. ML models that could predict the compressive strength (CS) of MC-based concrete that contained FA and RHA were built. Gene expression programming (GEP) and multi-expression programming (MEP) were used to build these models. Additionally, models were evaluated by calculating R^2 values, carrying out statistical tests, creating Taylor's diagram, and comparing theoretical and experimental readings. When comparing the MEP and GEP models, MEP yielded a slightly better-fitted model and better prediction performance ($R^2 = 0.96$, mean absolute error = 0.646, root mean square error = 0.900, and Nash–Sutcliffe efficiency = 0.960). According to the sensitivity analysis, the prediction of CS was most affected by curing age and MC content, then by FA and RHA contents. Incorporating waste materials such as marble powder, RHA, and FA into building materials can help reduce environmental impacts and encourage sustainable development.

1 Introduction

The production of ordinary Portland cement (OPC) accounts for around 5–8% of the world's CO₂ emissions, and current predictions indicate that this percentage will increase by 8% by the year 2050 [1,2], which raises doubts about the feasibility of reaching the zero-emission target set out by the Paris Agreement. It is critical to identify less damaging alternatives to OPC to decrease CO₂ emissions from the OPC industry. Conversely, marble has become an increasingly popular material around the world. About half of the marble that is mined is wasted due to blasting processes employed in quarrying [3,4]. Because no proper method of disposal is in place, waste is still dispersed throughout the quarries. Tiles and other valuable stones are created from marble, which is transported in massive blocks from quarries to processing facilities. It is wasteful to trim and refine raw blocks. Around 20% of these blocks are reduced to fine powder, while the exact percentage varies by processing method [5]. Typically, open spaces surrounding factories are used for depositing the slurry. As the mixture dries into a fine powder, it poses health risks, including skin irritation, lung cancer, and eye pain. Adding sludge to water also makes it more polluted [6,7]. One negative aspect of plant diversity is the drying out of older trees and hedges caused by small marble atoms on shrub and vegetative leaves [8–10].

OPC contains 60–65% CaO, 20–25% SiO₂, and 4–8% Al₂O₃, as stated by Neville and Brooks [11]. The majority of the calcium oxide used to make cement comes from limestone, as the researcher has already mentioned repeatedly. Limestone and waste marble powder (WMP) have several molecular components, including a high CaO content [12,13]. Therefore, Khan *et al.* [14] created a binding material such as marble cement (MC) using WMP and a silica-rich substance such as clay, as depicted in Figure 1. The phases that make up MC were determined to be 3.7%

* **Corresponding author: Ji Zhou**, College of Civil and Environmental Engineering, Hunan University of Science and Engineering, Yongzhou, China, e-mail: hnkjxyzj@huse.edu.cn

* **Corresponding author: Tao Cheng**, School of Civil Engineering, Hubei Polytechnic University, Huangshi, 435003, China, e-mail: drchtao@126.com

* **Corresponding author: Jiandong Huang**, School of Civil Engineering, Guangzhou University, Guangzhou, China, e-mail: jiandong.huang@hotmail.com

Qiong Tian: College of Civil and Environmental Engineering, Hunan University of Science and Engineering, Yongzhou, China

Yijun Lu: School of Civil Engineering, Guangzhou University, Guangzhou, China

Shutong Song, Liming Yang: China Construction Railway Investment Rail Traffic Construction Co., Ltd., Guangzhou, China

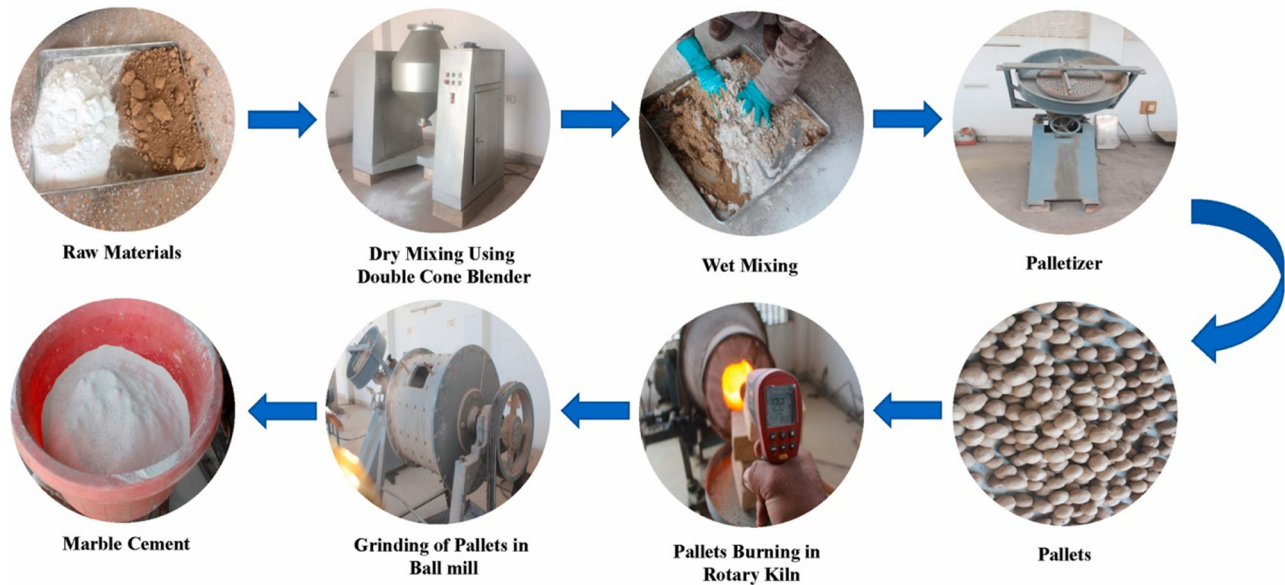


Figure 1: Production of MC [19].

C_3S , 23.11% free lime, and 52.51% C_2S , as reported by Khan *et al.* in an X-ray diffraction study [14]. Sluggish cooling led to low C_3S concentrations and high C_2S and CaO concentrations. The cold air was softly blown over the MC pallets. Then, at about $1,100^\circ C$, C_3S changed back into C_2S and CaO . By rapidly cooling the cement clinkers during OPC manufacture, the C_3S that is generated at high temperatures is preserved [15]. The shorter setting time compared to OPC was attributed to the increased free lime content in MC, which allowed the cement to avoid slaking (the formation of $Ca(OH)_2$) and use less water. The OPC increases as the concentration of free lime rises. The enormous pressure that develops within the cement matrix causes mortar and concrete made with this cement to deteriorate with time. Due to the pulverization of a considerable amount of $Ca(OH)_2$, the compressive strength (CS) of cement falls as the allowed lime level increases [16–18].

As a partial substitute for MC, researchers suggested using pozzolanic materials to boost the consumption of free lime content. Various pozzolanic applications are now being explored for a variety of waste byproducts [20,21]. Powdered coal is the fuel used in thermal coal power units. Bottom ash and fly ash (FA) are the byproducts of burning it. Emissions from vehicles carry FA in its microscopic particle form. Precipitators are put in place before the chimney to extract it from the exhaust gases. Precipitators remove FA from exhaust gases before entering the chimney [22,23]. Worldwide, rice is also cultivated. “Rice husk” describes the hard outside layer that protects individual grains of rice [24]. After being burned at a controlled

temperature, this waste material becomes ash. The pozzolanic properties of rice husk ash (RHA), a byproduct of finely powdered ash, are unparalleled [25].

The importance of CS has led to substantial study of cement-based materials (CBMs) [26–29]. Important information on the CBM’s characteristics may be found in its CS [30,31]. Many mechanical and durability characteristics are related to the concrete’s CS in some way [32–34]. By creating analytical models for material strength, analysts are aiming to reduce unnecessary testing and associated expenses. To replicate the material properties, which are obtained by regression analysis, a plethora of conventional models, including best-fit curves, are used. Traditional regression techniques may mistakenly assume the material’s intrinsic behavior when dealing with CBMs because of their non-linear nature [35–37]. The use of artificial intelligence (AI) methods, particularly supervised machine learning (ML), is propelling the development of more sophisticated models in this area [38–40]. In order to generate reliable predictions, these models rely on input features and undergo experimental verification of their accuracy. The application of ML algorithms to predict CBM properties is on the rise [41–44].

This work aimed to forecast the CS of FA and RHA-based MC concrete (FR-MCC) using ML algorithms, specifically multi-expression programming (MEP) and gene expression programming (GEP). This forecast is based on information found in previously published works. Several measures were used to assess the efficacy of ML procedures, such as the R^2 coefficient, statistical tests, and the dispersion of

expected results. Investigating the efficacy of ML techniques in reliably projecting material attributes was the driving force behind this study. An exploratory experiment or an analysis of existing databases can yield the dataset that is required by ML methods [45,46]. By examining this information, ML algorithms may gain a more accurate understanding of the material's qualities. By combining experimental data with four input parameters, ML methods' ability to predict the CS of FR-MCC was assessed [47,48]. Through the application of sensitivity analysis, additional exploration into the relevance of raw materials was conducted [49,50]. The newly gathered features and developed ML models could be used to improve the current sustainable materials database or to guide the formulation of CBM blends, among other potential applications.

2 Methods of research

2.1 Collecting and analyzing data

Using MEP and GEP techniques, this research analyzed a dataset comprising 500 data points from an experimental inquiry with the purpose of forecasting the CS of FR-MCC [51]. With an initial set of 84 data points amplified to 500 points, this study predicted the CS of FR-MCC using four input parameters: MC, FA, RHA, and curing age (CA). The Python code that was executed followed a certain protocol for expanding the dataset's data points. First, it opens a file dialog box built on Tkinter so the user can choose a database file. The selection of the file is then followed by its import into a Pandas DataFrame, and the script checks the current point count. A new file is generated that contains the dataset that is the end result of merging the synthetic

data with the original DataFrame. As the data are being supplemented, the script provides illuminating statements. Among the information included in these declarations are the total number of data points added, the amount of synthetic data points, and the precise location of the saved file. Furthermore, the script takes into consideration situations when either no file is chosen or resampling is necessary. With the aid of data preparation, the data were collected and organized. One popular approach to data preparation is to use it as a cushion to get past a major problem with the well-known process of getting new insights out of old data. Eliminating extraneous information and background noise from data is what data preparation is all about. The model analysis made use of regression and error-distribution approaches. Table 1 displays the outcomes of multiple descriptive statistics that were calculated using this dataset. In addition, validation was used to assess the efficacy of the models that were used. Histograms show the distribution of frequencies for different values in Figure 2(a)–(e). As a whole, the dataset's frequency distribution can be described by combining its constituent components' distributions. One technique to gain insight into the dataset is to construct a relative frequency distribution. This will show you how common certain values are and how often they appear.

Overfitting occurs frequently in the field of mathematical dataset simulation [30,52]. This happens when a model does a good job of reproducing the input data during training and development but fails miserably when faced with input parameters that are different from both sets. While the model does a great job of mimicking the known dataset, it can produce wildly inaccurate results when asked to predict values for non-standard input parameters [53,54]. The authors used regularization (L1 and L2) and other strategies to punish complexity and discourage

Table 1: Data variable statistical descriptions

Statistical parameters	MC (kg·m ⁻³)	FA (kg·m ⁻³)	RHA (kg·m ⁻³)	Age (days)	CS (MPa)
Mean	241.48	45.62	40.90	170.32	14.38
Standard error	1.83	2.35	2.30	5.68	0.20
Median	230.00	0.00	0.00	136.50	14.70
Mode	230.00	0.00	0.00	91.00	20.05
Standard deviation	40.83	52.63	51.35	126.93	4.50
Sample variance	1667.24	2769.81	2636.97	16110.58	20.24
Kurtosis	-0.05	-1.44	-1.25	-1.18	-1.06
Skewness	0.84	0.49	0.65	0.56	-0.31
Range	131.00	131.00	131.00	336.00	16.19
Minimum	197.00	0.00	0.00	28.00	5.81
Maximum	328.00	131.00	131.00	364.00	22.00
Sum	120738.00	22811.00	20451.00	85162.00	7190.66
Count	500.00	500.00	500.00	500.00	500.00

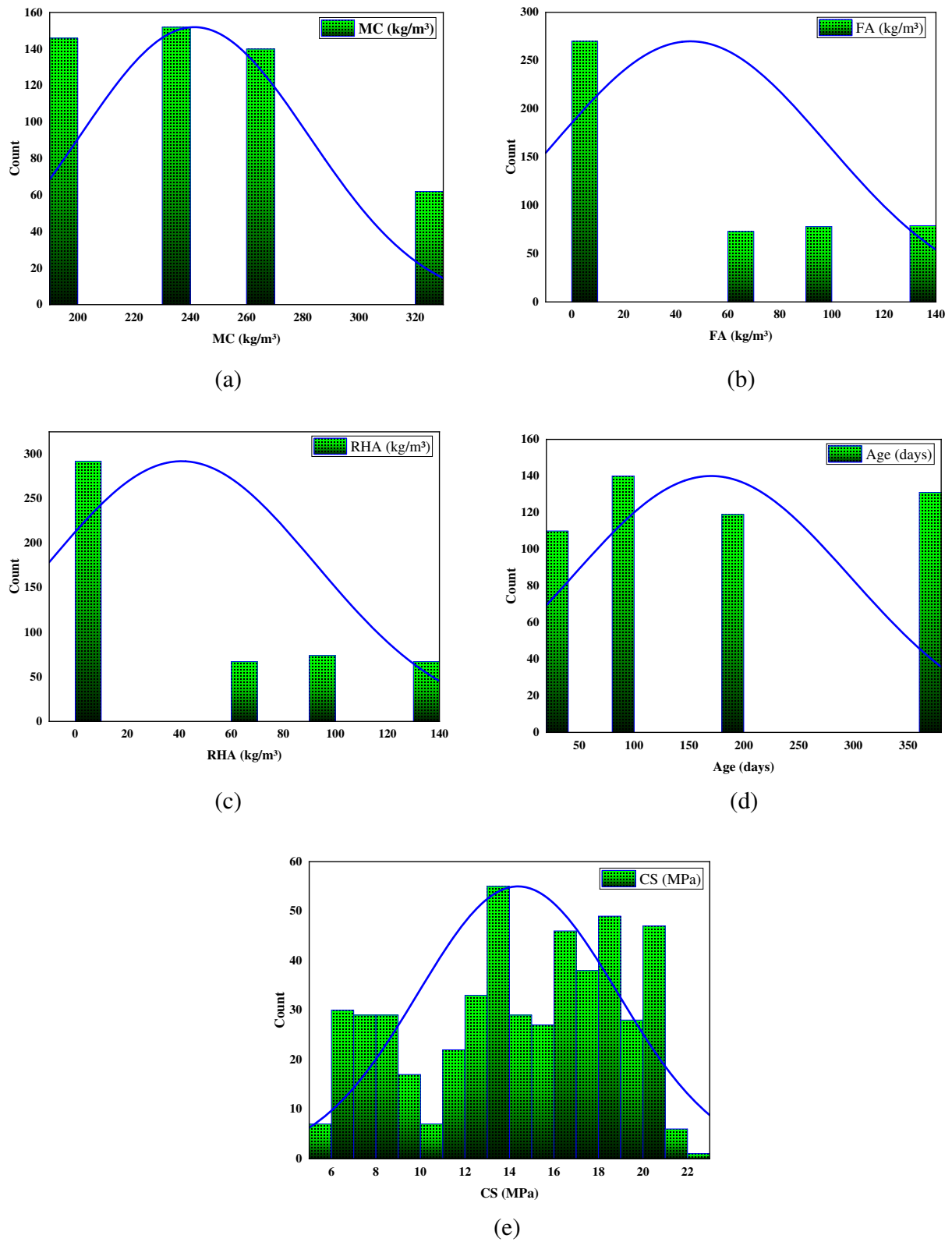


Figure 2: Frequency distribution of the input and output features of the database: (a) MC, (b) FA, (c) RHA, (d) Age, and (e) CS.

overemphasis on specific features in order to minimize overfitting in the constructed model. To evaluate the model on new data, the dataset was split into two parts: training and validation. In order to protect the training data from fitting noise, early stopping was instituted by keeping an eye on validation performance and ending training if there was no progress. All of these steps added up to a generalizable model that can handle new, unknown data with ease and without overfitting [55,56].

2.2 ML algorithm application

A highly controlled setting was used to test the CS of FR-MCC. In order to obtain the outcome (CS), four inputs were required. Using advanced ML methods such as GEP and MEP, predictions were made for the CS of FR-MCC. It is usual practice to evaluate outcomes by feeding them into ML algorithms. A total of 70% of the data were used to train the ML models, while 30% were set aside for testing. As a measure of the model's effectiveness, the R^2 score of the predicted outcomes quantified the degree to which the actual values differed from the expected ones; a smaller R^2 number indicates a larger discrepancy [36,57,58]. Statistical testing and error evaluations are two of the methods used to verify the

model's accuracy [59,60]. While the hyperparameters for both the GEP and MEP models are presented in Table 2, Figure 3 gives a simple representation of a scenario model.

2.2.1 GEP model

Taking cues from Darwin's idea of evolution, Holland created the genetic algorithm (GA) [62]. The genomic process, which is defined by the order of GAs, is followed by the observation of chromosomes of persistent length. One unique GA approach is that Koza coined the term "gene programming" [63,64]. Generalized problem-solving (GP) uses GAs to generate an evolutionary model [65]. The ability to substitute non-linear structures, such as parse trees, for binary strings of fixed length gives GP its flexibility. By leveraging naturally occurring genetic components, including reproduction, crossover, and alteration, AI software tackles reproductive difficulties, drawing inspiration from Darwin's theory [66]. By progressively removing inefficient programs from succeeding iterations, GP aims to achieve its purpose. Similar to the previous example, cleaning the area by removing trees that do not match properly is an important part of implementing the chosen plan. Nevertheless, early convergence is protected by the evolutionary process [66,67]. It is necessary to specify five critical

Table 2: Predefined model factors for MEP and GEP (parameters similar to [61])

MEP		GEP	
Parameters	Settings	Parameters	Settings
Number of generations	250	Stumbling mutation	0.00141
Problem type	Regression	Constant per gene	10
Terminal set	Problem input	Inversion rate	0.00546
Replication number	15	Head size	8
Operators/variables	0.5	Data type	Floating number
Number of treads	2	Two-point recombination rate	0.00277
Number of generations	500	Chromosomes	200
Mutation probability	0.01	Linking function	Addition
Error	MSE, MAE	Lower bound	-10
Cross-over probability	0.9	Upper bound	10
Number of subpopulations	50	Mutation rate	0.00138
Subpopulation size	100	Genes	4
Number of runs	15	Leaf mutation	0.00546
Function set	+, −, ×, ÷, square root	General	CS
Code length	40	IS transposition rate	0.00546
		RIS transposition rate	0.00546
		One-point recombination rate	0.00277
		Function set	+, −, ×, ÷, square root
		Gene recombination rate	0.00277
		Gene transposition rate	0.00277
		Random chromosomes	0.0026

factors before using the GP approach. Among these, you may find the following: a list of mandatory domain tasks, an assessment of fitness, a list of primary functional operators (including population size and crossover), and the establishment of outcomes according to the method-specific criteria [66]. Although GP incorporates recurrent model building, a crossover genetic processor is mostly responsible for parse tree formation [48]. The need for non-linear GP forms to serve as both genotype and phenotype has resulted in the development of intricate expressions that stand in for desirable features [67].

GEP is a variation on GP that Ferreira initially proposed [67]. The GEP model incorporates static-length-lined chromosomes into parse trees in accordance with the population-generation theory. GP employs simple, fixed-length chromosomes to encrypt medium-sized software; GEP is an improved version of that. The ability to predict complicated and non-linear issues with high reliability is

one of GEP's many benefits [68,69]. The fitness function, the last set, and the conditions for termination are defined in the same way as in GP. Chromosomes are generated at random by the GEP technique; however, they are designated as such before they are produced using the "Karva" dialectal. GEP relies on a fixed-length line as its fundamental basis. On the other hand, GP's code processing of data displays parse trees of varying lengths. Initially, these unique cords are described as genomes of static length. Subsequently, they stand in for chromosomes by means of non-linear manifestation/parse trees that have pronged morphologies varying in size [66]. These genotypes and the small number of phenol strains each have their own unique genetic code [67]. The necessity for expensive structural changes or duplications is eradicated by GEP's capacity to maintain the genome from one generation to another.

In a typical chromosome, the "head" and the "tail" are the two complementary regions. Surprisingly, this phenomenon can occur when multiple genes are produced by a single chromosome [66]. Logic, mathematics, arithmetic, and Boolean operations are encoded in these genes. A programmer assigns specific functions to cells in the genetic code. Karva, a novel language, can decipher the contents of these chromosomes and use that information to build empirical formulas. A basic revolution marks the beginning of the trip on the Expression Tree (ET), which starts at Karva. According to Eq. (1), ET arranges the nodes in the underlying layer [68]. In terms of both amount and duration, the total number of ETs can affect GEP gene K-expression:

$$ET_{GEP} = \log \left(i - \frac{3}{j} \right). \quad (1)$$

The fact that GEP's findings are not dependent on any previous relationships makes it a sophisticated ML method. A GEP mathematical equation goes through a number of steps, as seen in Figure 4. There is no change to a person's chromosomal count at delivery. Upon confirmation that these chromosomes are ETs, a thorough assessment of the health of every individual can be conducted. Reproductive privileges are bestowed upon the healthiest and strongest individuals. The best solution is found by taking the most skilled people through an iterative process. Reproduction, alteration, and overlap are the three generations of genetic processes that have resulted in the ultimate numerical expression once they have been applied.

2.2.2 MEP model

Due to its reliance on linear chromosomes, the MEP represents a state-of-the-art, model linear-based GP technique.

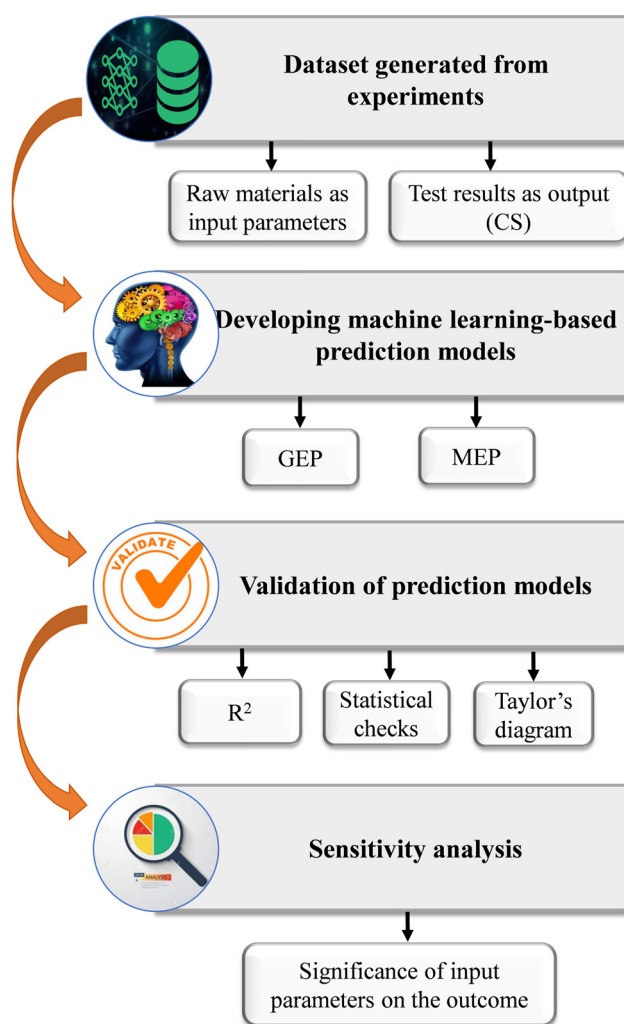


Figure 3: Comprehensive overview of the studied method.

In terms of their core software, the GEP and MEP are very similar to one another. MEP's capacity to encode numerous software components (substitutes) into a sole chromosome sets it apart from its predecessor, the GP method. Using fitness analysis to choose the optimal chromosome yields the desired result [70,71]. According to Oltean and Grosan, this happens when a bipolar system recombines to form two new offspring, with each offspring choosing one parent [72]. Figure 5 demonstrates that the process will keep running until the termination condition is met or until the best program is found. Mutations in newborns happen here. A number of components can be combined using the MEP model, much as with the GEP model. Criteria that are important in MEP include the number of functions, the number of subpopulations, the length of the algorithm or code, and the possibility of crossover [73]. Assessing and accounting for a population when its size equals the total number of packages is a time-consuming and complex process. Mathematical expression sizes are heavily impacted by the code length. The amount of MEP parameters needed to build a trustworthy model of rheological properties is shown in Table 2.

Both approaches rely heavily on literature datasets during the modeling and evaluation stages [74,75]. Famous linear GP approaches such as the GEP and the MEP are superior at predicting the properties of ecological concrete,

according to certain studies. According to Grosan and Abraham, the optimal neural network-based strategy was a hybrid of lined genomic programming and maximum entropy programming (MEP) [76]. The GEP's method of operation is marginally more intricate than that of the MEP [73]. Notwithstanding MEP's lower density compared to GEP [77], there are a few key differences: (i) MEP explicitly encodes function argument references, (ii) non-coding units are not required to be presented at a set point in the genes, and (iii) MEP allows code reprocess. It is commonly believed that the GEP chromosome has greater competence due to the signs found at its "head" and "tail" that facilitate the writing of syntactically accurate software programs [72]. Therefore, additional comprehensive assessments of these genomic strategies for engineering problems are urgently required.

2.3 Validation of models

A test set was used to statistically test the models that were created using GEP and MEP. Using equations from previous

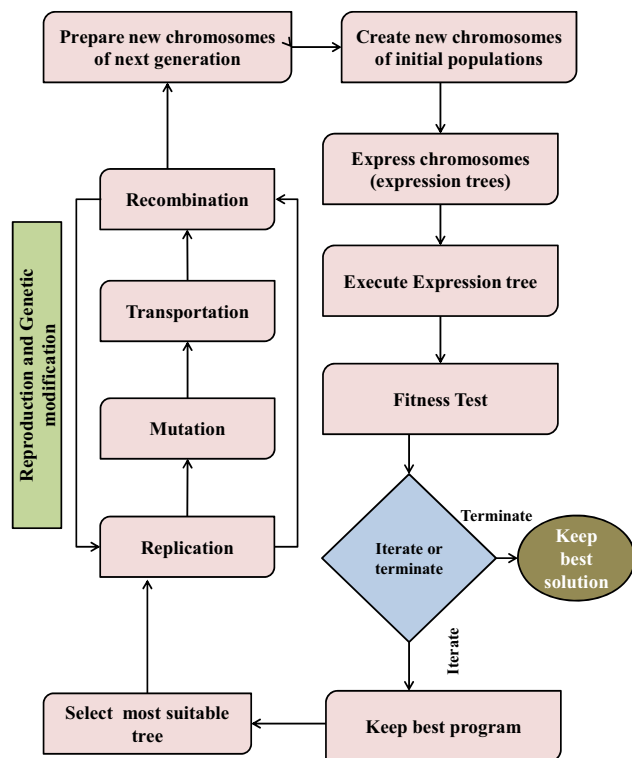


Figure 4: GEP procedure's method flowchart [61].

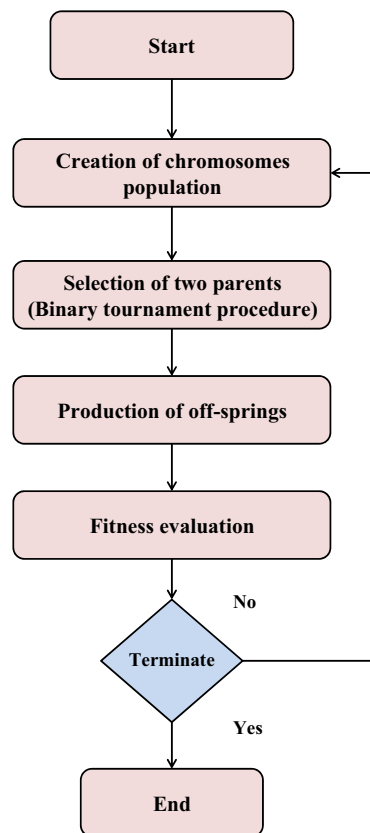


Figure 5: Process flow diagram for MEP operation [61].

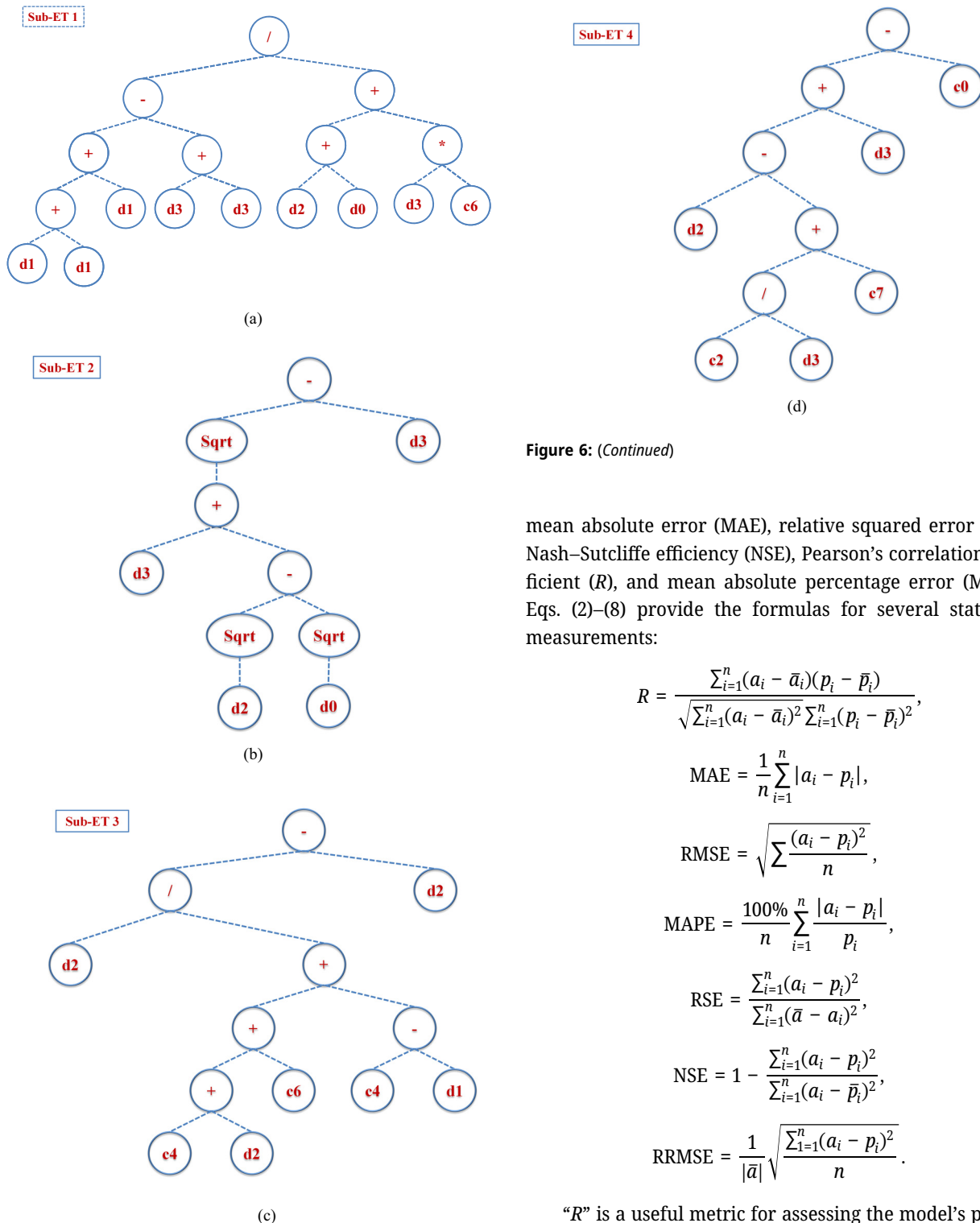


Figure 6: (Continued)

mean absolute error (MAE), relative squared error (RSE), Nash–Sutcliffe efficiency (NSE), Pearson’s correlation coefficient (R), and mean absolute percentage error (MAPE). Eqs. (2)–(8) provide the formulas for several statistical measurements:

$$R = \frac{\sum_{i=1}^n (a_i - \bar{a}_i)(p_i - \bar{p}_i)}{\sqrt{\sum_{i=1}^n (a_i - \bar{a}_i)^2 \sum_{i=1}^n (p_i - \bar{p}_i)^2}}, \quad (2)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |a_i - p_i|, \quad (3)$$

$$\text{RMSE} = \sqrt{\sum_{i=1}^n \frac{(a_i - p_i)^2}{n}}, \quad (4)$$

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^n \frac{|a_i - p_i|}{p_i}, \quad (5)$$

$$\text{RSE} = \frac{\sum_{i=1}^n (a_i - p_i)^2}{\sum_{i=1}^n (\bar{a} - a_i)^2}, \quad (6)$$

$$\text{NSE} = 1 - \frac{\sum_{i=1}^n (a_i - p_i)^2}{\sum_{i=1}^n (a_i - \bar{p}_i)^2}, \quad (7)$$

$$\text{RRMSE} = \frac{1}{|\bar{a}|} \sqrt{\frac{\sum_{i=1}^n (a_i - p_i)^2}{n}}. \quad (8)$$

Figure 6: ET schematic for CS-GEP model: (a) sub-ET 1, (b) sub-ET 2, (c) sub-ET 3, (d) sub-ET 4. d0: MC, d1: FA, d2: RHA, and d3: CA.

studies [75,78–81], seven arithmetical measures were computed for each of the models: relative root mean square error (RRMSE), mean root mean square error (RMSE),

“ R ” is a useful metric for assessing the model’s predictive capability, where “ n ” stands for the overall count of data points, “ a_i ” and “ p_i ” denote the i th tangible and projected values, and “ $\bar{a}_i$ ” and “ $\bar{p}_i$ ” denote the mean tangible and projected values, correspondingly. A strong correlation between expected and actual production volumes is indicated by a high value of R [10,82,83]. No matter how you divide or multiply, the component R will stay the same.

The R^2 statistic, which provides a more accurate estimation of the actual value, was computed between the actual results and the expected outcomes. When R^2 values are close to 1, it indicates that the process of developing the model is quite successful [84,85]. Just as MAE and RMSE showed great improvement with progressively larger mistakes, the proposed model shows even greater performance with smaller errors, and both approach zero with larger errors [86,87]. However, upon closer inspection, it became apparent that continuous and smooth databases are where MAE truly excels [88,89]. When the computed error numbers are reduced, it is generally believed that the model is performing effectively:

$$a20\text{-index} = \frac{m20}{M}. \quad (9)$$

This takes into consideration an expected or experimental value between 0.80 and 1.20, where M is the number of specimens in the dataset and $m20$ is the number of entries, as provided in Eq. (9) [90]. The optimal $a20$ -index values, according to the prediction model, would be 1%. The proposed 20-index offers the advantage of a physical engineering approach, revealing the percentage of samples that correspond to expected values within a $\pm 20\%$ uncertainty range of experimental data.

The Taylor diagram, in conjunction with statistical validation, is one of the most useful tools for evaluating the predictive power of a model. In order to determine which models are more credible and accurate, this figure plots their divergence from the truth, which serves as the reference point [91,92]. Three metrics can be used in order to ascertain the optimal location for a model. These metrics are the standard deviation, which is depicted by the axis, the correlation coefficient, which is represented by the radial lines, and the RMSE, which is represented by the circular lines that are centered at the actual value point. The most reliable model is the one that has the best track record of accurately predicting outcomes [91,93].

3 Results and analysis

3.1 CS-GEP models

Figure 6(a)–(d) shows how the GEP technique predicts FR-MCC's CS using ET models based on chromosomal number and head size data. The $\div$, $\times$, $-$, $+$, and square root mathematical operations are used to create most CS sub-ETs in FR-MCC. The output is a mathematical formula once the

sub-ETs have been encrypted. The equations that are produced as a result of these models (Eqs. (10)–(14)) can be used to make predictions regarding the future CS of FR-MCC by making use of the presented input values. Under ideal conditions and with sufficient observations, the resulting model performs better than a perfect model would. Figure 7(a) illustrates the use of lines of statistical significance (CS) for the purpose of comparing the predicted values of the model with the values that were observed in the testing set. According to the great degree of agreement between the expected and observed values ($R^2 = 0.95$), the GEP approach is effective in estimating the CS of FR-MCC. Figure 7(b) shows the absolute error plotted against experimental data to show that there can be a difference between the GEP model and actual outcomes.

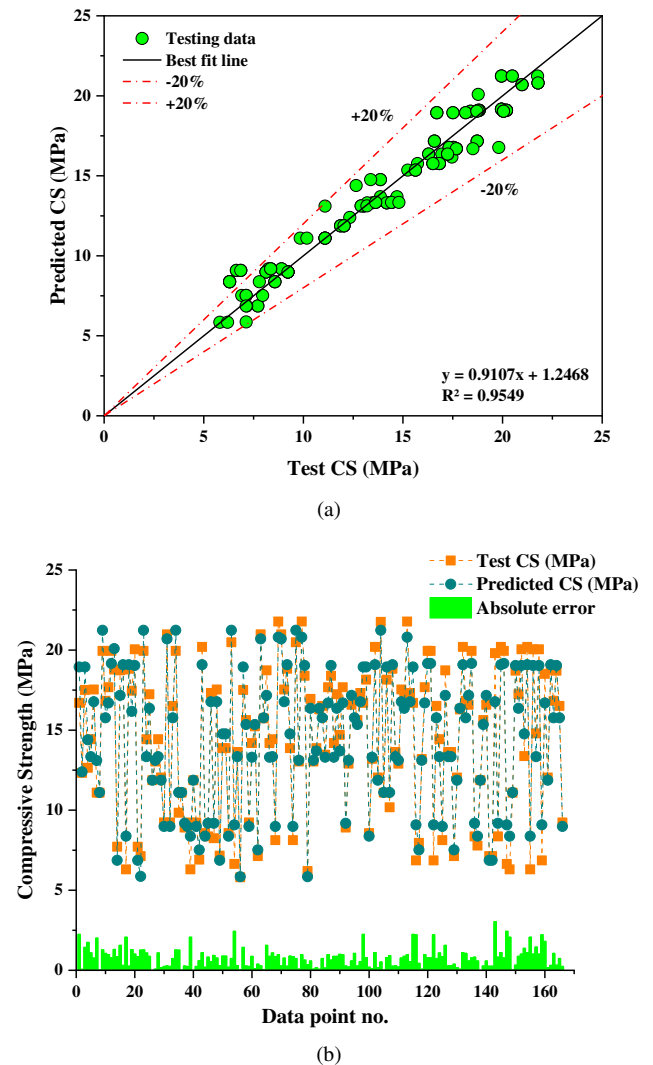


Figure 7: GEP model for CS in FR-MCC: (a) the relationship between tested and predicted CS and (b) the error distribution in tested and predicted CS.

The fact that the absolute errors range from 0.001 to 3.040 MPa shows that the experimental data agree well with the predictions made by the GEP equation. An average of 0.792 MPa is the inaccuracy. Figure 8 shows that the error values follow a bell-shaped distribution, with 49 measurements over 1.0 MPa, 56 readings between 0.5 and 1.0 MPa, and 61 readings below 0.5 MPa. The occurrence of maximal error frequencies is extremely rare:

$$CS(\text{MPa}) = A + B + C + D, \quad (10)$$

$$A = \frac{(3FA - 2CA)}{(RHA + MC) - \frac{CA}{4.723}}, \quad (11)$$

$$B = (\sqrt{CA + (\sqrt{RHA} - \sqrt{MC})}) - CA, \quad (12)$$

$$C = \left(\frac{RHA}{((-11.714 + RHA) - 9.883) + (-11.714 - FA)} \right) - RHA, \quad (13)$$

$$D = \left(\left(RHA - \left(\frac{11.364}{CA} + 1.299 \right) \right) + CA \right) + 4.580, \quad (14)$$

where MC is the marble cement, FA is the fly ash, RHA is the rice husk ash, CA is the curing age, and CS is the compressive strength.

3.2 CS-MEP model

After analyzing the MEP data and taking into account the impact of the four independent components, an empirical method was created to estimate the CS of FR-MCC. The final mathematical model that comes out of this procedure is shown in Eq. (15):

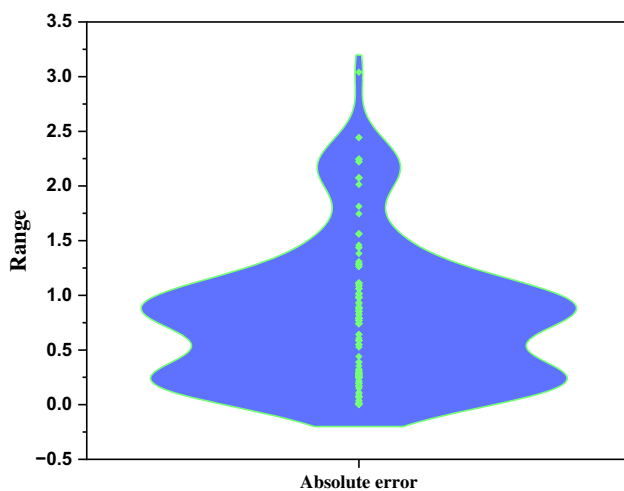


Figure 8: Violin plot for GEP models' error distribution.

$$CS(\text{MPa}) = \sqrt{CA + RHA} + \frac{CA - MC}{CA} + \frac{2\left(FA - \frac{MC}{FA}\right) + MC + CA + RHA}{MC}, \quad (15)$$

where MC is the marble cement, FA is the fly ash, RHA is the rice husk ash, CA is the curing age, and CS is the compressive strength.

Figure 9(a) shows that the MEP model is well trained, capable of handling oversimplification, and performs adequately on novel, untested data, with an R^2 of 0.96. It appears that the CS-MEP model is marginally more accurate than the CS-GEP model, according to its higher R^2 value. Figure 9(b) displays the results of an analysis of

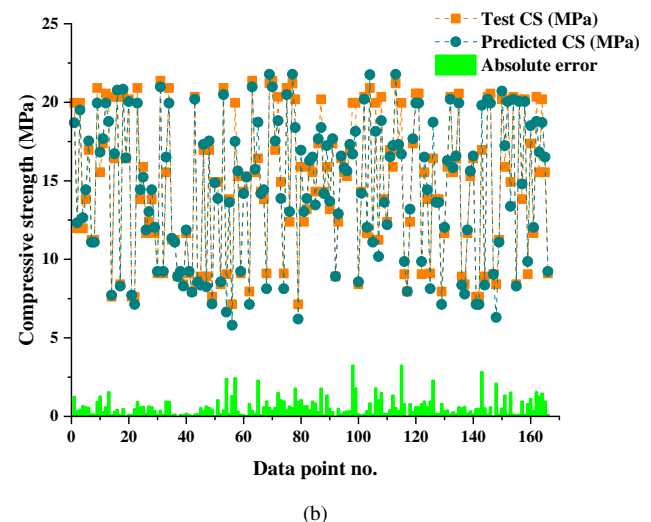
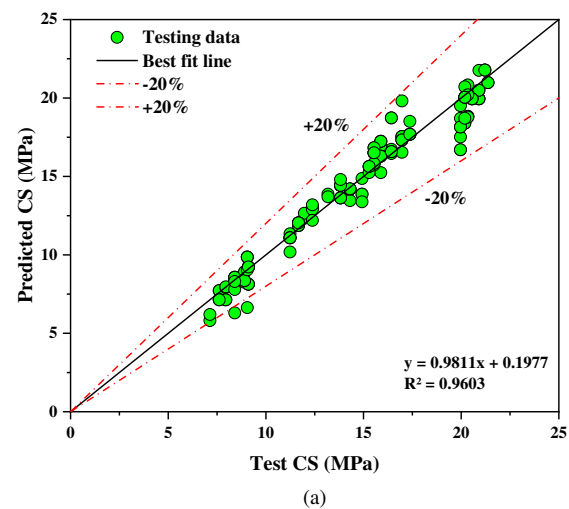


Figure 9: MEP model for CS in FR-MCC: (a) the relationship between tested and predicted CS and (b) the error distribution in tested and predicted CS.

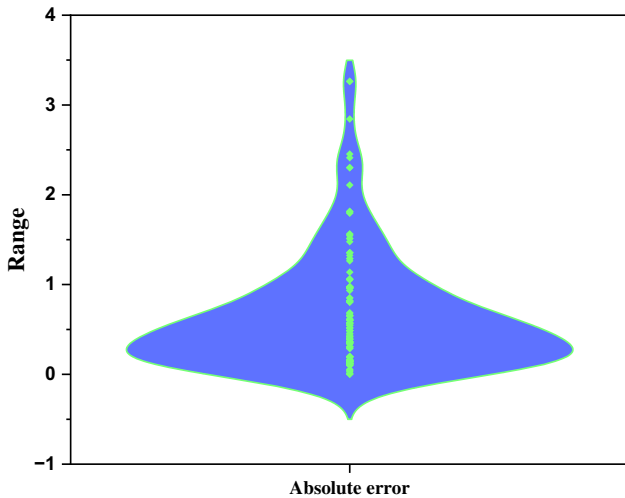


Figure 10: Violin plot for MEP models' error distribution.

absolute differences between the goal and observed values performed in MEP simulations. The presented data show that the error margin for MEP estimations ranged from 0.002 to 3.264 MPa, with an average of 0.647 MPa. With 86 values below 0.5 MPa, 50 values between 0.5 and 1.0 MPa, and 30 values over 1.0 MPa, the mean error values were also below 3.500 MPa. Remember that the MEP model predicts less degree of outcome variability than the GEP model when considering the most extreme values. Both the MEP and GEP models might be incredibly accurate predictors. Implementing the MEP equation leads to a decrease in both the correlation coefficient and the error standard deviations. There has been much use of the MEP equation because of its generalizability and its condensed form. The MEP model appears to be better than the GEP model because it has a higher correlation coefficient and lower error levels (as shown in Figure 10), comparable to earlier research of a similar kind [94,95]. Some researchers have also developed prediction models using different ML techniques for various properties of concrete [96–98].

Table 3: Findings gained by statistical examination

Property	CS (MPa)	
	GEP	MEP
MAE	0.792	0.646
MAPE	6.50	4.60
RMSE	0.992	0.900
R	0.977	0.980
RSE	0.297	0.229
NSE	0.953	0.960
RRMSE	0.642	0.342
a20-index	0.840	0.940

3.3 Validation of model

Based on the previously described Eqs. (2)–(8), Table 3 displays the results of effectiveness and error computations (RMSE, MAE, R , RSE, NSE, and RRMSE). The created models' prediction accuracy is higher when their error values are smaller. The CS-GEP model's MAE values were 0.792 MPa, while the counterpart CS-MEP model's values were 0.646 MPa, a little reduction. CS-GEP model's MAPE value of 6.60% was significantly reduced to 4.60% in the corresponding CS-MEP model. Additionally, additional error-based statistical metrics, such as RMSE, RSE, and RRMSE, showed an analogous trend. Additionally, error-based validation was used in order to evaluate the effectiveness of the constructed models. The Pearson's coefficient (R) and the NSE were the two measures that were used. An increase in a model's efficiency indicates an improvement in its forecast accuracy. The CS-GEP model had an NSE of 0.953, whereas the matching CS-MEP model had a slightly higher NSE of 0.960. When measured with Pearson's coefficient (R), the models that were constructed produced findings that were comparable to one another. As can be seen in Figure 11, a Taylor diagram is used to compare all of the different forecasting models. As far as predicting the CS of FR-MCC is concerned, the MEP models are relatively near, whereas the GEP models are relatively far away. The MEP model outperformed other ML-based techniques in predicting the CS of FR-MCC, according to prior study, since it had the lowest standard deviation, highest efficiency, lowest error, and best R^2 .

3.4 Sensitivity analysis

The research aims to investigate the effect of different input parameters on CS prediction for FR-MCC. The input

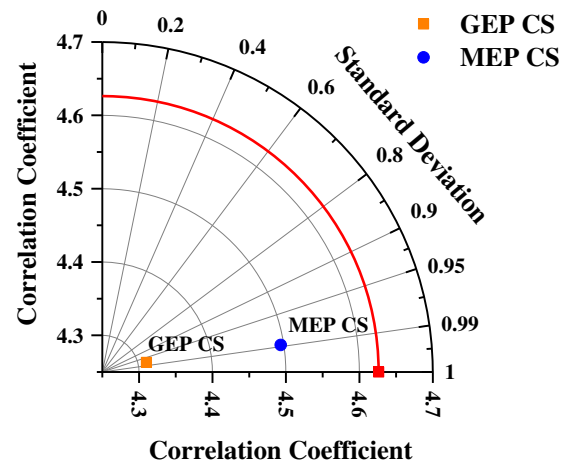


Figure 11: Taylor diagram for CS models.

factors are strongly correlated with the expected output [99]. Figure 12 shows how each variable affects the CS of FR-MCC, giving us a glimpse into the future of concrete and building materials in general. The highest impact, 65%, on predicting the CS of FR-MCC came from CA, followed by MC (23.0%), RHA (9.0%), and FA (3%). The amount of model parameters and data points used in sensitivity studies was directly related to the results. The results of the analysis were shown to be affected differentially by various input parameters, such as the quantities of concrete mix when the ML technique was applied. The relative importance of the model's input parameters was determined using Eqs. (16) and (17):

$$N_i = f_{\max}(x_i) - f_{\min}(x_i), \quad (16)$$

$$S_i = \frac{N_i}{\sum_{j=1}^n N_j}, \quad (17)$$

where $f_{\max}(x_i)$ represents the highest predicted value across all i th outputs and $f_{\min}(x_i)$ represents the lowest.

4 Discussions

The GEP and MEP models that were used in this investigation ensure that the predictions will be specific to FR-MCC. This is due to the fact that these models are only able to accept values from a limited range of four input parameters. Since all of the models employ the same unit measurements and testing technique, the CS predictions they

produce are reliable. These models use mathematical equations to help us better understand mix design and how each input parameter affects it. The projected models might not work well if the composite analysis uses more than the four inputs. Their intended purpose must be well aligned with the inputs used to train these models; otherwise, they might not yield the expected results. If the input parameters' units are changed or inconsistent, the models could under- or over-predict the outcomes. The models can only work if the unit sizes remain constant. There are numerous applications of ML models in the construction industry, including strength prediction, quality assurance, risk assessment, predictive maintenance, and energy efficiency improvements. Still, there are a few problems with these models. One is that they use human input, which can lead to inaccurate results and inaccurate data. To address these limitations and improve ML-based solutions, future research could look into integrating internet of things devices, creating hybrid models, using explainable AI techniques, considering sustainability, and customizing data generation and distribution for specific industries, among other things. Technological improvements have the potential to revolutionize the construction industry. By enhancing efficiency, interpretability, transparency, and informed decision-making, these advancements could reduce project delays, increase safety, and promote more sustainable practices. The findings of this study may lead to a shift toward greener construction methods and an increase in the usage of long-term, eco-friendly materials.

5 Study limitations and suggestions for future research

Taking four variables into account, this study used a dataset of 500 records to forecast CS. One potential strategy to improve the models' performance in future research is to add more records from experiments to the dataset. With a larger dataset, the model can make more accurate predictions, which boosts confidence in the results. While GEP and MEP models were used in this study, hybrid ML approaches such as genetic algorithm-particle swarm optimization and random forest-artificial neural network should be investigated in future analyses. In addition, there is hope for enhancing model performance with individual and ensemble techniques, including support vector machine, decision tree, bagging, and boosting. It makes sense to combine these hybrid approaches, and doing so might greatly improve prediction abilities. Even though they were not used in this work, post hoc explanatory techniques such as SHapely Additive

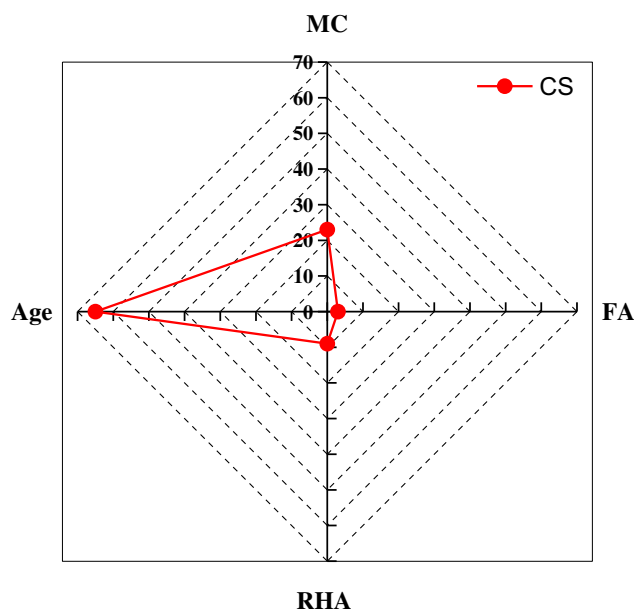


Figure 12: Sensitivity analysis radar plot.

explanations, local interpretable model-agnostic explanations, and partial dependence plot can provide valuable insights into the ML model's predictions. Additional information about the interpretability of models can be gleaned from future studies that use similar methods. Much of the current research on using ML to forecast concrete qualities has focused on mechanical considerations. On the other hand, research into important aspects, including concrete microstructure, dynamic properties such as fatigue, and durability, is noticeably inadequate. For a more complex understanding of concrete performance, additional study is needed to use ML approaches to thoroughly investigate these elements affecting durability.

6 Conclusions

The objective of this study was to examine the CS of FR-MCC using MEP and GEP. The models underwent training, testing, and validation using 500 sets of CS data collected from laboratory tests. The following are the key findings that were discovered by the study:

- To estimate CS for FR-MCC, the GEP method was sufficiently accurate ($R^2 = 0.95$), but the MEP method was more exact ($R^2 = 0.96$).
- For the GEP method, the average discrepancy between the predicted and actual CS (errors) was 0.792 MPa, while for the MEP method, it was 0.647 MPa. With these error rates, it was clear that the MEP technique was better at predicting the CS of FR-MCC than the GEP model.
- The models' efficacy has been validated statistically. In contrast to the MEP model's 4.60% MAPE, the GEP model's MAPE was 6.50%. The MEP model was more predictable than the GEP model, which had an MAE of 0.792 MPa, with a value of 0.646 MPa.
- Sensitivity analysis showed that the prediction of CS of FR-MCC was most affected by CA at 65%, followed by MC at 23%, RHA at 9%, and FA at 3%.

What makes GEP and MEP so crucial for feature prediction in other databases is the unique mathematical method they provide. Quickly evaluating, improving, and rationalizing the proportioning of concrete mixtures is possible with the mathematical models that scientists and engineers can apply to this work.

Acknowledgments: The authors gratefully acknowledge the support of the Natural Science Foundation of Hunan and Hunan Provincial Transportation Technology Project.

Funding information: This research was supported by the Natural Science Foundation of Hunan (Grant No. 2023JJ50418) and Hunan Provincial Transportation Technology Project (Grant No. 202109). The authors are grateful for this support.

Author contributions: Q.T.: conceptualization, methodology, formal analysis, writing-original draft. Y.L.: data acquisition, software, methodology, writing, reviewing, and editing. J.Z.: investigation, funding acquisition, supervision, writing, reviewing, and editing. S.S.: formal analysis, resources, methodology, writing, reviewing, and editing. L.Y.: formal analysis, visualization, writing, reviewing, and editing. T.C.: validation, investigation, supervision, writing, reviewing, and editing. J.H.: conceptualization, supervision, project administration, writing, reviewing, and editing. All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Conflict of interest: The authors state no conflict of interest.

Data availability statement: The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

- [1] Castro-Alonso, M. J., L. E. Montañez-Hernandez, M. A. Sanchez-Muñoz, M. R. Macias Franco, R. Narayanasamy, and N. Balagurusamy. Microbially induced calcium carbonate precipitation (MICP) and its potential in bioconcrete: microbiological and molecular concepts. *Frontiers in Materials*, Vol. 6, 2019, id. 126.
- [2] Habert, G., S. A. Miller, V. M. John, J. L. Provis, A. Favier, A. Horvath, et al. Environmental impacts and decarbonization strategies in the cement and concrete industries. *Nature Reviews Earth & Environment*, Vol. 1, 2020, pp. 559–573.
- [3] Khan, Z., M. Umar, K. Shahzada, and A. Ali. Utilization of marble dust in fired clay bricks. *Environmental Monitoring*, Vol. 17, 2017, pp. 1–10.
- [4] Ahmad, T., M. Hussain, M. Iqbal, A. Ali, W. Manzoor, H. Bibi, et al. Environmental, energy, and water footprints of marble tile production chain in a life cycle perspective. *Sustainability*, Vol. 14, 2022, id. 8325.
- [5] Bilir, T., Ö. Karadağ, and B. F. Aygün. Waste marble powder. In *Sustainable concrete made with ashes and dust from different sources*, Elsevier, 2022, pp. 479–506.
- [6] Fawad, M., F. Ullah, W. Shah, M. Irshad, Q. Mehmood, A. A. Tahir, et al. Impacts of the marble waste slurry on ground water quality and its reuse potential. *Fresenius Environmental Bulletin*, Vol. 30, 2021, pp. 2077–2086.
- [7] Liang, X., X. Yu, B. Xu, C. Chen, G. Ding, Y. Jin, et al. Storage stability and compatibility in foamed warm-mix asphalt containing recycled asphalt pavement binder. *Journal of Materials in Civil Engineering*, Vol. 36, 2024, id. 04024062.

- [8] Kushwah, E. R. P. S. Scientific disposal system of marble slurry for clean and green environment. *International Journal of Engineering Sciences & Research Technology*, Vol. 3, 2014, pp. 500–503.
- [9] Zornoza, R., A. Faz, D. M. Carmona, S. Martínez-Martínez, and J. A. Acosta. Plant cover and soil biochemical properties in a mine tailing pond five years after application of marble wastes and organic amendments. *Pedosphere*, Vol. 22, 2012, pp. 22–32.
- [10] Huang, J. D., M. M. Zhou, H. W. Yuan, M. M. S. Sabri, and X. Li. Towards sustainable construction materials: A comparative study of prediction models for green concrete with metakaolin. *Buildings*, Vol. 12, No. 6, 2022, id. 772.
- [11] Neville, A. M. and J. J. Brooks. *Concrete technology*, Vol. 438, Longman Scientific & Technical, England, 1987.
- [12] Bouazza, N., A. El Mrihi, and A. Maâte. Geochemical assessment of limestone for cement manufacturing. *Procedia Technology*, Vol. 22, 2016, pp. 211–218.
- [13] Kore, S. D. and A. K. Vyas. Impact of marble waste as coarse aggregate on properties of lean cement concrete. *Case studies in construction materials*, Vol. 4, 2016, pp. 85–92.
- [14] Khan, M. A., B. Khan, K. Shahzada, S. W. Khan, N. Wahab, and M. I. Ahmad. Conversion of waste marble powder into a binding material. *Civil Engineering Journal*, Vol. 6, 2020, pp. 431–445.
- [15] Wesselsky, A. and O. M. Jensen. Synthesis of pure Portland cement phases. *Cement and Concrete Research*, Vol. 39, 2009, pp. 973–980.
- [16] Mtarfi, N. H., Z. Rais, and M. Taleb. Effect of clinker free lime and cement fineness on the cement physicochemical properties. *Journal of Materials and Environmental Science*, Vol. 8, 2017, pp. 2541–2548.
- [17] Huang, J. D., J. Zhang, and Y. Gao. Evaluating the clogging behavior of pervious concrete (PC) using the machine learning techniques. *CMES-Computer Modeling in Engineering & Sciences*, Vol. 130, 2022, pp. 805–821.
- [18] Wang, R., J. Zhang, Y. Lu, S. Ren, and J. Huang. Towards a reliable design of geopolymer concrete for green landscapes: A comparative study of tree-based and regression-based models. *Buildings*, Vol. 14, No. 3, 2024, id. 615.
- [19] Khan, M. A., S. Ayub Khan, B. Khan, K. Shahzada, F. Althoeay, and A. F. Deifalla. Investigating the feasibility of producing sustainable and compatible binder using marble waste, fly ash, and rice husk ash: A comprehensive research for material characteristics and production. *Results in Engineering*, Vol. 20, 2023, id. 101435.
- [20] Sepehri, A. and M.-H. Sarrafzadeh. Effect of nitrifiers community on fouling mitigation and nitrification efficiency in a membrane bioreactor. *Chemical Engineering and Processing-Process Intensification*, Vol. 128, 2018, pp. 10–18.
- [21] Lao, J.-C., B.-T. Huang, L.-Y. Xu, M. Khan, Y. Fang, and J.-G. Dai. Seawater sea-sand Engineered Geopolymer Composites (EGC) with high strength and high ductility. *Cement and Concrete Composites*, Vol. 138, 2023, id. 104998.
- [22] Sabri, M. Z. H. A., R. A. Malek, A. A. Omar, and K. N. Ismail. Study of fly ash concrete exposed to elevated temperature, *Key Engineering Materials*, Vol. 908, 2022, pp. 645–650.
- [23] American Coal Ash, A. *Fly ash facts for highway engineers*, US Department of Transportation, Federal Highway Administration, 2003.
- [24] Fuad, M. Y. A., Z. Ismail, Z. A. M. Ishak, and A. K. M. Omar. Rice husk ash. *Plastics Additives: An AZ Reference*, 1998, pp. 561–566.
- [25] Kone, B., J. N. Mwero, and E. K. Ronoh. Experimental effect of cassava starch and rice husk ash on physical and mechanical properties of concrete. *International Journal of Engineering Trends and Technology*, Vol. 70, 2022, pp. 343–350.
- [26] Yuan, X., Y. Tian, W. Ahmad, A. Ahmad, K. I. Usanova, A. M. Mohamed, et al. Machine learning prediction models to evaluate the strength of recycled aggregate concrete. *Materials*, Vol. 15, 2022, id. 2823.
- [27] Nguyen, N.-H., T. P. Vo, S. Lee, and P. G. Asteris. Heuristic algorithm-based semi-empirical formulas for estimating the compressive strength of the normal and high performance concrete. *Construction and Building Materials*, Vol. 304, 2021, id. 124467.
- [28] Emad, W., A. S. Mohammed, A. Bras, P. G. Asteris, R. Kurda, Z. Muhammed, et al. Metamodel techniques to estimate the compressive strength of UHPFRC using various mix proportions and a high range of curing temperatures. *Construction and Building Materials*, Vol. 349, 2022, id. 128737.
- [29] Huang, J., M. Zhou, J. Zhang, J. Ren, N. I. Vatin, and M. M. S. Sabri. Development of a new stacking model to evaluate the strength parameters of concrete samples in laboratory. *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, Vol. 46, 2022, pp. 4355–4370.
- [30] Zhu, F., X. Wu, Y. Lu, and J. Huang. Strength reduction due to acid attack in cement mortar containing waste eggshell and glass: A machine learning-based modeling study. *Buildings*, Vol. 14, 2024, id. 225.
- [31] Zhu, F., X. Wu, Y. Lu, and J. Huang. Strength estimation and feature interaction of carbon nanotubes-modified concrete using artificial intelligence-based boosting ensembles. *Buildings*, Vol. 14, 2024, id. 134.
- [32] Singh, N., P. Kumar, and P. Goyal. Reviewing the behaviour of high volume fly ash based self compacting concrete. *Journal of Building Engineering*, Vol. 26, 2019, id. 100882.
- [33] Althoeay, F. Compressive strength reduction of cement pastes exposed to sodium chloride solutions: Secondary ettringite formation. *Construction and Building Materials*, Vol. 299, 2021, id. 123965.
- [34] Shi, X., S. Chen, Q. Wang, Y. Lu, S. Ren, and J. Huang. Mechanical framework for geopolymer gels construction: An optimized LSTM technique to predict compressive strength of fly ash-based geopolymer gels concrete. *Gels*, Vol. 10, No. 2, 2024, id. 148.
- [35] Awoyera, P. O. Nonlinear finite element analysis of steel fibre-reinforced concrete beam under static loading. *Journal of Engineering Science and Technology*, Vol. 11, 2016, pp. 1669–1677.
- [36] Huang, J., J. Zhang, X. Li, Y. Qiao, R. Zhang, and G. S. Kumar. Investigating the effects of ensemble and weight optimization approaches on neural networks' performance to estimate the dynamic modulus of asphalt concrete. *Road Materials and Pavement Design*, Vol. 24, 2023, pp. 1939–1959.
- [37] Tian, Q., Y. J. Lu, J. Zhou, S. T. Song, L. M. Yang, T. Cheng, et al. Exploring the viability of AI-aided genetic algorithms in estimating the crack repair rate of self-healing concrete. *Reviews on Advanced Materials Science*, Vol. 63, No. 1, 2024, id. 179.
- [38] Amin, M. N., W. Ahmad, K. Khan, M. N. Al-Hashem, A. F. Deifalla, and A. Ahmad. Testing and modeling methods to experiment the flexural performance of cement mortar modified with eggshell powder. *Case Studies in Construction Materials*, Vol. 18, 2023, id. e01759.
- [39] Emad, W., A. S. Mohammed, R. Kurda, K. Ghafor, L. Cavaleri, S. M. A. Qaidi, et al. Prediction of concrete materials compressive strength using surrogate models, *Structures*, Vol. 46, 2022, pp. 1243–1267.
- [40] Asteris, P. G., P. C. Roussis, and M. G. Douvika. Feed-forward neural network prediction of the mechanical properties of sandcrete materials. *Sensors*, Vol. 17, 2017, id. 1344.

- [41] Javed, M. F., M. N. Amin, M. I. Shah, K. Khan, B. Iftikhar, F. Farooq, et al. Applications of gene expression programming and regression techniques for estimating compressive strength of bagasse ash based concrete. *Crystals*, Vol. 10, 2020, id. 737.
- [42] Zou, B., Y. Wang, M. N. Amin, B. Iftikhar, K. Khan, M. Ali, et al. Artificial intelligence-based optimized models for predicting the slump and compressive strength of sustainable alkali-derived concrete. *Construction and Building Materials*, Vol. 409, 2023, id. 134092.
- [43] Shah, S., M. Houda, S. Khan, F. Althoey, M. Abuhussain, M. A. Abuhussain, et al. Mechanical behaviour of E-waste aggregate concrete using a novel machine learning algorithm: Multi expression programming (MEP). *Journal of Materials Research and Technology*, Vol. 25, 2023, pp. 5720–5740.
- [44] Nazar, S., J. Yang, X.-E. Wang, K. Khan, M. N. Amin, M. F. Javed, et al. Estimation of strength, rheological parameters, and impact of raw constituents of alkali-activated mortar using machine learning and SHapely Additive exPlanations (SHAP). *Construction and Building Materials*, Vol. 377, 2023, id. 131014.
- [45] Huang, J. D., M. M. Zhou, M. M. S. Sabri, and H. W. Yuan. A novel neural computing model applied to estimate the dynamic modulus (DM) of asphalt mixtures by the improved beetle antennae search. *Sustainability*, Vol. 14, No. 10, 2022, id. 5938.
- [46] Xu, W. J., X. Huang, Z. J. Yang, M. M. Zhou, and J. D. Huang. Developing hybrid machine learning models to determine the dynamic modulus (E^*) of asphalt mixtures using parameters in Witczak 1-40D model: A comparative study. *Materials*, Vol. 15, No. 5, 2022, id. 1791.
- [47] Huang, J. D., X. Li, J. Zhang, Y. N. Sun, and J. L. Ren. Determining the Rayleigh damping parameters of flexible pavements for finite element modeling. *Journal of Vibration and Control*, Vol. 28, 2022, pp. 3181–3194.
- [48] Huang, J. D., J. Zhang, and Y. Gao. Intelligently predict the rock joint shear strength using the support vector regression and firefly algorithm. *Lithosphere*, Vol. 2021, 2021, id. 2467126.
- [49] Wang, Q., T. Cheng, Y. Lu, H. Liu, R. Zhang, and J. Huang. Underground mine safety and health: A hybrid MERECo-CoCoSo system for the selection of best sensor. *Sensors*, Vol. 24, 2024, id. 1285.
- [50] Wang, Q. A., C. Zhang, Z. G. Ma, J. D. Huang, Y. Q. Ni, and C. Zhang. SHM deformation monitoring for high-speed rail track slabs and Bayesian change point detection for the measurements. *Construction and Building Materials*, Vol. 300, 2021, id. 124337.
- [51] Khan, S. A., M. A. Khan, M. N. Amin, M. Ali, F. Althoey, and F. Alsharari. Sustainable alternate binding material for concrete using waste materials: A testing and computational study for the strength evaluation. *Journal of Building Engineering*, Vol. 80, 2023, id. 1285.
- [52] Zhou, J., Z. Su, S. Hosseini, Q. Tian, Y. Lu, H. Luo, et al. Decision tree models for the estimation of geo-polymer concrete compressive strength. *Mathematical Biosciences and Engineering*, Vol. 21, 2024, pp. 1413–1444.
- [53] Armaghani, D. J. and P. G. Asteris. A comparative study of ANN and ANFIS models for the prediction of cement-based mortar materials compressive strength. *Neural Computing and Applications*, Vol. 33, 2021, pp. 4501–4532.
- [54] Asteris, P. G., A. D. Skentou, A. Bardhan, P. Samui, and K. Pilakoutas. Predicting concrete compressive strength using hybrid ensembling of surrogate machine learning models. *Cement and Concrete Research*, Vol. 145, 2021, id. 106449.
- [55] Ji, Z., M. Zhou, Q. Wang, and J. Huang. Predicting the international roughness index of JPCP and CRCP rigid pavement: A Random Forest (RF) Model Hybridized with Modified Beetle Antennae Search (MBAS) for higher accuracy. *Computer Modeling in Engineering & Sciences*, Vol. 139, 2024, pp. 1557–1582.
- [56] Wu, X. P., F. Zhu, M. M. Zhou, M. M. S. Sabri, and J. D. Huang. Intelligent design of construction materials: A comparative study of AI approaches for predicting the strength of concrete with blast furnace slag. *Materials*, Vol. 15, 2022, id. 4582.
- [57] Zhang, J., R. Wang, Y. Lu, and J. Huang. Prediction of compressive strength of geopolymer concrete landscape design: Application of the Novel Hybrid RF–GWO–XGBoost algorithm. *Buildings*, Vol. 14, 2024, id. 591.
- [58] Huang, J. D., G. S. Kumar, J. L. Ren, J. F. Zhang, and Y. T. Sun. Accurately predicting dynamic modulus of asphalt mixtures in low-temperature regions using hybrid artificial intelligence model. *Construction and Building Materials*, Vol. 297, 2021, id. 123655.
- [59] Huang, J., M. M. Sabri, D. V. Ulrikh, M. Ahmad, and K. A. Alsaffar. Predicting the compressive strength of the cement-fly Ash–Slag ternary concrete using the Firefly Algorithm (FA) and Random Forest (RF) hybrid machine-learning method. *Materials*, Vol. 15, 2022, id. 4193.
- [60] Wang, R., J. Zhang, Y. Lu, and J. Huang. Towards designing durable sculptural elements: Ensemble learning in predicting compressive strength of fiber-reinforced nano-silica modified concrete. *Buildings*, Vol. 14, 2024, id. 396.
- [61] Amin, M. N., W. Ahmad, K. Khan, and A. F. Deifalla. Optimizing compressive strength prediction models for rice husk ash concrete with evolutionary machine intelligence techniques. *Case Studies in Construction Materials*, Vol. 18, 2023, id. e02102.
- [62] Holland, J. H. *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*, MIT Press, Cambridge, MA, USA, 1992.
- [63] Koza, J. On the programming of computers by means of natural selection. *Genetic programming*, MIT Press, Cambridge, MA, USA, 1992.
- [64] Huang, J. D., M. M. Zhou, J. Zhang, J. L. Ren, N. I. Vatin, and M. M. S. Sabri. The use of GA and PSO in evaluating the shear strength of steel fiber reinforced concrete beams. *KSCE Journal of Civil Engineering*, Vol. 26, 2022, pp. 3918–3931.
- [65] Gholampour, A., T. Ozbakkaloglu, and R. Hassanli. Behavior of rubberized concrete under active confinement. *Construction and Building Materials*, Vol. 138, 2017, pp. 372–382.
- [66] Topcu, I. B. and M. Saridemir. Prediction of compressive strength of concrete containing fly ash using artificial neural networks and fuzzy logic. *Computational Materials Science*, Vol. 41, 2008, pp. 305–311.
- [67] Ferreira, C. *Gene expression programming: mathematical modeling by an artificial intelligence*, Vol. 21, Springer, Warsaw, Poland, 2006.
- [68] Gandomi, A. H., G. J. Yun, and A. H. Alavi. An evolutionary approach for modeling of shear strength of RC deep beams. *Materials and Structures*, Vol. 46, 2013, pp. 2109–2119.
- [69] Gandomi, A. H., S. K. Babanajad, A. H. Alavi, and Y. Farnam. Novel approach to strength modeling of concrete under triaxial compression. *Journal of Materials in Civil Engineering*, Vol. 24, 2012, pp. 1132–1143.
- [70] Wang, H.-L. and Z.-Y. Yin. High performance prediction of soil compaction parameters using multi expression programming. *Engineering Geology*, Vol. 276, 2020, id. 105758.
- [71] Iqbal, M. F., M. F. Javed, M. Rauf, I. Azim, M. Ashraf, J. Yang, et al. Sustainable utilization of foundry waste: Forecasting mechanical

- properties of foundry sand based concrete using multi-expression programming. *Science of the Total Environment*, Vol. 780, 2021, id. 146524.
- [72] Oltean, M. and C. Grosan. A comparison of several linear genetic programming techniques. *Complex Systems*, Vol. 14, 2003, pp. 285–314.
- [73] Fallahpour, A., E. U. Olugu, and S. N. Musa. A hybrid model for supplier selection: integration of AHP and multi expression programming (MEP). *Neural Computing and Applications*, Vol. 28, 2017, pp. 499–504.
- [74] Alavi, A. H., A. H. Gandomi, M. G. Sahab, and M. Gandomi. Multi expression programming: A new approach to formulation of soil classification. *Engineering with Computers*, Vol. 26, 2010, pp. 111–118.
- [75] Mohammadzadeh S. D., S.-F. Kazemi, A. Mosavi, E. Nasseralshariati, and J. H. M. Tah. Prediction of compression index of fine-grained soils using a gene expression programming model. *Infrastructures*, Vol. 4, 2019, id. 26.
- [76] Grosan, C., A. Abraham. Stock market modeling using genetic programming ensembles. In *Genetic Systems Programming: Theory and Experiences*, Springer, 2006, pp. 131–146.
- [77] Oltean, M. and D. Dumitrescu. Multi expression programming. *Journal of Genetic Programming and Evolvable Machines*, 2002.
- [78] Iqbal, M. F., Q.-f. Liu, I. Azim, X. Zhu, J. Yang, M. F. Javed, et al. Prediction of mechanical properties of green concrete incorporating waste foundry sand based on gene expression programming. *Journal of Hazardous Materials*, Vol. 384, 2020, id. 121322.
- [79] Shahin, M. A. *Genetic programming for modelling of geotechnical engineering systems*, Springer, Cham, Berlin, Germany, 2015.
- [80] Çanakcı, H., A. Baykasoğlu, and H. Güllü. Prediction of compressive and tensile strength of Gaziantep basalts via neural networks and gene expression programming. *Neural Computing and Applications*, Vol. 18, 2009, pp. 1031–1041.
- [81] Alade, I. O., M. A. Abd Rahman, and T. A. Saleh. Predicting the specific heat capacity of alumina/ethylene glycol nanofluids using support vector regression model optimized with Bayesian algorithm. *Solar Energy*, Vol. 183, 2019, pp. 74–82.
- [82] Alade, I. O., A. Bagudu, T. A. Oyehan, M. A. Abd Rahman, T. A. Saleh, and S. O. Olatunji. Estimating the refractive index of oxygenated and deoxygenated hemoglobin using genetic algorithm–support vector regression model. *Computer Methods and Programs in Biomedicine*, Vol. 163, 2018, pp. 135–142.
- [83] Liang, X. M., X. Yu, G. Y. Ding, Y. Jing, and J. D. Huang. Environmental and mechanical effects of rubberised open-graded asphalt mixtures incorporating with titanium dioxide: a laboratory investigation. *International Journal of Pavement Engineering*, Vol. 24, 2023, id. 2241604.
- [84] Zhang, W., R. Zhang, C. Wu, A. T. C. Goh, S. Lacasse, Z. Liu, et al. State-of-the-art review of soft computing applications in underground excavations. *Geoscience Frontiers*, Vol. 11, 2020, pp. 1095–1106.
- [85] Alavi, A. H., A. H. Gandomi, H. C. Nejad, A. Mollahasani, and A. Rashed. Design equations for prediction of pressuremeter soil deformation moduli utilizing expression programming systems. *Neural Computing and Applications*, Vol. 23, 2013, pp. 1771–1786.
- [86] Kisi, O., J. Shiri, and M. Tombul. Modeling rainfall-runoff process using soft computing techniques. *Computers & Geosciences*, Vol. 51, 2013, pp. 108–117.
- [87] Alade, I. O., M. A. Abd Rahman, and T. A. Saleh. Modeling and prediction of the specific heat capacity of Al₂O₃/water nanofluids using hybrid genetic algorithm/support vector regression model. *Nano-Structures & Nano-Objects*, Vol. 17, 2019, pp. 103–111.
- [88] Shahin, M. A. Use of evolutionary computing for modelling some complex problems in geotechnical engineering. *Geomechanics and Geoengineering*, Vol. 10, 2015, pp. 109–125.
- [89] Huang, J. D., M. M. Zhou, H. W. Yuan, M. M. S. Sabri, and X. Li. Prediction of the compressive strength for cement-based materials with metakaolin based on the hybrid machine learning method. *Materials*, Vol. 15, 2022.
- [90] Asteris, P. G., M. Koopialipoor, D. J. Armaghani, E. A. Kotsonis, and P. B. Lourenço. Prediction of cement-based mortars compressive strength using machine learning techniques. *Neural Computing and Applications*, Vol. 33, 2021, pp. 13089–13121.
- [91] Band, S. S., E. Heggy, S. M. Bateni, H. Karami, M. Rabiee, S. Samadianfard, et al. Groundwater level prediction in arid areas using wavelet analysis and Gaussian process regression. *Engineering Applications of Computational Fluid Mechanics*, Vol. 15, 2021, pp. 1147–1158.
- [92] Taylor, K. E. Summarizing multiple aspects of model performance in a single diagram. *Journal of geophysical research: atmospheres*, Vol. 106, 2001, pp. 7183–7192.
- [93] Huang, J. D. and J. H. Xue. Optimization of SVR functions for flyrock evaluation in mine blasting operations. *Environmental Earth Sciences*, Vol. 81, 2022, id. 434.
- [94] Jin, C., Y. Qian, S. A. Khan, W. Ahmad, F. Althoey, B. S. Alotaibi, et al. Investigating the feasibility of genetic algorithms in predicting the properties of eco-friendly alkali-based concrete. *Construction and Building Materials*, Vol. 409, 2023, id. 134101.
- [95] Jin, C., Y. Qian, K. Khan, A. Ahmad, M. N. Amin, F. Althoey, et al. Predicting the damage to cementitious composites due to acid attack and evaluating the effectiveness of eggshell powder using interpretable artificial intelligence models. *Materials Today Communications*, Vol. 37, 2023, id. 107333.
- [96] Ali, M., A. Kumar, A. Yvaz, and B. Salah. Central composite design application in the optimization of the effect of pumice stone on lightweight concrete properties using RSM. *Case Studies in Construction Materials*, Vol. 18, 2023, id. e01958.
- [97] Ali, M., M. I. Khan, F. Masood, B. T. Alsulami, B. Bouallegue, R. Nawaz, et al. Central composite design application in the optimization of the effect of waste foundry sand on concrete properties using RSM. *Structures*, Vol. 46, 2022, pp. 1581–1594.
- [98] Ali, M., S. Abbas, B. Salah, J. Akhter, W. Saleem, S. Haruna, et al. Investigating optimal confinement behaviour of low-strength concrete through quantitative and analytical approaches. *Materials*, Vol. 14, 2021, id. 4675.
- [99] Ahmad, A., K. A. Ostrowski, M. Maślak, F. Farooq, I. Mehmood, and A. Nafees. Comparative study of supervised machine learning algorithms for predicting the compressive strength of concrete at high temperature. *Materials*, Vol. 14, 2021, id. 4222.