Research Article

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

Exploring the viability of AI-aided genetic algorithms in estimating the crack repair rate of self-healing concrete

https://doi.org/10.1515/rams-2023-0179
received December 09, 2023; accepted January 10, 2024

Abstract: As a potential replacement for traditional concrete, which has cracking and poor durability issues, self-healing concrete (SHC) has been the research subject. However, conducting lab trials can be expensive and time-consuming. Therefore, machine learning (ML)-based predictions can aid improved formulations of self-healing concrete. The aim of this work is to develop ML models that could analyze and forecast the rate of healing of the cracked area (CrA) of bacteria- and fiber-containing SHC. These models were constructed using gene expression programming (GEP) and multi-expression programming (MEP) tools. The discrepancy between expected and desired results, statistical tests, Taylor's diagram, and R^2 values were additional metrics used to assess the constructed models. A SHapley Additive exPlanations (SHAP) approach was used to evaluate which input attributes were highly relevant. With R^2 = 0.93, MAE = 0.047, MAPE = 12.60%, and RMSE = 0.062, the GEP produced somewhat worse predictions than the MEP (R^2 = 0.93, MAE = 0.033, MAPE = 9.60%, and RMSE = 0.044). Bacteria had an indirect (negative) relationship with the CrA of SHC, while fiber had a direct (positive) association, according to the SHAP study. The SHAP study might help researchers and companies figure out how much of each raw material is needed for SHCs. Therefore, MEP and GEP models can be used to generate and test SHC compositions based on bacteria and polymeric fibers.

Keywords: machine learning, self-healing concrete, crack repair

*** 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

1 Introduction

For numerous reasons, including its low cost, superior compressive strength, and excellent workability, concrete is the most used building material on the planet [1–3]. Unfortunately, fractures can appear in concrete at any time, which reduces its durability [3–5]. In addition, fixing fractured concrete structures is an expensive and time-consuming ordeal because of all the complicated operations involved [6]. In recent years, there has been a lot of interest in creating concrete with a self-healing ability to patch fractures; this is because discovering an effective repair method might reduce maintenance costs, environmental impact, and energy usage [7–10]. It is possible to grade the severity of cracks in concrete. According to Bayar and Bilir [11] and Rasol *et al.* [6], concrete's durability and resistance are significantly impacted by internal cracks and cracks bigger than 0.2 mm, which are classified as serious. In addition, concrete structures can lose stiffness and have their service lives shortened due to cracks, which can reduce their penetration resistance against chlorides [12]. Engineers have been looking for ways to fix concrete fractures and prevent them from occurring.

The most important component in crack mending is the precipitation of calcium carbonate ($CaCO_3$). Feng *et al.* [13], Huang *et al.* [14], Jamshidi *et al.* [15], Rasol *et al.* [6], and Su *et al.* [16] all found that by incorporating materials for example organic and artificial fibers, mineral blends, and micro-organisms into the concrete preparation process, it becomes possible for the concrete to self-heal through the precipitation of $CaCO_3$. Microbes that can withstand acidic conditions help fill the micro-cracks and bind

aggregates like sand and gravel to concrete by microbiologically precipitating calcium carbonate [13]. Due to the irreversible nature of the hardening and hydration processes, any dosage errors in the complicated blend of cement, water, coarse aggregate, and fine aggregate, which is typically enhanced with additives, can result in extremely costly mistakes while making concrete. Furthermore, these mistakes might reduce the structure's durability, which in turn reduces its future usability and significantly impacts building costs. A simplified representation of self-healing events in concrete is shown in Figure 1.

Therefore, to get around these mistakes, researchers are always looking for new kinds of concrete to meet the demanding requirements of the construction business [15,18–20]. Forecasting models that can estimate the properties of concrete are also being developed. Machine learning (ML) and artificial intelligence (AI) have emerged as popular techniques for predictive model construction, joining computational simulation and statistical estimation. ML modeling is quickly replacing traditional methods as the go-to for building predictive models in the age of big data and data-driven science and engineering [21–24]. Commonly, ML is concerned with training data samples that incorporate various techniques to construct a learning model that can enhance itself when presented with fresh data [25–27]. In this way, ML can be of service to civil engineers in their estimation of concrete and other material qualities. Concerning both time and money, this is preferable. Civil engineers have begun to focus on ML and AI methods for predicting concrete's mechanical properties [28–31]. These methods include decision trees, support vector machines, random forests, multi-expression programming (MEP), and gene expression programming (GEP) [32–36]. Furthermore, these methods can produce accurate results. Additionally, civil engineers have trained and utilized ML algorithms to create novel ultra-high-performance concrete, utilize artificial neural networks to forecast the strength of fiber-reinforced polymers, and ascertain the ultimate buckling stress of composite cylinders with varying stiffness [19,37–39]. Thus, self-healing concrete (SHC) can likewise benefit from ML modeling in its design and development processes. Huang *et al.* [40] and Zhuang *et al.* [41] created and published ML models to predict how bacteria affect the self-healing capacity of cementitious materials, taking into account the substantial literature on the subject. While these investigations did a good job of identifying the types of bacteria involved, they failed to consider how other aggregates, such as fibers and polymers, would affect the cementitious materials' self-healing capacities.

Using ML algorithms, namely, MEP and GEP, this study intended to predict how quickly cracked areas (CrAs) will be repaired in SHC that has been altered with polymer fibers (*e.g.*, PVA and PP fibers) and two alkali-resistant bacteria (Bacillus alcalophilus and Bacillus cereus). The data used for this prediction come from the existing literature. To evaluate ML algorithms' efficacy, many metrics were used, including the R^2 coefficient, statistical tests, and the dispersion of anticipated results. The rationale for this research was to find out how well ML methods work for accurately predicting material properties. ML methods necessitate a dataset, which can be generated by exploratory experiments or by analyzing existing databases. ML models could have a better idea of material properties by analyzing this dataset. The capacity of ML techniques to forecast CrA on SHC was evaluated using experimental data and six input factors. Further investigation into

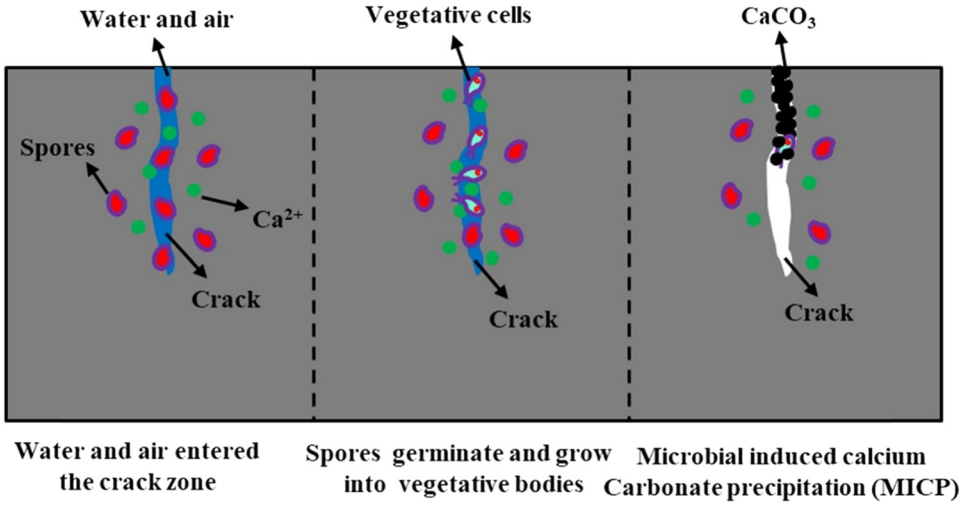


Figure 1: A graphic representation of a self-healing fracture [17].

the significance of raw materials was carried out through the use of SHapley Additive exPlanations (SHAP) analysis. Improving the existing SHC database and maybe guiding the formulation of SHC mixes are two possible uses for the newly acquired features and the built ML models.

2 Methods of research

2.1 Collecting and analyzing data

Predicting the CrA of the SHC was the goal of this research, which employed MEP and GEP simulations to examine a collection of 684 data points gathered from an experiment [13,16,42]. This study predicted the CrA centered on six input factors such as cement (OPC), sand (FA), Water (W), time (T), bacteria (B), and fiber (F). Data preparation was crucial in collecting and organizing the information. Data preparation for data mining is a common tactic in the famous knowledge discovery from data technique, which helps to circumvent a major hurdle. Data preparation is to remove unnecessary features and noise from the data to make it more understandable and less complicated [43–45]. The automatic scaling and standardization of datasets can significantly affect ML models if performed without proper data preparation. Because the model overvalues factors with greater numerical scales, it overestimates their impact on prediction. Without sufficient normalization, the model may struggle to weigh feature contributions, skewing learning, and restricting generalization [46,47]. When analyzing the model, regression and error-distribution techniques were used. Several descriptive metrics were computed using these data, and their findings are shown in Table 1. The validity of the used models has also been evaluated for their effectiveness. The frequency of different values is shown by the histograms in Figure 2. A dataset's total frequency breakdown can be described by integrating its component distributions. One way to see how common certain values are in a set is to look at their relative frequency distribution.

2.2 ML modeling

CrA of SHC was examined in the laboratory. The production of CrA, the end product, required six inputs. To forecast the CrA of concrete, sophisticated ML methods like GEP and MEP were employed. Results are often evaluated using ML algorithms concerning input data. 70% of the data were used for training the ML models, while only 30% were used for testing. The R^2 value of the predicted result proved that the model was accurate. A small R^2 number specifies a large disparity [48,49], while a large value indicates a high degree of agreement between the expected and actual results. Regularization techniques such as dropout or L1/L2 regularization were probably employed to handle over-fitting in the suggested model. It is likely that generalizability across different data subsets was evaluated by cross-validation [50,51]. Reliability in real-world applications was ensured through careful hyper-parameter tweaking and validation set monitoring, which helped strike a compromise between model complexity and robustness. Improving performance outside of the training datasets requires this focus on over-fitting [52,53]. The accuracy of the model was confirmed by several means, including statistical tests and evaluations of errors. The hyper-parameters for the MEP model are displayed in Table 2. Figure 3 shows a flow diagram of the research strategy. The initial phase of this procedure entailed

Table 1: Data analysis of a collection of variables

Parameters	Cement (kg·m^{-3})	Sand (kg·m^{-3})	Water (kg·m^{-3})	Time (h)	Bacteria	Fiber	CrA (mm)
Mean value	600.00	1200.00	220.00	336.00	0.83	1.17	0.39
Standard error	0.00	0.00	1.91	7.55	0.03	0.04	0.01
Median	600.00	1200.00	220.00	336.00	0.50	1.00	0.36
Mode	600.00	1200.00	270.00	0.00	0.00	0.00	0.44
Standard deviation	0.00	0.00	50.04	197.57	0.90	1.07	0.22
Sample variance	0.00	0.00	2503.66	39033.07	0.81	1.14	0.05
Range	0.00	0.00	100.00	672.00	2.00	3.00	0.96
Minimum	600.00	1200.00	170.00	0.00	0.00	0.00	0.00
Maximum	600.00	1200.00	270.00	672.00	2.00	3.00	0.96
Sum	410400	820800	150480	229824	570	798	269
Count	684.00	684.00	684.00	684.00	684.00	684.00	684.00

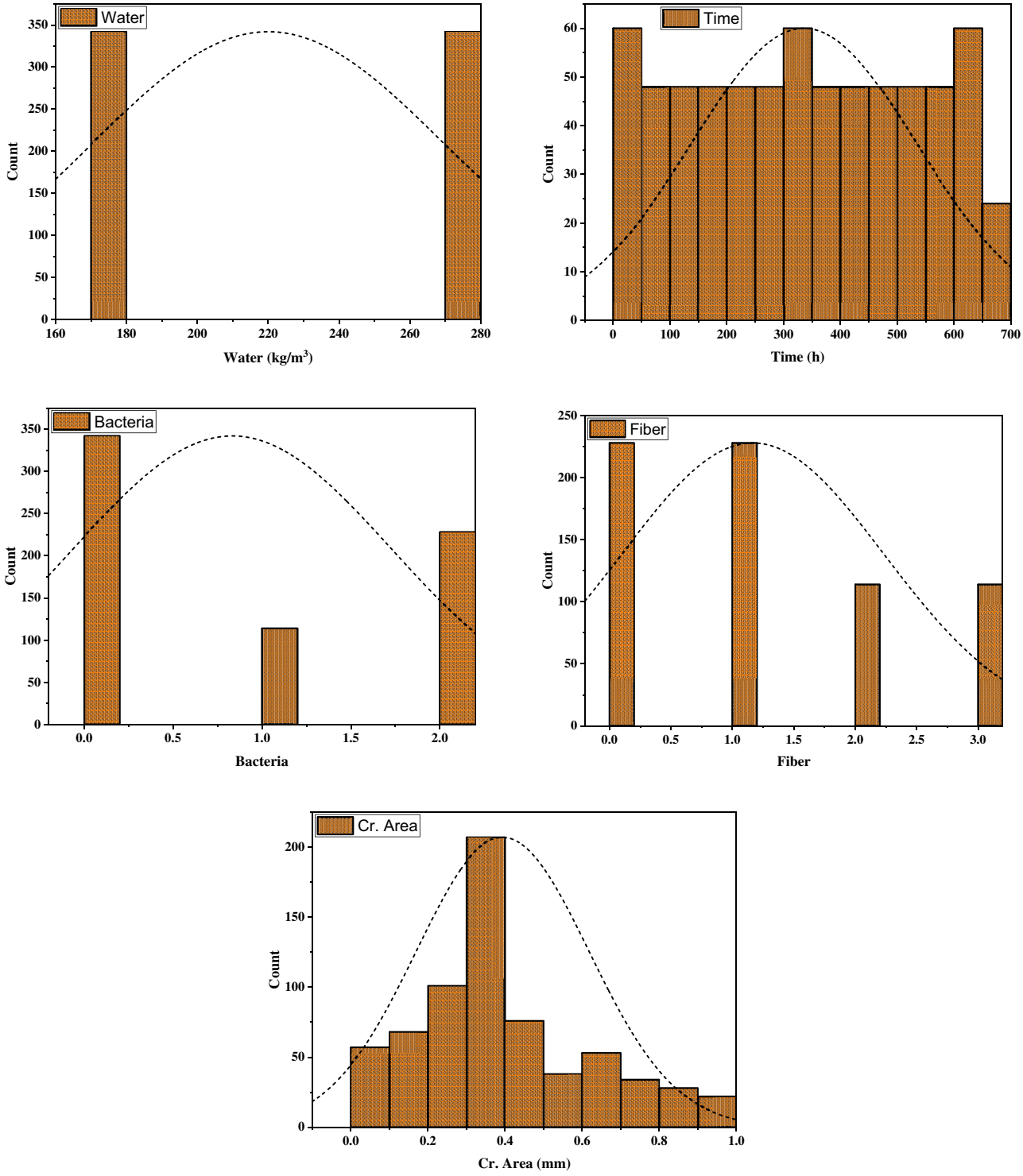


Figure 2: Frequency distribution of database input/output features.

generating a sample dataset. This pivotal preliminary stage established the groundwork for subsequent stages through the provision of essential data required for model training and evaluation. Subsequently, the models were constructed utilizing ML techniques, *i.e.*, MEP and GEP. These techniques enhanced the resilience and precision of the models, guaranteeing a holistic strategy for managing varied datasets. Afterward, model validation was carried out through the utilization of Taylor's diagram and statistical parameters. This stage is of utmost importance in evaluating the efficacy

Table 2: Set of parameters for the MEP and GEP techniques

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

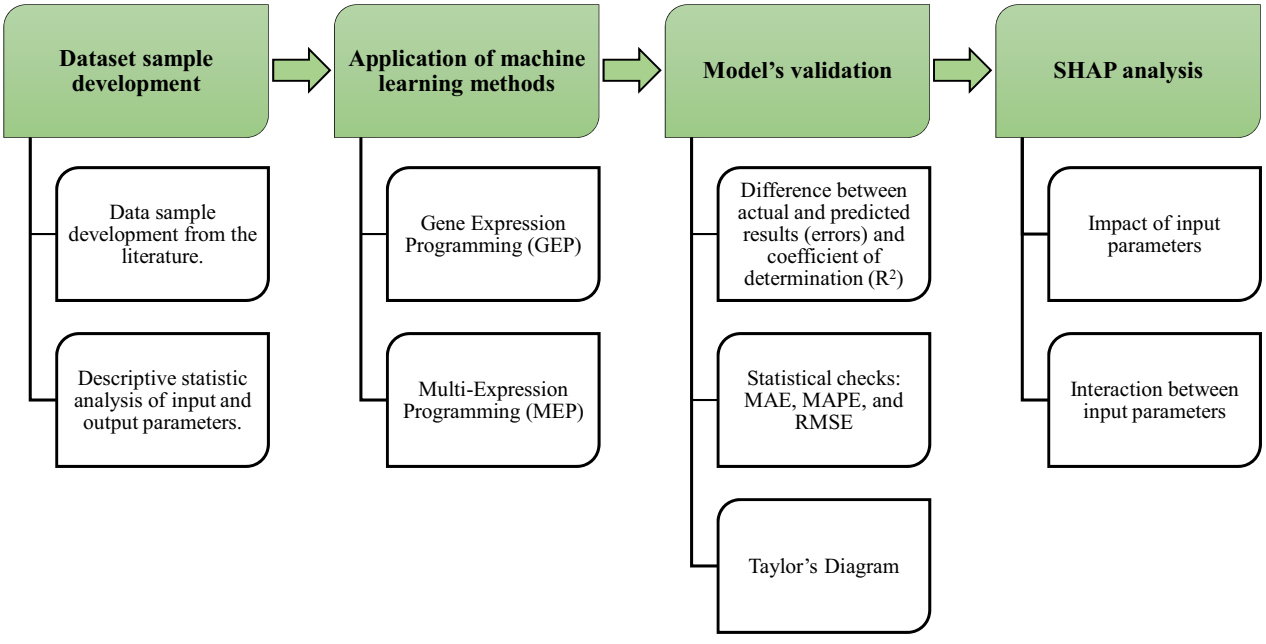


Figure 3: An overview of the ML methodology.

and dependability of the models. Statistical parameters provide numerical assessments of the accuracy of a model, whereas Taylor's diagram illustrates the model's visual capability of detecting patterns and fluctuations within the data [54,55]. Finally, SHAP analysis was performed in order to ascertain and comprehend the influence that input variables have on the outputs of the developed models. By revealing the extent to which each input variable influences the model's predictions, SHAP values facilitated the explanation and interpretation of the model's decision-making procedure.

2.2.1 GEP model

Originating from Darwin's theory of evolution, the genetic algorithm (GA) was developed by Holland [56]. Chromosomes that are continuously longer indicate that the genomic process has been completed, which is signified by a succession of GAs. "Gene programming" is the name that Koza gave to a creative GA technique [57]. Generalized problem-solving (GP) uses GAs to generate an evolutionary model [58]. The flexibility of GP comes from its ability to utilize nonlinear structures like parse trees rather than constant-length binary strings. In agreement with Darwin's concept, well-recognized AI software solves reproduction-connected challenges using naturally occurring genomic components, such as reproduction, crossover, and change [59–61]. In GP, a plan is made to eliminate wasteful programs from the next iteration. Like in the prior example, replanting the area using the chosen technique involves cutting down the unwanted trees. Evolution, in contrast, safeguards early convergence [59,62]. Five crucial parameters need to be defined before the GP method can be applied. Essential tasks within the area, evaluation of fitness, main functional operators (such as populace size and verge), and outcomes generated by technique-particular endpoints [59,63]. A crossover genetic processor is responsible for most of the parse tree development, even if GP's model construction happens repeatedly [64]. Due to their dual role as genotype and phenotype, nonlinear GP forms complicate the manifestation of desired traits [62].

GEP is a variation in GP that Ferreira initially proposed [62]. The GEP model incorporates static-length lined chromosomes into parse trees in accordance with the population-generation theory. The original GP uses simple, fixed-length chromosomes to encrypt medium-sized software; GEP is a better form of that. A potential benefit of GEP is its ability to be used to develop arithmetical expressions that can accurately forecast multifaceted and nonlinear hitches [65,66]. Like GP, it has a fitness function, parameters, and a final set of conditions for termination. Although the GEP technique generates chromosomes with seemingly random numbers, they are recognized as such before production employing the "Karva" dialectal. A line of constant length is necessary for GEP to function. On the other hand, GP's code processing of data displays parse trees of varying lengths. The individual cords are defined as genomes of static length and then portray chromosomes using nonlinear appearance/construe trees with branched morphologies of different sizes [59]. There is distinct genetic information for each of these genotypes and a small number of additional phenol strains [62]. Expensive structural mutations or duplications are unnecessary, even if GEP can maintain genomic integrity from one generation to the next.

In a typical chromosome, the "head" and the "tail" are the two complementary regions. Amazingly, creatures with several genes can develop from a single chromosome [59]. Logic, mathematics, arithmetic, and Boolean processes are prearranged in these genetic factors. A genomic code operator assigns a cell a specific function. The development of empirical formulas is made possible by the fact that one newly discovered language, Karva, can deduce the contents of these chromosomes. Following the expression trees (ET), a leadership revolution occurs, and travel begins in Karva. Following the steps outlined in Eq. (1), ET positions the nodes in the underlying layer [65]. Even though GEP preserves genomic integrity, expensive structural mutations or duplications are unnecessary.

$$ET\text{GEP} = \log\left(i - \frac{3}{j}\right). \tag{1}$$

The fact that GEP's findings are not dependent on any previous relationships makes it a sophisticated ML method. Figure 4 depicts the steps that are used to create a GEP arithmetical expression. The number of chromosomes in a human being is fixed at birth. After confirming that these chromosomes are ETs, comprehensive health examinations can be carried out. Those who are physically and mentally fit have the upper hand when it comes to having children.

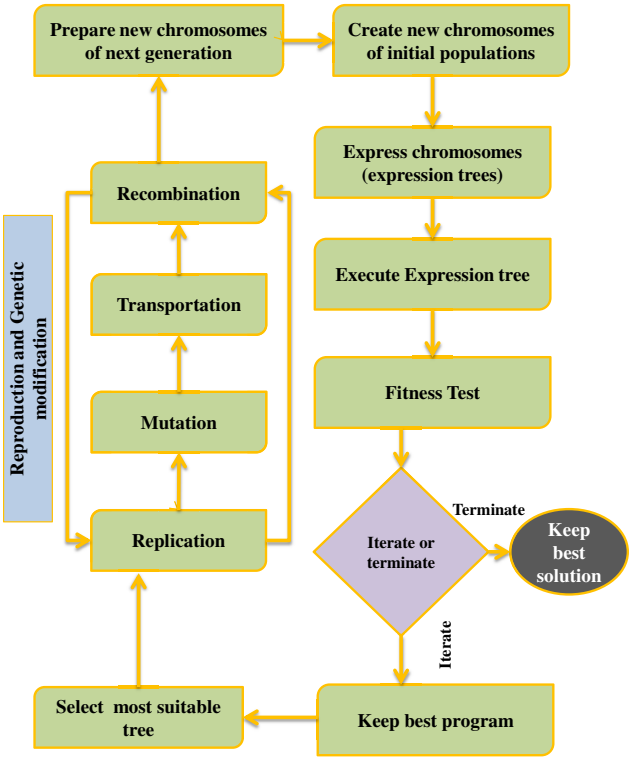


Figure 4: Flowchart of the GEP procedure [67].

An iterative procedure with the top experts yields the best answer. The numerical expression comes from mutation, crossing, and breeding over three generations.

2.2.2 MEP model

Considering it employs linear chromosomes, the MEP is a state-of-the-art, model linear-based GP method. The MEP and GEP are functionally identical with respect to their central software. Something that sets MEP apart from its forerunner, the GP approach, is its capacity to encode many software components (alternatives) into a single chromosome. Using fitness analysis to choose the optimal chromosome yields the desired result [68,69]. Oltean and Grosan [70] define this phenomenon as the process by which a bipolar scheme recombines to produce two diverse offspring [71]. Figure 5 demonstrates that the process will keep running until the termination form is met or until the best program is found. Here are the mutations that affect newborns. Similar to the GEP paradigm, the MEP model allows for the integration of several elements. 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 [72–74]. When there are as many packages as there are people in the population, evaluating them becomes more tedious and difficult. The size of the created mathematical expressions is affected by the code length, which is another crucial component. A comprehensive set of MEP parameters is required to accurately represent rheological properties, as shown in Table 2.

Both approaches rely heavily on literature datasets during the modeling and evaluation stages [75,76]. Popular linear GP methodologies like the MEP and GEP are deemed by some scholars to be superior for predicting the properties of viable concrete. Linguistic programming, in conjunction with maximum likelihood estimation (MEP), was determined by Grosan and Abraham to be the most effective neural network-based strategy [71,77]. The GEP's method of operation is marginally more intricate than that of the MEP [72]. Regardless of GEP having a higher density than MEP [78], dissimilarities encompass the capacity to reuse code in MEP, (i) the explicit encoding of function argument references in the MEP and (ii) the requirement that non-coding components not be shown at a static point inside the genes. Many people think the GEP chromosome is more powerful because of the symbols at its "head" and "tail" that make it easier to write software with the correct syntax [70]. This necessitates a more thorough evaluation of each of these genetic approaches to engineering challenges.

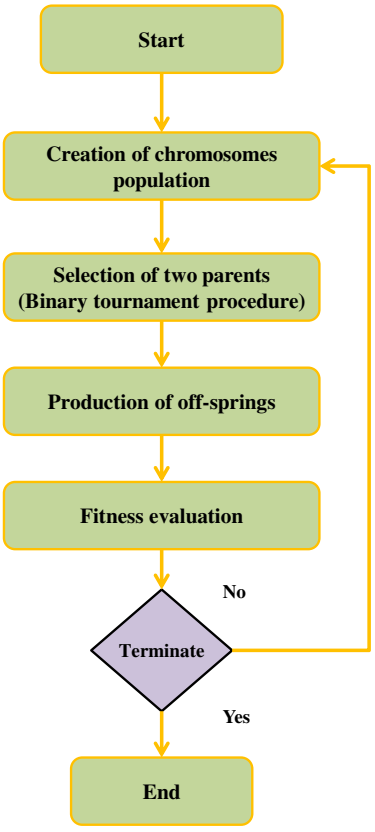


Figure 5: Flowchart of the MEP procedure [67].

2.3 Models validation

A test set was used for statistical testing of models that were constructed using GEP and MEP. Each created model had seven different statistical metrics computed [76,79–82]: Nash–Sutcliffe efficiency (NSE), Pearson's correlation coefficient (R), mean absolute error (MAE), relative squared error (RSE), mean absolute percentage error (MAPE), root mean square error (RMSE), and relative root mean square error (RRMSE). All these statistical measures have their formulations in Eqs (2)–(8).

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

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|P_i - Ti|, \tag{3}$$

$$\mathrm{RMSE} = \sqrt{\sum \frac{(P_i - T_i)^2}{n}}, \tag{4}$$

$$\mathrm{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\frac{|P_i - Ti|}{T_i}, \tag{5}$$

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

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

$$\mathrm{RRMSE} = \frac{1}{|\bar{a}|}\sqrt{\frac{\sum_{1=1}^{n}(a_i - p_i)^2}{n}}, \tag{8}$$

where n is the total data points number, a_i and p_i are the ith experimental and projected values, correspondingly; a_i also represents the mean experimental and projected values. The relationship coefficient, abbreviated as R, is a common way to measure a model's projection power (ai and pi). A high value of R indicates a robust relationship between the predicted and actual output amounts [83]. However, division and multiplication do not affect component R. R^2 was calculated using actual and projected outcomes since it better estimates the real value. R^2 values near 1 indicate a more effective model development process [84,85]. Similarly, when confronted with progressively more severe errors, both MAE and RMSE performed quite well. Less significant errors result in higher performance from the generated model and MAE and RMSE that are closer to zero [86,87]. However, upon closer inspection, it became apparent that continuous and smooth databases are where MAE truly excels [88]. When the values of the errors computed above are smaller, the model often performs better.

$$a20\text{-index} = \frac{m20}{M}. \tag{9}$$

An experimental value or anticipated value between 0.80 and 1.20 is taken into account, where M is the number of dataset samples and $m20$ is the number of samples. [89]. An ideal predictive model would anticipate that the a20-index values will be one hundredth of a percent. The suggested 20-index has the benefit of a physical engineering method, showing what proportion of samples match anticipated values within a ±20% margin of error from experimental data.

In conjunction with statistical validation, the Taylor diagram is among the most useful tools for determining a model's predictive power. 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 [90,91]. Standard deviation (x- and y-axes), correlation coefficient (radial lines), and RMSE indicate model placement. The model with the highest accuracy rate is the most reliable [90].

3 Results and discussion

3.1 CrA-GEP model

The GEP method yielded ETs-based models that calculated the CrA by deducing mathematical correlations from the chromosomal number and head size (Figure 6). The widely held sub-ETs in the SHC's CrA are built using the five arithmetical operations: ÷, x, −, +, and square root. An equation is the result of encrypting the GEP model's sub-ETs. It is possible to forecast the future CrA of SHC using the input data and the output value of these equations (Eqs (10)–(14)). With sufficient data, the produced model surpasses an ideal model operating under perfect circumstances. In Figure 7(a), the correlation between experimental and anticipated CrA is illustrated visually. A perfect match to the data is depicted by the solid black line, and the dotted lines reflect the percentage deviation (20%) from the perfect fit. The GEP model's predictions for CrA were quite similar to the measured values. The CrA of SHC was effectively determined using the GEP method, with an R^2 of 0.91 and 83% of its predictions within the 20% threshold, suggesting considerably higher accuracy. Figuring out how far the GEP model could be from reality is done by plotting the absolute error *vs* experimental data in Figure 7(b). The results showed that the GEP equation's predictions are quite close to the experimental results, with an absolute error that ranges from 0.00 to 0.182 mm and averages out to 0.047 mm. Figure 8 shows that the error values followed a bell-shaped spreading; 44 readings were less than 0.01 mm, 93 readings were between 0.01 and 0.05 mm, and 91 readings were greater than 0.05 mm. Importantly, maximal error frequencies really do not happen very often.

$$\mathrm{CrA(mm)} = A + B + C + D, \tag{10}$$

$$A = \frac{\left(\frac{B + W + T}{3.603}\right)}{-5.853W + (B \times \mathrm{FA})}, \tag{11}$$

$$B = \left(\frac{T}{(W + T)^{*}7.371}\right)^{*}(5.406 + F) + (-3.900 + B), \tag{12}$$

$$C = \frac{\left(\frac{(\mathrm{OPC} - W) - (12.609 - T)}{(8.644 + F) + F}\right)}{W}, \tag{13}$$

$$D = -1.663 \times F, \tag{14}$$

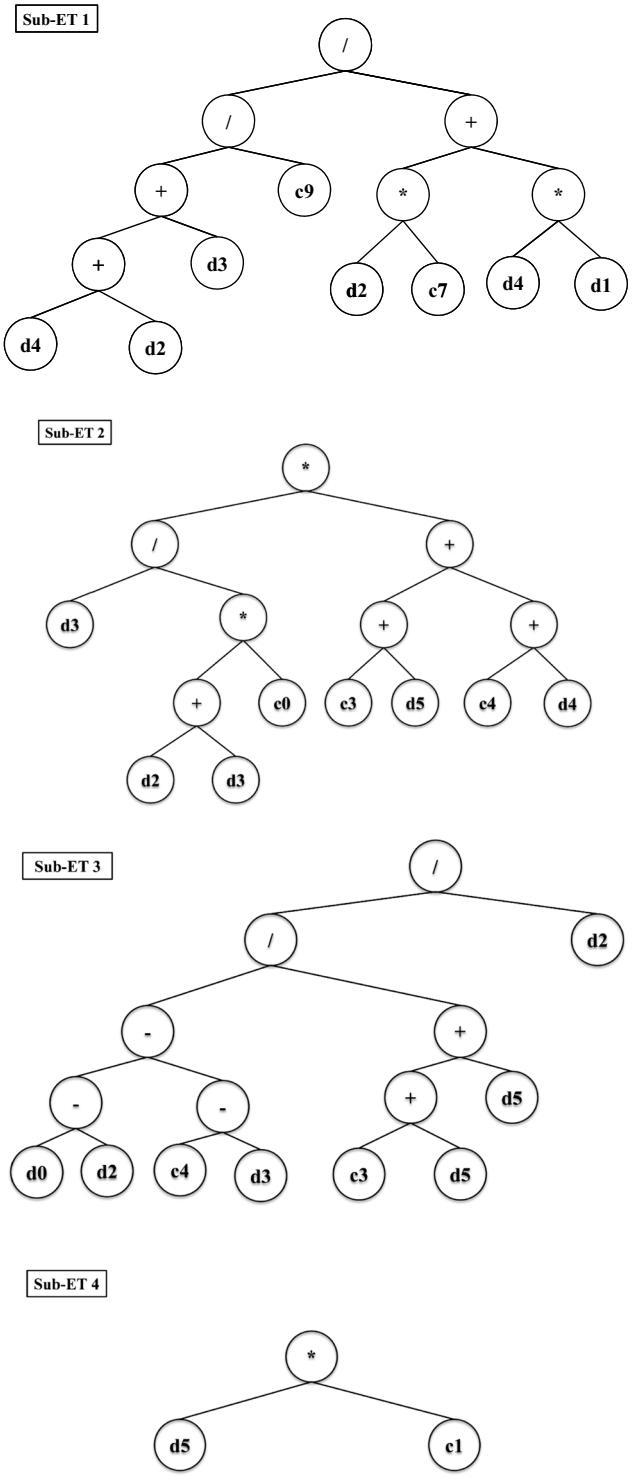
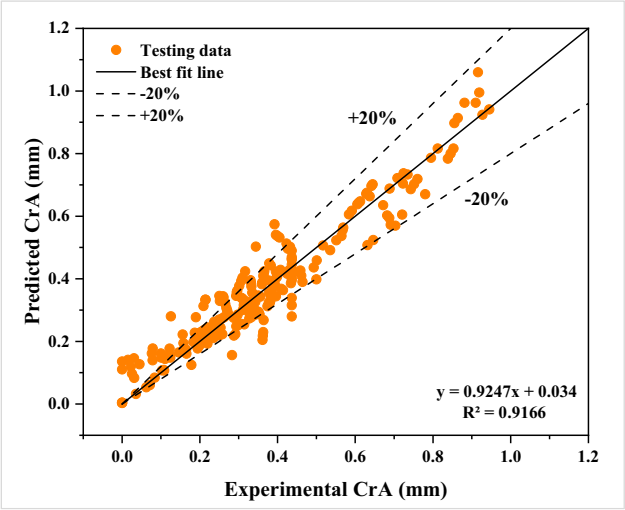
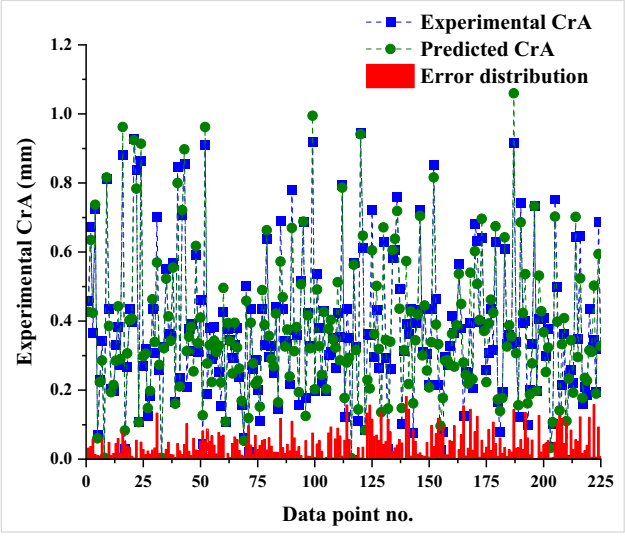


Figure 6: Expression tree diagram of the CrA-GEP model.

where OPC is the ordinary Portland cement, FA is the sand, W is the water, T is the time, B is the bacteria, F is the fiber, and CrA is the cracked area.



(a)



(b)

Figure 7: CrA-GEP approach: (a) relationship between projected and experimental CrA values and (b) dispersion of expected and experimental CrA values as well as errors.

3.2 CrA-MEP model

After analyzing the MEP results to account for the impact of the six independent components, an empirical formula was derived to determine the CrA of SHC. Eq. (15) displays the final set of mathematical equations that were modeled.

$$\mathrm{CrA}(mm) = \frac{W}{(B - W)(\sqrt{W}(B - W))}\left(2T - 2\frac{T}{\sqrt{W}} \times B + \mathrm{OPC} - \frac{FA}{\sqrt{W}}\right), \tag{15}$$

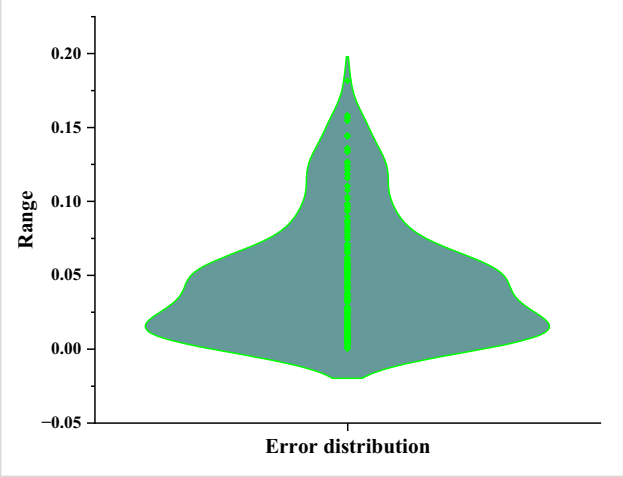
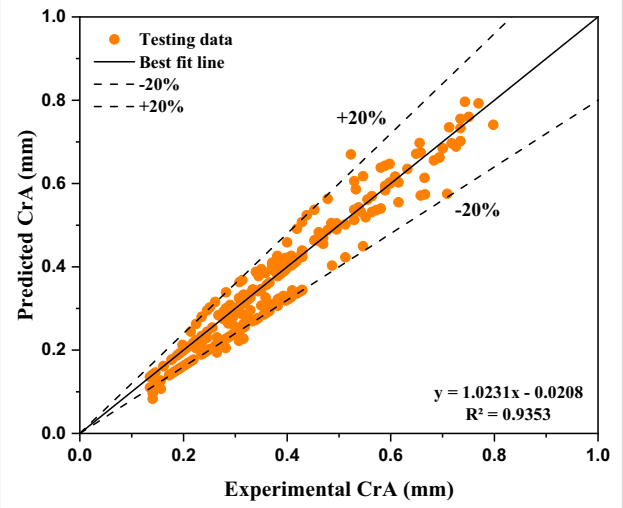
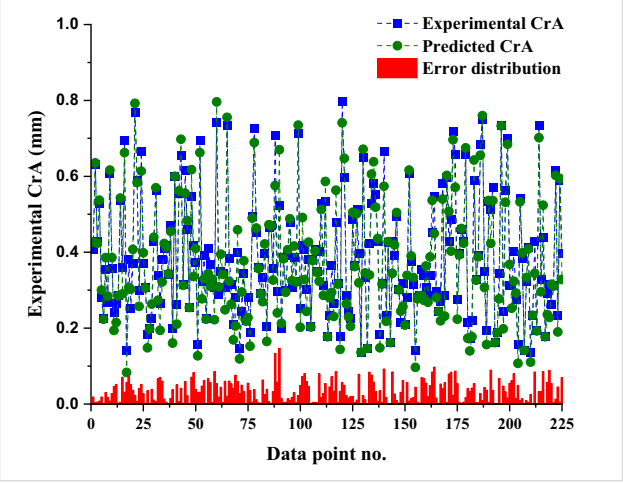


Figure 8: Violin plot for the GEP model's error distribution.



(a)



(b)

Figure 9: CrA-MEP approach: (a) relationship between projected and experimental CrA values; (b) dispersion of expected and experimental CrA values as well as errors.

where OPC is the ordinary Portland cement, FA is the sand, W is the water, T is the time, B is the bacteria, F is the fiber, and CrA is the cracked area.

As demonstrated in Figure 9(a) by an R^2 of 0.93, the MEP model is both well-trained and capable of handling oversimplification. Additionally, it exhibits satisfactory performance on novel, untested data. The CrA-MEP model outperforms the CrA-GEP model in terms of accuracy, as shown by its higher R^2 value. Whereas, in Figure 9(a), the solid black line represents a perfect fit to the data, and the dotted lines represent the percent deviation (20%) from the perfect fit. The MEP model's predictions for CrA were quite similar to the measured values. The CrA of SHC was effectively determined using the GEP method, with 95% of its predictions within the 20% threshold, suggesting exceptionally higher accuracy. Figure 9(b) displays the results of an analysis of absolute differences between the goal and observed values performed in MEP simulations. Rendering to the provided evidence, the MEP forecast brim of error was an average of 0.033 mm and varied between 0.00 and 0.147.60 mm. A total of 71 error values were less than 0.01 mm, 90 were between 0.01 and 0.05 mm, and 67 were greater than 0.05 mm, bringing the overall error values below 0.150 mm. The MEP model outperforms the GEP model in terms of extreme value prediction. As shown in the violin plot in Figure 10, the MEP model reduces both the correlation coefficient and the standard deviations of the errors. The MEP equation is often used since it is both generalizable and concise. In comparison to the GEP model, the MEP model seems to be better because of its greater correlation coefficient and lower error levels.

3.3 Validation of the models

Findings of the efficacy and error metrics (RRMSE, NSE, MAE, RSE, RMSE, and R) are presented in Table 3, which is based on the calculations done using the aforementioned Eqs (2)–(9). The lower the error values of the models, the more accurate the predictions. It was discovered that the MAE value for the CrA-GEP model was 0.047 mm, which significantly reduced to 0.033 mm, for the corresponding CrA-MEP model. Whereas, the 12.6% MAPE value for the CrA-GEP model considerably dropped to 9.60% in the analogous CrA-MEP model. Further, additional error-based statistical metrics, such as

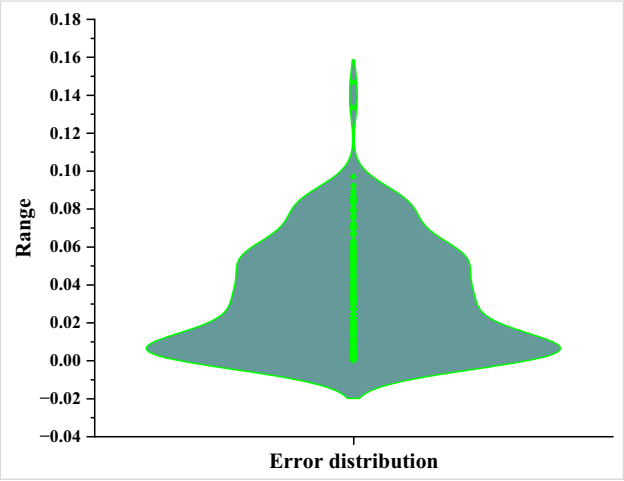


Figure 10: Violin plot for the MEP model's error distribution.

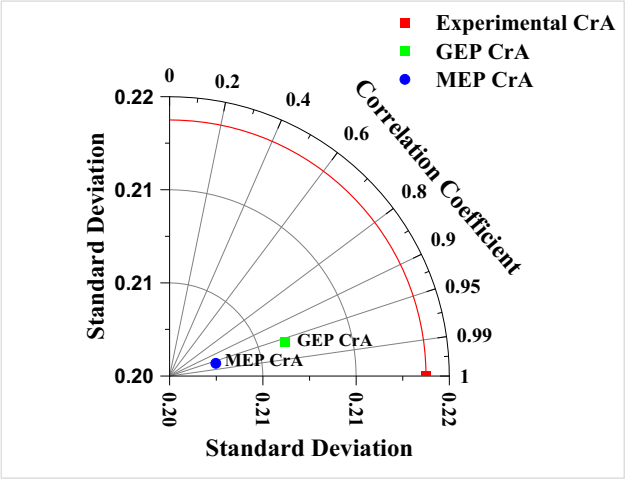


Figure 11: Taylor diagram for the models' validation.

RMSE, RSE, and RRMSE, showed a comparable pattern. Two measures, NSE and R, were used to assess the efficiency of the production alongside validation depending on errors. The accuracy of a model's predictions is directly proportional to its efficiency rating. In the CrA-GEP model, the NSE value was 0.916, while in the comparable CrA-MEP model, it increased to 0.921. When looking at the produced models with Pearson's coefficient (R), the results were similar. Moreover, the a20-index for the MEP model was 0.95 as compared to GEP's 0.83, validating the higher precision of the MEP model. The constructed forecasting models GEP and MEP are compared in Figure 11, which is a Taylor diagram. In the forecasting of CrA of SHC, the MEP model can be seen much closer to the experimental line as compared to the GEP model. So, as previously pointed out, the MEP approach has the best ML-grounded strategy for predicting the CrA of SHC because of its high efficiency, minimal standard deviation, low error, and high R^2.

3.4 SHAP analysis outcomes

The effect of different raw materials on the CrA of SHC was explored. From one dataset to another, the SHAP tree interpreter is employed to provide further details regarding the local SHAP explanations and the overall feature effects. Figure 12 shows the violin SHAP graph findings for all raw materials and how they affect the CrA of SHC. This graph uses different shades to represent the different variables, and the x-axis SHAP value represents the relative contribution of each raw ingredient. A negative association between input bacteria and output CrA of SHC was observed as per the SHAP analysis plot in Figure 12, evident by the more high-intensity red dots on the negative side of the plot as compared to the lesser low-intensity blue dots on the positive side of the plot. It clearly illustrates that increasing the bacteria after a certain limit will result in the decrease in CrA in the SHC. Figure 13 exhibits the interdependencies of bacteria and fiber. According to Figure 13(a), up to the bacteria value 1, the CrA increases, but it drops significantly afterward. The SHAP study confirms the same relationship between bacteria and CrA as previous research in the same field [92,93]. Additionally, the SHAP analysis plot in Figure 13(b) shows that the fiber and CrA of SHC had a stronger positive (direct) link, as indicated by the higher intensity of the red dots on the positive side of the plot compared to the lower intensity of the blue dots on the negative side. These data strongly imply that CrA values decrease with the increase in the fiber content at the beginning, but then rise with further enhancements. The SHAP analysis for the rest of the inputs is not provided as their impact on the CrA of SHC was not obvious since there was not enough input value variation in the used dataset. Remember that the ingredients and sample size used in this research might affect the outcomes [94,95]. A wide range of outcomes could be achieved by varying the input parameters and sample size.

Table 3: Results acquired by statistical examination

Property	CrA-GEP	CrA-MEP
MAE (mm)	0.047	0.033
MAPE (%)	12.6	9.60
RMSE (mm)	0.062	0.044
R	0.957	0.967
RSE (mm)	0.262	0.234
NSE	0.916	0.921
RRMSE (mm)	0.556	0.442
a20 index	0.830	0.950

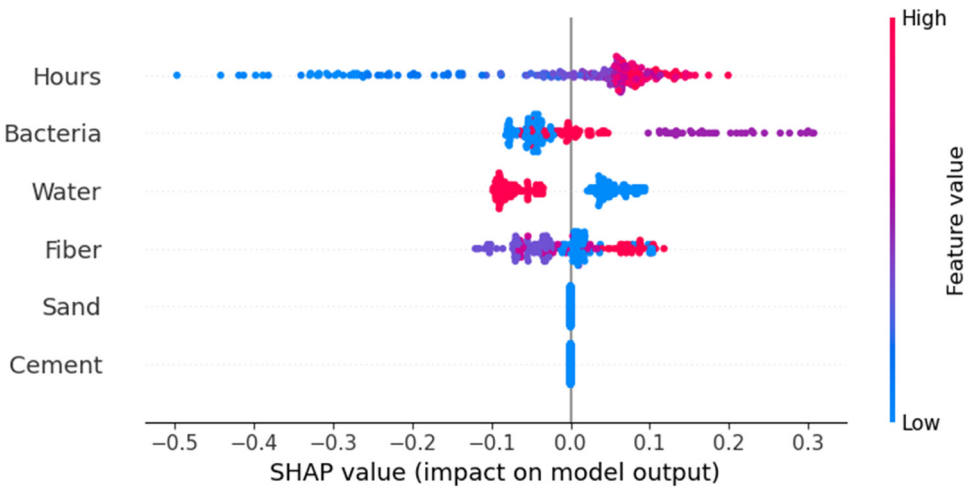


Figure 12: The significance and impact of input elements are suggested by the SHAP plot.

4 Discussions

The GEP and MEP models used in this work guarantee that the predictions will be exclusive to SHC because they can only take values from a narrow set of six input parameters. Since all of the models employ the same unit measurements and testing procedure, their CrA predictions are reliable. The models employ mathematical equations to gain a deeper understanding of the mix design and the impact of each input parameter. The projected models might not operate if the composite analysis contains more than six parameters. These models might not work as expected if the data used to train them are drastically different from what they are supposed to do. The models' accuracy in predicting results is dependent on how consistent or altered the units of the input parameters are. It is critical to maintain consistent unit sizes for the models to function. Predicting material strength, assuring quality, assessing risk, performing predictive maintenance, and improving energy efficiency are just a few of the many applications of ML models in the construction sector. Nevertheless, these models have a number of issues; for example, they rely on human input, employ erroneous data, and are not necessarily correct. To address these restrictions and improve ML-grounded solutions, future research could look into integrating IoT devices, creating hybrid models, using explainable AI techniques, considering sustainability, and customizing data generation and distribution for specific industries, among other things. Thanks to improvements in efficiency, interpretability, transparency, and informed decision-making, as well as increased levels of safety and fewer project delays, the construction sector stands to gain substantially from these technological breakthroughs. This study's results have the

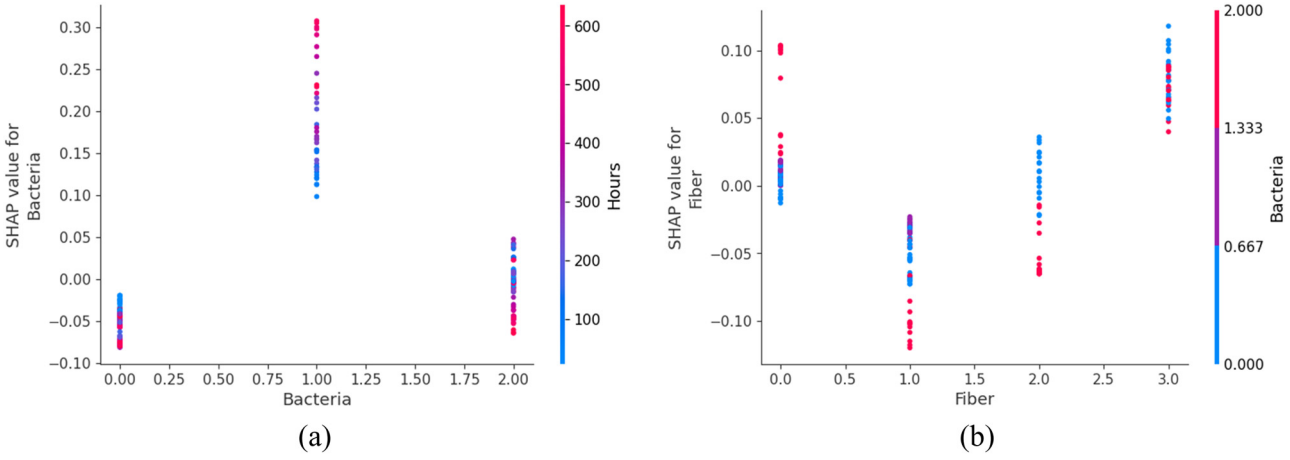


Figure 13: Input factors relationships for the CrA of SHC: (a) bacteria and (b) fiber.

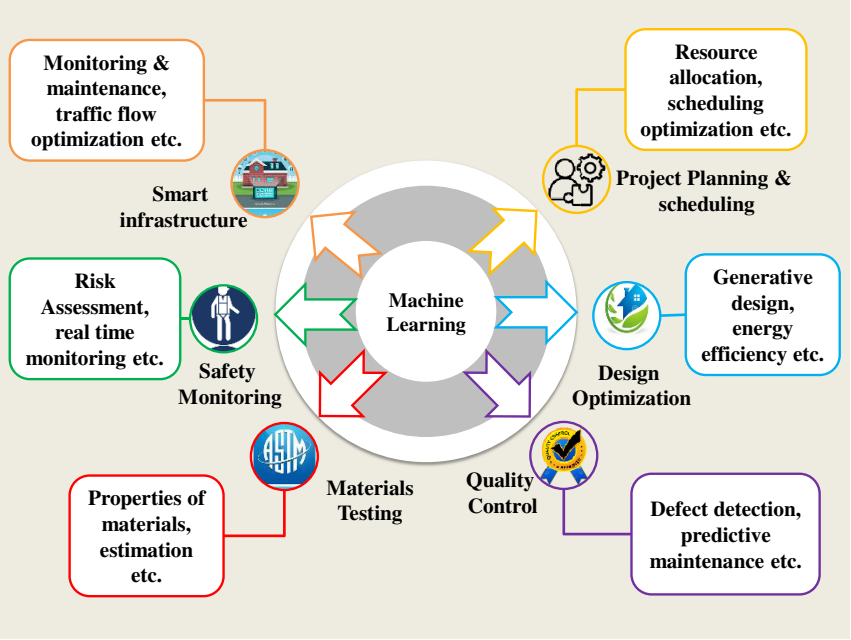


Figure 14: ML applications in civil engineering.

potential to encourage more sustainable building practices by increasing the use of SHC in the building sector. Figure 14 displays the applications of ML in the field of engineering.

5 Conclusion

The goal of this study is to employ MEP and GEP to analyze and predict the CrA of SHC. Training, testing, and validation of the developed models were carried out using 684 sets of SHC's CrA data acquired *via* laboratory tests. Here are the key findings of the study:

1) The MEP method provided more accurate results (R^2 = 0.93) for the CrA estimation of SHC, whereas the GEP approach was sufficiently accurate (R^2 = 0.91).
2) For both the GEP and MEP methods, the average discrepancy between experimental and projected CrA (errors) was 0.047 and 0.033 mm, respectively. These error rates further demonstrated the accuracy of the GEP model, while the MEP method provided a more precise prediction of SHC's CrA.
3) The models' efficacy has been validated statistically. The R^2 and error rates of ML models have been enhanced. The MAPE for the GEP model was 12.6%, whereas that for the MEP model was 9.60%. With an RMSE of 0.062 mm, the MEP model outperformed the GEP model by a narrow margin. Other areas of validating the model's performance were bolstered by these decisions.
4) SHAP analysis showed that the relation between bacteria and CrA of SHC is in-direct (negative), which means that increasing the bacteria (above value 1) would result in the decrease in CrA of SHC. Whereas, the relation of fiber and CrA as per the SHAP analysis was more direct.

The unique mathematical formula provided by GEP and MEP is what makes it so important for feature prediction in other databases. Scientists and engineers can use the mathematical models that come out of this study to quickly evaluate, improve, and rationalize the proportioning of SHC mixtures. Each of the 684 datasets had six parameters. Researchers may add new experiment data to improve models. This research used GEP and MEP models. However, hybrid ML methods like GA-PSO and RF-ANN, as well as individual/standalone and ensemble algorithms like SVM, DT, and boosting, could be studied further. These hybrid approaches may improve model functionality and prediction, making them realistic to adopt.

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, and writing-original draft. Y.L.: data acquisition, software, methodology, and writing – reviewing and editing. J.Z.: investigation, funding acquisition, supervision, and writing – reviewing and editing. S.S.: formal analysis, resources, methodology, and writing – reviewing and editing. L.Y.: formal analysis, visualization, and writing – reviewing and editing. T.C.: validation, investigation, supervision, and writing – reviewing, and editing. J. H.: conceptualization, supervision, project administration, and 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] He, X., M. N. Amin, K. Khan, W. Ahmad, F. Althoey, and N. I. Vatin. Self-healing concrete: A scientometric analysis-based review of the research development and scientific mapping. *Case Studies in Construction Materials*, Vol. 17, 2022, id. e01521.

[2] Wiktor, V. and H. M. Jonkers. Quantification of crack-healing in novel bacteria-based self-healing concrete. *Cement and Concrete Composites*, Vol. 33, 2011, pp. 763–770.

[3] 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.

[4] Yuan, X., Q. Cao, M. Nasir Amin, A. Ahmad, W. Ahmad, F. Althoey, et al. Predicting the crack width of the engineered cementitious materials via standard machine learning algorithms. *Journal of Materials Research and Technology*, Vol. 24, 2023, pp. 6187–6200.

[5] 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.

[6] Rasol, M. A., V. Pérez-Gracia, M. Solla, J. C. Pais, F. M. Fernandes, and C. Santos. An experimental and numerical approach to combine Ground Penetrating Radar and computational modeling for the identification of early cracking in cement concrete pavements. *Ndt & E International*, Vol. 115, 2020, id. 102293.

[7] Hossain, M. R., R. Sultana, M. M. Patwary, N. Khunga, P. Sharma, and S. J. Shaker. Self-healing concrete for sustainable buildings. A review. *Environmental Chemistry Letters*, Vol. 20, 2022, pp. 1265–1273.

[8] Karthiga, N. and R. Praveena. Performance of bacteria on self-healing concrete and its effects as carrier. *Materials Today: Proceedings*, Vol. 65, 2022, pp. 1987–1989.

[9] Luo, M., C.-X Qian, and R.-Y Li. Factors affecting crack repairing capacity of bacteria-based self-healing concrete. *Construction and building materials*, Vol. 87, 2015, pp. 1–7.

[10] Shah, K. W. and G. F. Huseien. Biomimetic self-healing cementitious construction materials for smart buildings. *Biomimetics*, Vol. 5, 2020, id. 47.

[11] Bayar, G. and T. Bilir. A novel study for the estimation of crack propagation in concrete using machine learning algorithms. *Construction and building materials*, Vol. 215, 2019, pp. 670–685.

[12] Nodehi, M., T. Ozbakkaloglu, and A. Gholampour. A systematic review of bacteria-based self-healing concrete: Biomineralization, mechanical, and durability properties. *Journal of Building Engineering*, Vol. 49, 2022, id. 104038.

[13] Feng, J., B. Chen, W. Sun, and Y. Wang. Microbial induced calcium carbonate precipitation study using Bacillus subtilis with application to self-healing concrete preparation and characterization. *Construction and Building Materials*, Vol. 280, 2021, id. 122460.

[14] Huang, X., J. Ge, S. Kaewunruen, and Q. Su. The self-sealing capacity of environmentally friendly, highly damped, fibre-reinforced concrete. *Materials*, Vol. 13, 2020, id. 298.

[15] Jamshidi, M., M. El-Badry, and N. Nourian. Improving concrete crack segmentation networks through CutMix data synthesis and temporal data fusion. *Sensors*, Vol. 23, 2023, id. 504.

[16] Su, Y., C. Qian, Y. Rui, and J. Feng. Exploring the coupled mechanism of fibers and bacteria on self-healing concrete from bacterial extracellular polymeric substances (EPS). *Cement and Concrete Composites*, Vol. 116, 2021, id. 103896.

[17] Zheng, T., C. Qian, and Y. Su. Influences of different calcium sources on the early age cracks of self-healing cementitious mortar. *Biochemical Engineering Journal*, Vol. 166, 2021, id. 107849.

[18] Congro, M., V. M. de Alencar Monteiro, A. L. T. Brandão, B. F. dos Santos, D. Roehl, and F. de Andrade Silva. Prediction of the residual flexural strength of fiber reinforced concrete using artificial neural networks. *Construction and Building Materials*, Vol. 303, 2021, id. 124502.

[19] Marani, A., A. Jamali, and M. L. Nehdi. Predicting ultra-high-performance concrete compressive strength using tabular generative adversarial networks. *Materials*, Vol. 13, 2020, id. 4757.

[20] 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.

[21] Himanen, L., A. Geurts, A. S. Foster, and P. Rinke. Data-driven materials science: status, challenges, and perspectives. *Advanced Science*, Vol. 6, 2019, id. 1900808.

[22] Pollice, R., G. dos Passos Gomes, M. Aldeghi, R. J. Hickman, M. Krenn, C. Lavigne, et al. Data-driven strategies for accelerated materials design. *Accounts of Chemical Research*, Vol. 54, 2021, pp. 849–860.

[23] Asteris, P. G., K. G. Kolovos, M. G. Douvika, and K. Roinos. Prediction of self-compacting concrete strength using artificial neural networks. *European Journal of Environmental and Civil Engineering*, Vol. 20, 2016, pp. s102–s122.

[24] Jaf, D. K. I., P. I. Abdulrahman, A. S. Mohammed, R. Kurda, S. M. A. Qaidi, and P. G. Asteris. Machine learning techniques and multi-scale models to evaluate the impact of silicon dioxide (SiO_2) and calcium oxide (CaO) in fly ash on the compressive strength of green concrete. *Construction and Building Materials*, Vol. 400, 2023, id. 132604.

[25] Juan, Y., Y. Dai, Y. Yang, and J. Zhang. Accelerating materials discovery using machine learning. *Journal of Materials Science & Technology*, Vol. 79, 2021, pp. 178–190.

[26] 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.
[27] Ji, Z., M. M. Zhou, Q. Wang, and J. D. 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. *CMES-Computer Modeling in Engineering & Sciences*, Vol. 139, No. 2, 2024, pp. 1557–1582.
[28] Asteris, P. G., P. B. Lourenço, P. C. Roussis, C. E. Adami, D. J. Armaghani, L. Cavaleri, et al. Revealing the nature of metakaolin-based concrete materials using artificial intelligence techniques. *Construction and Building Materials*, Vol. 322, 2022, id. 126500.
[29] Asteris, P. G., A. D. Skentou, A. Bardhan, P. Samui, and P. B. Lourenço. Soft computing techniques for the prediction of concrete compressive strength using non-destructive tests. *Construction and Building Materials*, Vol. 303, 2021, id. 124450.
[30] Abdalla, A. and A. S. Mohammed. Hybrid MARS-, MEP-, and ANN-based prediction for modeling the compressive strength of cement mortar with various sand size and clay mineral metakaolin content. *Archives of Civil and Mechanical Engineering*, Vol. 22, 2022, id. 194.
[31] Abdalla, A. and A. Salih. Implementation of multi-expression programming (MEP), artificial neural network (ANN), and M5P-tree to forecast the compression strength cement-based mortar modified by calcium hydroxide at different mix proportions and curing ages. *Innovative Infrastructure Solutions*, Vol. 7, 2022, id. 153.
[32] Koopialipoor, M., P. G. Asteris, A. S. Mohammed, D. E. Alexakis, A. Mamou, and D. J. Armaghani. Introducing stacking machine learning approaches for the prediction of rock deformation. *Transportation Geotechnics*, Vol. 34, 2022, id. 100756.
[33] Ahmed, H. U., A. S. Mohammed, R. H. Faraj, A. A. Abdalla, S. M. A. Qaidi, N. H. Sor, et al. Innovative modeling techniques including MEP, ANN and FQ to forecast the compressive strength of geo-polymer concrete modified with nanoparticles. *Neural Computing and Applications*, Vol. 35, 2023, pp. 12453–12479.
[34] 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.
[35] Zhou, J., X. Shen, Y. Qiu, X. Shi, and K. Du. Microseismic location in hardrock metal mines by machine learning models based on hyperparameter optimization using Bayesian optimizer. *Rock Mechanics and Rock Engineering*, Vol. 56, 2023, pp. 8771–8788.
[36] Zhou, J., S. Huang, and Y. Qiu. Optimization of random forest through the use of MVO, GWO and MFO in evaluating the stability of underground entry-type excavations. *Tunnelling and Underground Space Technology*, Vol. 124, 2022, id. 104494.
[37] Mahjoubi, S., R. Barhemat, W. Meng, and Y. Bao. AI-guided auto-discovery of low-carbon cost-effective ultra-high performance concrete (UHPC). *Resources, Conservation and Recycling*, Vol. 189, 2023, id. 106741.
[38] Kaveh, A., A. Dadras Eslamlou, S. M. Javadi, and N. Geran Malek. Machine learning regression approaches for predicting the ultimate buckling load of variable-stiffness composite cylinders. *Acta Mechanica*, Vol. 232, 2021, pp. 921–931.
[39] Huang, J. D., Y. T. Sun, and J. F. Zhang. Reduction of computational error by optimizing SVR kernel coefficients to simulate concrete compressive strength through the use of a human learning optimization algorithm. *Engineering with Computers*, Vol. 38, 2022, pp. 3151–3168.
[40] Huang, X., J. Sresakoolchai, X. Qin, Y. F. Ho, and S. Kaewunruen. Self-healing performance assessment of bacterial-based concrete using machine learning approaches. *Materials*, Vol. 15, 2022, id. 4436.
[41] Zhuang, X. and S. Zhou. The prediction of self-healing capacity of bacteria-based concrete using machine learning approaches. *Computers, Materials & Continua*, Vol. 59, 2019, pp. 57–77.
[42] Zhou, J., Z. L. Su, S. Hosseini, Q. Tian, Y. J. 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.
[43] Huang, J. D., P. G. Asteris, S. M. K. Pasha, A. S. Mohammed, and M. Hasanipanah. A new auto-tuning model for predicting the rock fragmentation: a cat swarm optimization algorithm. *Engineering with Computers*, Vol. 38, 2022, pp. 2209–2220.
[44] Huang, J. D., M. Koopialipoor, and D. J. Armaghani. A combination of fuzzy delphi method and hybrid ANN-based systems to forecast ground vibration resulting from blasting. *Scientific Reports*, Vol. 10, 2020, id. 19397.
[45] Huang, J. D., G. S. Kumar, and Y. T. Sun. Evaluation of workability and mechanical properties of asphalt binder and mixture modified with waste toner. *Construction and Building Materials*, Vol. 276, 2021, id. 122230.
[46] Huang, J. D., G. S. Kumar, J. L. Ren, Y. T. Sun, Y. J. Li, and C. G. Wang. Towards the potential usage of eggshell powder as bio-modifier for asphalt binder and mixture: workability and mechanical properties. *International Journal of Pavement Engineering*, Vol. 23, 2022, pp. 3553–3565.
[47] Ma, H. X., J. D. Liu, J. Zhang, and J. D. Huang. Estimating the compressive strength of cement-based materials with mining waste using support vector machine, decision tree, and random forest models. *Advances in Civil Engineering*, Vol. 2021, 2021.
[48] Lee, B. C. and D. M. Brooks. Accurate and efficient regression modeling for microarchitectural performance and power prediction. *ACM Sigops operating systems review*, Vol. 40, 2006, pp. 185–194.
[49] 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, No. 1, 2024, id. 225.
[50] 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.
[51] Zhang, H., Q. Chang, S. Li, and J. D. Huang. Determining the efficiency of the sponge city construction pilots in China based on the DEA-malmquist model. *International Journal of Environmental Research and Public Health*, Vol. 19, 2022, id. 11195.
[52] Huang, J. D., T. H. Duan, Y. Zhang, J. D. Liu, J. Zhang, and Y. W. Lei. Predicting the permeability of pervious concrete based on the beetle antennae search algorithm and random forest model. *Advances in Civil Engineering*, Vol. 2020, 2020, id. 8863181.
[53] Cui, K., J. Chang, M. M. S. Sabri, and J. D. Huang. Toughness, reinforcing mechanism, and durability of hybrid steel fiber reinforced sulfoaluminate cement composites. *Buildings*, Vol. 12, No. 8, 2022, id. 1243.
[54] Tian, Q., Z. L. Su, N. Fiorentini, J. Zhou, H. Luo, Y. J. Lu, et al. Ensemble learning models to predict the compressive strength of geopolymer concrete: a comparative study for geopolymer

composition design. *Multiscale and Multidisciplinary Modeling Experiments and Design*, Vol. 6, 2023.
[55] Wang, Q. A., J. Zhang, and J. D. Huang. Simulation of the compressive strength of cemented tailing backfill through the use of firefly algorithm and random forest model. *Shock and Vibration*, Vol. 2021, 2021, id. 5536998.
[56] Holland, J. H. *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*, MIT Press, 1992.
[57] Koza, J. On the programming of computers by means of natural selection. *Genetic programming*, MIT Press, Cambridge, USA, 1992.
[58] Gholampour, A., T. Ozbakkaloglu, and R. Hassanli. Behavior of rubberized concrete under active confinement. *Construction and Building Materials*, Vol. 138, 2017, pp. 372–382.
[59] Topcu, I. B. and M. Sarıdemir. 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.
[60] 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, 2022, id. 5938.
[61] Huang, J., G. S. Kumar, J. Ren, J. Zhang, and Y. Sun. Accurately predicting dynamic modulus of asphalt mixtures in low-temperature regions using hybrid artificial intelligence model. *Construction and Building Materials*, Vol. 297, 2022, id. 123655.
[62] Ferreira, C. *Gene expression programming: mathematical modeling by an artificial intelligence*, Vol. 21, Springer, Heidelberg, Germany, 2006.
[63] 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, No. 10, 2022.
[64] 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.
[65] 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.
[66] 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.
[67] 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.
[68] 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.
[69] 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.
[70] Oltean, M. and C. Grosan. A comparison of several linear genetic programming techniques. *Complex Systems*, Vol. 28, 2020, pp. 499–504.
[71] Huang, J. D., T. H. Duan, Y. W. Lei, and M. Hasanipanah. Finite element modeling for the antivibration pavement used to improve the slope stability of the open-pit mine. *Shock and Vibration*, Vol. 28, 2020, pp. 499–504.
[72] 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.
[73] Huang, J. D., M. M.S. Sabri, D. V. Ulrikh, M. Ahmad, and K. A.M. 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.
[74] Huang, J. D., P. Leandri, G. Cuciniello, and M. Losa. Mix design and laboratory characterisation of rubberised mixture used as damping layer in pavements. *International Journal of Pavement Engineering*, Vol. 23, 2022, pp. 2746–2760.
[75] 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.
[76] 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.
[77] Grosan, C., A. Abraham. Stock market modeling using genetic programming ensembles. *In Genetic Systems Programming: Theory and Experiences*, Springer, Heidelberg, Germany, 2006, pp. 131–146.
[78] Oltean, M. and D. Dumitrescu. Multi expression programming. *Journal of Genetic Programming and Evolvable Machines*, 2002.
[79] 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.
[80] Shahin, M. A. *Genetic Programming for modelling of geotechnical engineering systems*, Springer, Cham, Germany, 2015.
[81] 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.
[82] 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.
[83] 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.
[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 Al_2O_3/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] 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.

[90] 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.

[91] Taylor, K. E. Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, Vol. 106, 2001, pp. 7183–7192.

[92] Jonkers, H. M. Bacteria-based self-healing concrete. *In-Genium*, 2021.

[93] Vijay, K., M. Murmu, and S. V. Deo. Bacteria based self healing concrete–A review. *Construction and building materials*, Vol. 152, 2017, pp. 1008–1014.

[94] 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.

[95] Zhu, F., X. P. Wu, M. M. Zhou, M. M. S. Sabri, and J. D. Huang. Intelligent design of building materials: development of an AI-based method for cement-slag concrete design. *Materials*, Vol. 15, 2022, id. 3833.