

Research Article

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

Supplementary cementitious materials-based concrete porosity estimation using modeling approaches: A comparative study of GEP and MEP

<https://doi.org/10.1515/rams-2023-0189>

received January 17, 2024; accepted February 16, 2024

Abstract: Using supplementary cementitious materials in concrete production makes it eco-friendly by decreasing cement usage and the corresponding CO₂ emissions. One key measure of concrete's durability performance is its porosity. An empirical prediction of the porosity of high-performance concrete with added cementitious elements is the goal of this work, which employs machine learning approaches. Binder, water/cement ratio, slag, aggregate content, superplasticizer (SP), fly ash, and curing conditions were considered as inputs in the database. The aim of this study is to create ML models that could evaluate concrete porosity. Gene expression programming (GEP) and multi-expression programming (MEP) were used to develop these models. Statistical tests, Taylor's diagram, R^2 values, and the difference between experimental and predicted readings were the metrics used to evaluate the models. With $R^2 = 0.971$, mean absolute error (MAE) = 0.348%, root mean square error (RMSE) = 0.460%, and Nash–Sutcliffe efficiency (NSE) = 0.971, the MEP provided a slightly better-fitted model and improved prediction performance when contrasted with the GEP, which had $R^2 = 0.925$, MAE = 0.591%, RMSE = 0.745%, and NSE = 0.923. Binder, water/binder ratio, curing conditions, and aggregate

content had a direct (positive) relationship with the porosity of concrete, while SP, fly ash, and slag had an indirect (negative) association, according to the SHapley Additive exPlanations study.

Keywords: porosity, concrete, prediction models

1 Introduction

In concrete, aggregates and hydrated cement paste come together to produce a composite material with the potential to form any shape. Any air that is not released during the process of concrete setting remains trapped inside the concrete matrix [1,2]. Furthermore, the occurrence of microscopic air pockets inside the concrete mass is a consequence of the delayed interaction between cement and water, as well as the evaporation of water during the setting process that was initially filled with water [3,4]. Both factors elucidate the reason behind the porous nature of concrete despite its durability in its solidified condition [5]. There is a lot of curiosity about how concrete's porosity affects its transport characteristics, including its electrical resistivity, migration of chloride ions, gas/water permeability, and carbon dioxide and oxygen diffusion coefficients [6]. Figure 1 shows the empirical link between the capillary porosity and the compressive strength and transport qualities (intrinsic permeability) of concrete. In general, the mechanical strength and permeability of concrete are both negatively affected by the increase in porosity. Hence, porosity is an important factor in predicting how long concrete constructions exposed to harsh environments will last and how well they will function [7,8].

Several factors influence the porosity of concrete, but one of the most important is the water-to-cement ratio (w/c), which helps the cement paste hydrate. A higher w/c results in hydrated cement paste with larger capillary pore volumes. Since fewer large-dimension pores are filled or joined by the pores of the calcium-silicate-hydrate (C–S–H) gel, as the curing period and hydration continue,

* **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

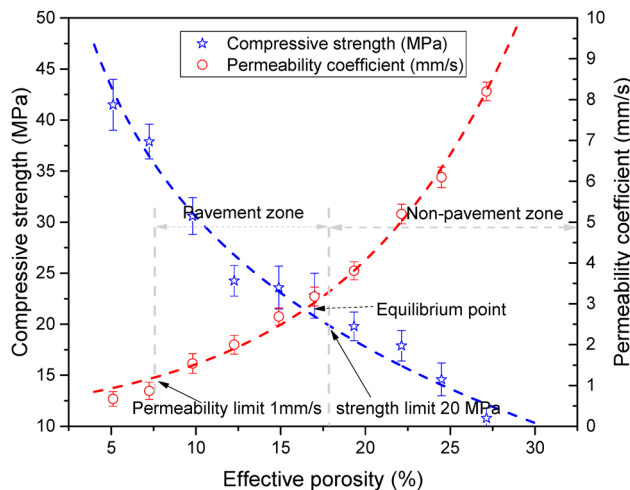


Figure 1: Variation in permeability coefficient and effective porosity of pervious concrete at the point of compressive strength equilibrium [7].

the porosity decreases. Centered on the cement's w/c ratio and hydration level, Powers developed a traditional model to determine the paste's volumetric composition when it has cured [9,10]. The interfacial transition zone (ITZ) aggregates and strengthens concrete's variability. Previous studies have established that when the size and fraction of coarse elements in concrete are increased, it clues to an upsurge in the permeability of the ITZ and a loss in the complete durability qualities of the material [11]. The CaCO_3 to fine aggregate weight ratio (CA/FA) is a key variable that influences concrete's porosity, permeability, and coarse-to-fine aggregate ratio [10,12,13].

Some studies have shown that supplemental cementitious materials (SCMs) can improve concrete strength and durability by partially replacing Portland cement [14,15]. Some SCMs that can aid in cleaner fabrication by plummeting CO_2 releases are fly ash, which is a byproduct of pig iron manufacturing, and ground granulated blast-furnace slag (GGBS), which is a consequence of coal powder combustion in thermoelectric power plants [16]. When added to concrete, GGBS causes a latent hydraulic reaction that mends the pore configuration and makes the matrix denser, both of which aid the cement's hydration process. Over time, this causes concrete's porosity to decrease and its compressive strength to increase. Cement hydrate mineralogy alterations in concrete could lead to increased chloride binding capabilities and elevated electrical resistivity [17,18]. Fly ash concrete is believed to have reduced permeability due to a combination of factors, including lower water quantity for particular workability and a better pore configuration brought about by the pozzolanic process. The pozzolanic reaction has long-term effects, which become increasingly noticeable in properly cured concrete as a result. Thus,

another important aspect that affects the porosity of high-performance concrete (HPC) is the curing state, which can be either air or water [19]. In addition, superplasticizer (SP) can help generate a denser pore structure while significantly reducing the quantity of mixing water needed [20,21].

An analytical model of the formation of the complex pore structure inside concrete can be hard due to differences in concrete composition and the cement hydration process, which is reliant on time [22]. One big problem is how little is known about the hydraulic and pozzolanic reactivities of slag and fly ash. Because SCMs can vary substantially in chemical composition, it is very difficult to predict what proportion of materials will be reactive. To forecast the chemical makeup and volume of fly ash concrete, Papadakis put forth a theoretical model [23,24]. All sorts of details are included in the model, including the reactant and product molar weights as well as the stoichiometry of the pozzolanic processes involving the hydration of Portland cement and fly ash. Unfortunately, the model does not take into account how porosity changes over time since it presumes that Portland cement and fly ash have fully hydrated and that all pozzolanic processes have been completed. To foretell how the degree of cement hydration will change over time, several researchers have put forward statistical models [25]. One way to monitor HPC's chloride diffusion coefficient and porosity is with the use of the empirical cement hydration model [26]. Using variables including fly ash percentage, micro silica concentration, and water/binder ratio (w/b), Khan [27] calculated the porosity of a concrete mixture at 28, 90, and 180 days using multivariate regression. Because of this, he was able to anticipate the permeability characteristics of HPC by experimentation.

Accurate predictions of HPC's mechanical qualities and durability have been made using statistical machine learning (ML) [28,29]. A recent example of ML's use in the concrete industry is the modeling of green concrete's mechanical and durability properties [30–33]. There is a strong agreement between the projected values and the concrete parameters measured experimentally, indicating that ML is a viable method for simulating concrete with complicated mixture compositions. The use of ML to forecast concrete porosity, however, has received scant attention [10,34,35]. To predict the porosity, transport tortuosity, and compressive strength of fly ash concrete, Boukhatem *et al.* developed a neural network modeling framework that takes into account the mix design components (aggregates, binder, SP, and water), fly ash content, and age as inputs [36]. Many ML-based techniques are being employed by engineers and researchers, including support vector machines, multi-expression programming (MEP), decision trees, random forests, and gene expression programming

(GEP) [37–39]. Furthermore, these methods can produce accurate results. Additionally, civil engineers have trained and utilized ML algorithms to create novel ultra-HPC, utilize artificial neural networks to forecast the strength of fiber-reinforced polymers, and ascertain the ultimate buckling stress of composite cylinders with varying stiffness [40–42]. Thus, HPC can likewise benefit from ML modeling in its design and development processes.

The aim of this study is to experimentally predict the porosity (P) of HPC incorporating SCMs by applying MEP and GEP. The first step was to compile a credible database of 240 records detailing concrete porosity data from existing literature. The effectiveness of ML algorithms was assessed using a variety of measures, 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 [43,44]. Datasets are essential for ML approaches; they can be created through exploratory experiments or mined from preexisting databases. By examining this information, ML algorithms may gain a more accurate understanding of the material's qualities. Experimental data and seven input criteria were used to evaluate the potential of ML approaches to anticipate the P of HPC. Using SHapley Additive exPlanations (SHAP) analysis, the importance of the raw ingredients was further explored. The newly collected features and developed ML models could be used to improve the existing SCM database or perhaps guide the design of HPC mixes. To evaluate the data-driven method against the traditional chemo-mechanical model for porosity prediction in concrete, the suggested case study might be used.

2 Methods of research

2.1 Collecting and analyzing data

Predicting the porosity (P) of the HPC was the goal of this research, which employed MEP and GEP simulations in

order to examine 240 data points collected from experiment research [45–50]. This study predicted the P centered on seven input factors such as GGBS, water-binder ratio (WBR), binder (B), fly ash (FA), SP, aggregate ratio (AR), and curing days (CD). The process of gathering and organizing the data relied heavily on data preparation. Preparing information for the purpose of data mining is a common tactic in the popular knowledge discovery from data technique that helps to circumvent a major hurdle. By removing unnecessary features and noise, data preparation aims to make the data easier to understand and work with. Several descriptive statistics were computed using these data, and their findings are shown in Table 1. The table includes the mean value, median, mode, standard deviation, and range of input and output parameters used in the database. It also contains the minimum and maximum values of all the parameters used for modeling. The validity of the used models has also been evaluated for their effectiveness [51]. The histograms in Figure 2(a)–(h) display the frequency of distinct parameters. A dataset's total frequency distribution can be described by integrating its component distributions. One way to see how common certain values are in a dataset is to look at their relative frequency distributions.

2.2 ML modeling

Through laboratory testing, the porosity (P) of HPC was determined. P , the final product, has seven inputs that went into its manufacturing. The use of advanced ML techniques such as GEP and MEP allowed for the prediction of HPC's P . When applied to input data, ML algorithms often yield evaluation results. ML model training consumed 70% of the total data, whereas testing was carried out on 30% of the data. Similar percentages of data split were identified in the literature [52,53]. This model worked as expected because the expected outcome had a high R^2 score. Disparity is high when R^2 is small [54–56], whereas a big number suggests that the predicted and actual outcomes

Table 1: Numerical interpretations of a set of variables

[illegible]

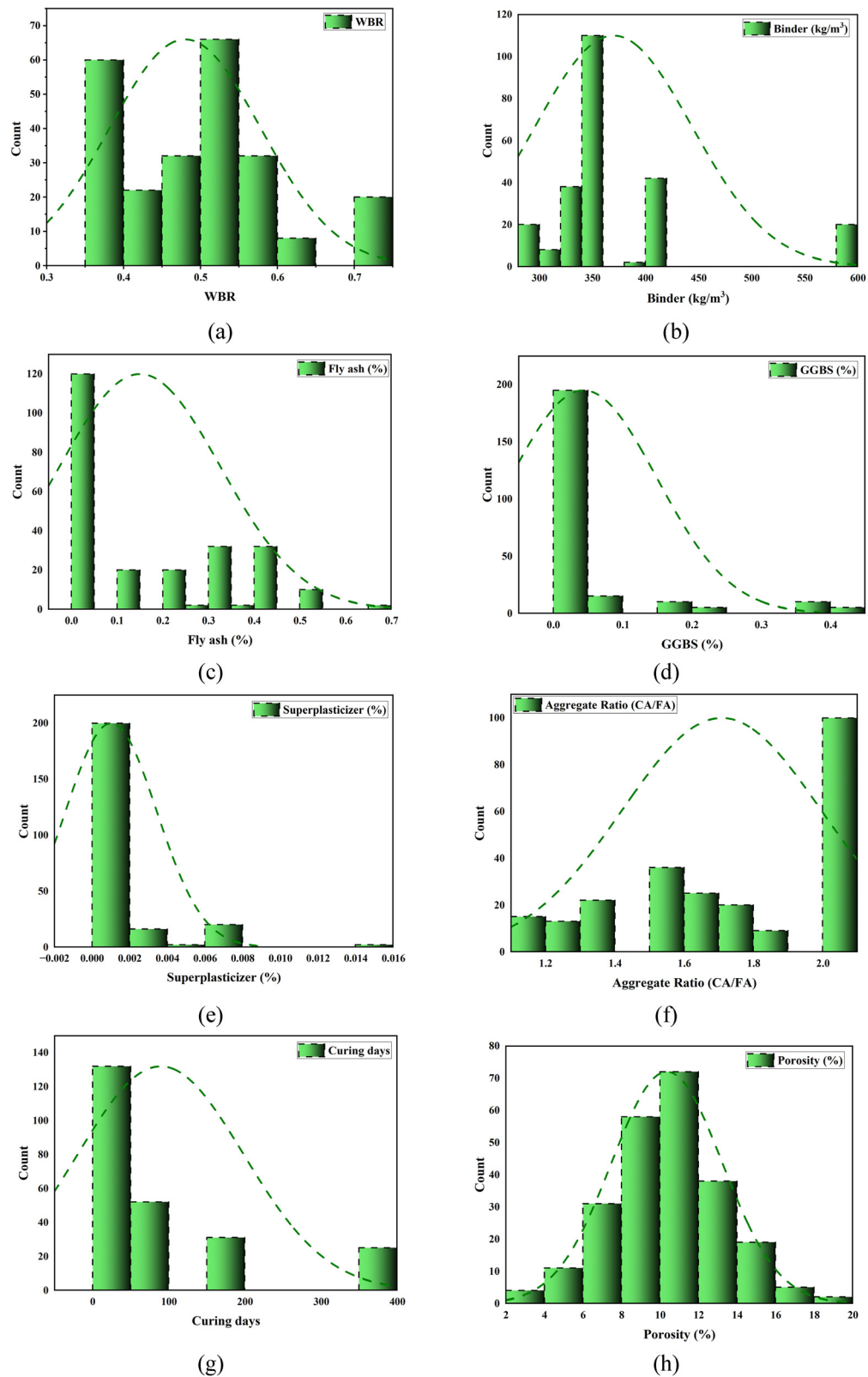


Figure 2: Distribution of database input features by frequency: (a) WBR, (b) binder, (c) fly ash, (d) GGBS, (e) SP, (f) AR, (g) CDs, and (h) Porosity.

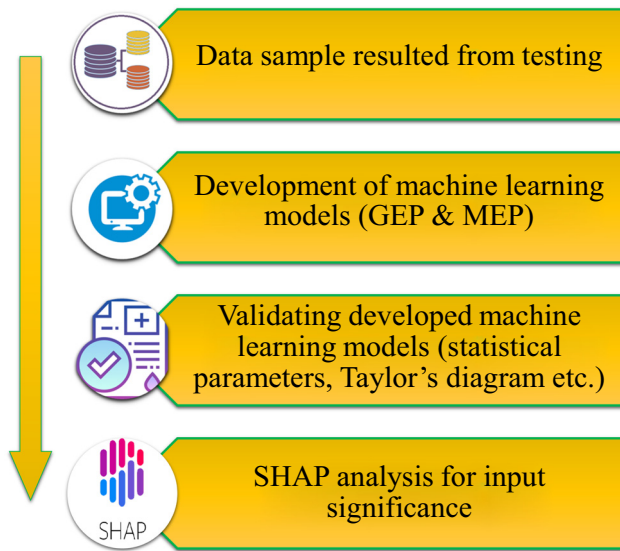


Figure 3: An overview of the ML methodology.

are very congruent. Statistical tests and evaluations of mistakes were among the methods used to verify the technique's accuracy. A streamlined representation of the event model is displayed in Figure 3. Table 2 displays the hyper-parameter parameters used by the MEP model.

2.2.1 GEP model

J. H. Holland created the genetic algorithm (GA) [43], out of which Darwin's theory of evolution emerged. The genomic progression is considered complete when a succession of GAs is observed, and continuously longer chromosomes serve as a marker for this completion. This innovative genetic algorithm technique is what Koza calls "genetic programming" [57]. GAs are employed in generalized problem-solving (GP) to construct an evolutionary framework [58]. Due to its ability to utilize nonlinear structures like parse trees, GP is highly adaptable and can be used instead of binary strings of constant length. In line with Darwin's theory, proven AI algorithms tackle issues connected to facsimiles by leveraging genetic components that occur naturally, such as procreation, overlap, and modification [59–61]. In GP, a plan is made to eliminate wasteful programs from the next iteration. Replanting the area using the chosen approach entails cutting down the unwanted trees, precisely as in the earlier illustration. Conversely, evolution safeguards early convergence [59,62]. Prior to implementing the GP approach, five critical parameters must be established, priority activities in the field, suitability valuation, principal serviceable operators (including cross-over and populace extent), and conclusions produced by

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

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

technique-precise endpoints [59]. A crossover genetic processor is responsible for most of the parse tree development, even if GP's model construction is repeated [47]. Nonlinear GP forms make it more difficult for desirable qualities to develop because they are both phenotype and genotype [62].

GEP is an alternative to the GP that Ferreira first suggested [62]. The GEP model incorporates static-length aligned genomes into parse trees, following the population-generation hypothesis. GEP is an upgraded variant of GP that uses basic, fixed-length chromosomes to encrypt software of moderate size. One advantage of GEP is that it may be used to create equations that can reliably predict outcomes for complicated and nonlinear issues [63,64]. 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 really identified as such before production by means of the “Karva” dialectal. A line of constant length is necessary for GEP to function. The data processing code in GP, on the other hand, shows parse trees of different lengths. Each cord is characterized as a genome with a fixed length, and its chromosomes are depicted by nonlinear manifestation/parse trees with varying sizes of pronged morphologies [59]. The genetic code of these genotypes and a small number of additional phenol strains are distinct from one another [62]. The GEP ensures that the genome is preserved from one generation to another, eliminating the need for costly structural mutations or duplications.

In a typical chromosome, the “head” and the “tail” are the two complementary regions. Amazingly, multi-gene entities evolved from a single genome [59]. These genes include the instructions for basic algebra, logic, and arithmetic as well as Boolean calculations. A hereditary code operator assigns a cell a specific function. One recently discovered language, Karva, can infer the contents of these chromosomes, which allows for the formulation of empirical formulas. A change in power takes place after the ET, and travel in Karva starts. By referring to Eq. (1), ET determines where to place the nodes in the layer below [63]. The degree and duration of GEP gene K-expression may be determined by the total number of ETs.

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

As an advanced ML algorithm, the results of GEP are independent of previous associations. A GEP mathematical equation goes through a number of steps, as seen in Figure 4. At birth, a person's chromosome number remains the same. After confirming that these chromosomes are ETs,

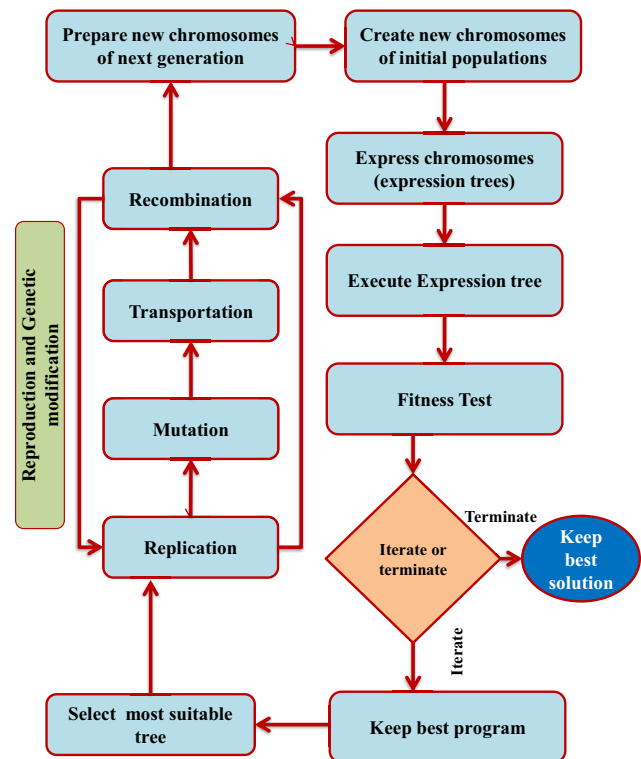


Figure 4: An approach to the GEP technique schematic diagram [65].

comprehensive health examinations can be carried out. The strongest and healthiest persons are given priority when it comes to having children. Using the most capable people in an iterative process leads to the best outcome. Mutation, cross-over, and breeding are the three generations of genetic processes that culminate in the final numerical manifestation.

2.2.2 MEP model

The MEP is a cutting-edge, exemplary linear-based GP approach since it utilizes linear chromosomes. There is a remarkable degree of software similarity between the MEP and the GEP. Unlike its predecessor, the GP technique, MEP may encode multiple software components (alternatives) into a single chromosome. You can achieve your goal by using fitness analysis to select the best chromosome [66,67]. According to Oltean and Grosan, this phenomenon occurs when a bipolar scheme recombines to create two different generations [68]. The process will continue running until either the termination form is satisfied or the best program is discovered, as shown in Figure 5. Mutations that impact infants are listed below. A number of components can be combined using the MEP model, just as in the GEP paradigm. The length of the algorithm or code, the number of

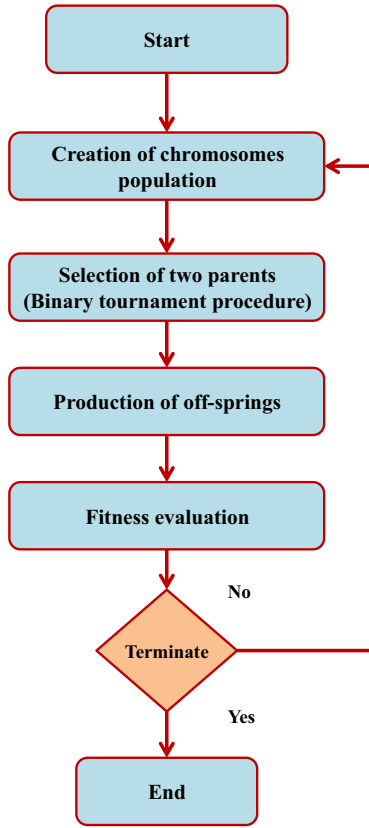


Figure 5: Method flowchart for the MEP procedure [65].

subpopulations, the number of functions, and the potential of crossover are all essential criteria in MEP [51,69]. Assessment gets more complex and time-consuming when the population size is equal to the total number of packages. Code length is another critical component that affects the size of the produced mathematical expressions. According to Table 2, a complete set of MEP parameters is required for an accurate representation of rheological properties.

Literature datasets are heavily used in both approaches' assessment and modeling stages [70,71]. When it comes to predicting the properties of practical concrete, some experts think that popular linear GP methods like the MEP and the GEP are better. The optimal neural network-based approach was found by Grosan and Abraham to be linguistic programming combined with maximum likelihood estimation [72]. The GEP's operation approach is slightly more complex than the MEP's [69]. Regardless that GEP has a higher density than MEP [73], 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. It is commonly believed that the GEP chromosome has better competency due to the codes located at its "head" and "tail" that facilitate the writing of syntactically

accurate software programs [68]. This necessitates a more thorough evaluation of each of these genetic approaches to engineering challenges.

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 [71,74–77]: mean absolute percentage error (MAPE), Pearson's correlation coefficient (R), mean absolute error (MAE), relative squared error (RSE), root mean square error (RMSE), Nash–Sutcliffe efficiency (NSE), and relative root mean square error (RRMSE). All of these statistical measures have their formulations in Eqs. (2)–(8).

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

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

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

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

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

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

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

where n is the total number of data points, a_i and p_i are the i th actual and predicted values, respectively; $\bar{a}_i$ also represents the average actual and predicted values. The relationship coefficient, abbreviated as R , is a common way to measure a model's projection power (a_i and p_i). A high value of R indicates a robust relationship between the predicted and actual output amounts [43,78,79]. Component R 's value is independent of both divisibility and multiplication. When comparing the actual and expected results, R^2 was computed since it offers a more accurate representation of the true value. A more effective method for constructing models is indicated by R^2 values that are closer to 1 [80,81]. 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 [82,83]. However, upon closer inspection, it became apparent that continuous and smooth databases are where MAE truly excels [84]. When the values of the errors computed above are smaller, the model often performs better.

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, the plots show their divergence from the truth, which serves as the reference point [85,86]. Model placement is indicated by three components: standard deviation (x- and y-axes), connection coefficient (outspread outlines), and RSME (circles focused on the true value point). A reliable model is one that consistently produces high-quality forecasts [18,85,87].

3 Results and discussion

3.1 P-GEP model

The GEP technique yielded ETs-based models that calculated porosity (P) by deducing mathematical correlations based on chromosomal number and head size, as shown in Figure 6. The numerical maneuvers ($\div$, $\times$, $-$, $+$, and square root) are used to construct most of the sub-ETs in the HPC's P . What comes out of encrypting the GEP method's sub-ETs is an arithmetical formula. Estimating the future P of HPC is possible using the input data and the yield value of these equalities Eqs. (9)–(13). With sufficient data, the produced model surpasses an ideal model operating under perfect circumstances. The solid blue line in Figure 7(a) depicts a perfect fit to the data, and the dotted lines reflect the percent deviation (20%) from the perfect match. This graphically shows the agreement between the experimental and projected P . P values predicted by the GEP model and those measured were very close. The GEP technique was quite effective in determining the HPC P ; it had an R^2 of 0.925 and predicted 97% of the time inside the 20% threshold, indicating significantly improved accuracy. Figure 7(b) shows the absolute error plotted against experimental results to show how far the GEP model could be from reality. The results showed that the GEP equation's predictions are quite close to the experimental data, with an MAE of 0.590% and a range of 0.00–1.564%. The error values

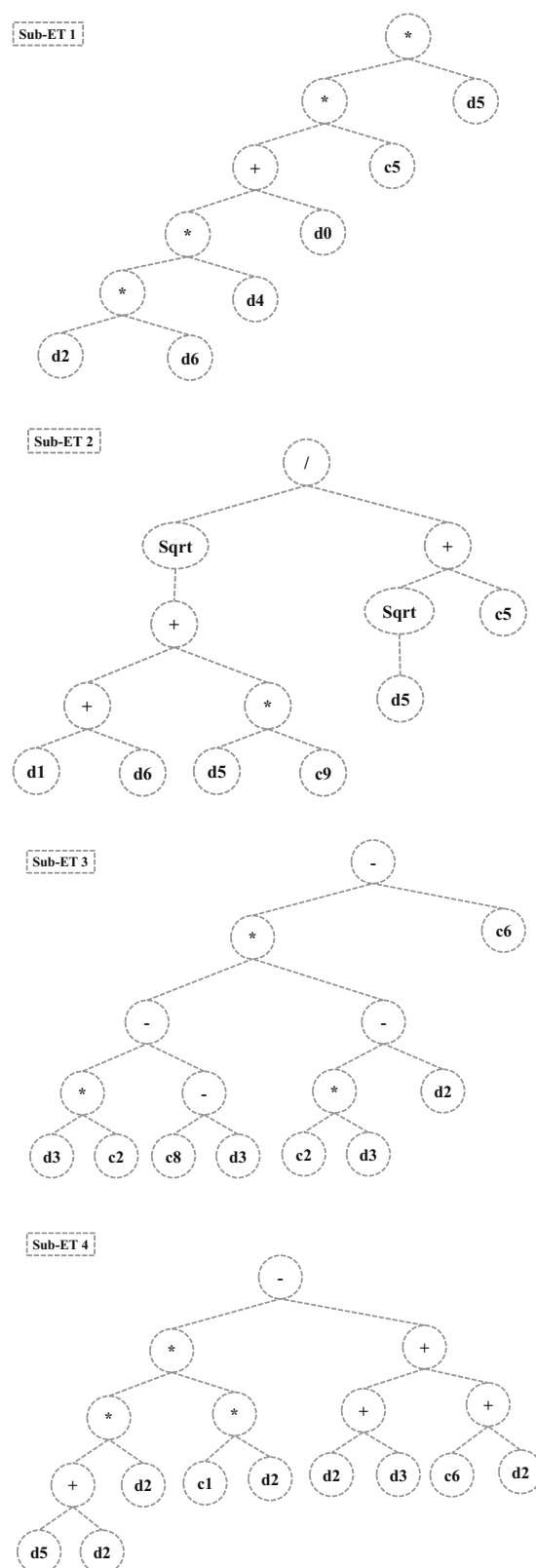


Figure 6: Model of the P-GEP expressed as an expression tree.

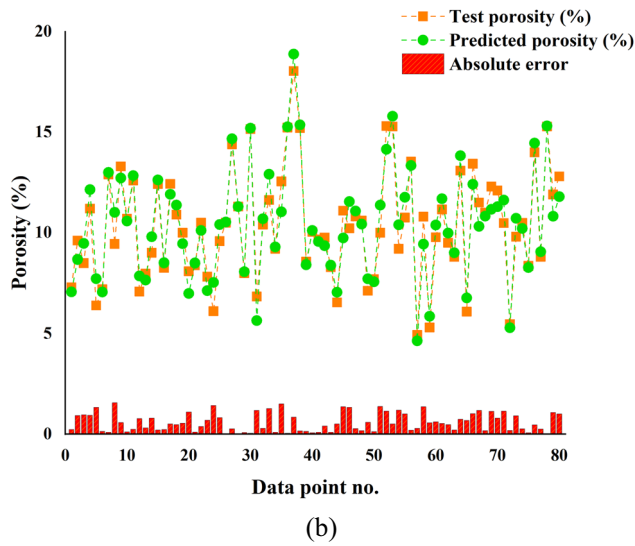
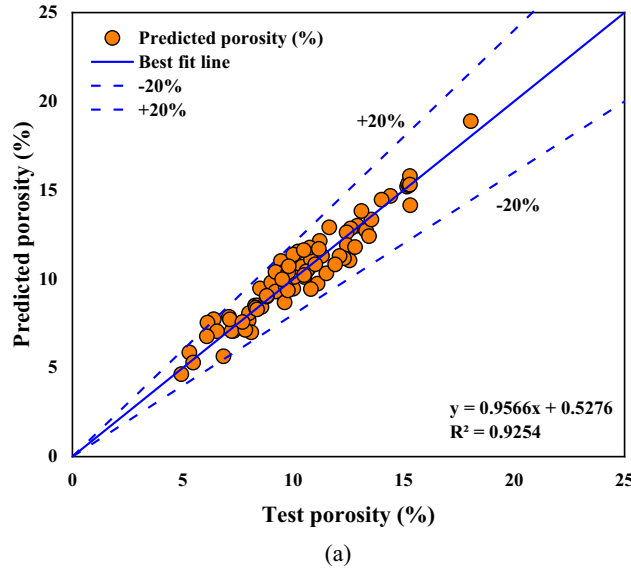


Figure 7: The P-GEP method entails: (a) a correlation between the anticipated and tested P -values and (b) a distribution of the expected and tested P -values, along with all errors.

spread out like a bell curve, as seen in Figure 8. Error readings for 30 results were below 0.5%, 21 were between 0.5 and 1.0%, and 21 were over 1.0%. You should know that maximal error frequencies are extremely rare.

$$P(\%) = A + B + C + D, \quad (9)$$

$$A = (((((FA \times CD) \times SP) + WBR) \times 7.137) \times AR), \quad (10)$$

$$B = \left[\frac{\sqrt{((B + CD) - 9.829 \times AR)}}{\sqrt{AR} - 4.128} \right], \quad (11)$$

$$C = (((S \times -5.184) - (-2.681 \times S)) \times ((-5.184 \times S) - FA)) + 2.189, \quad (12)$$

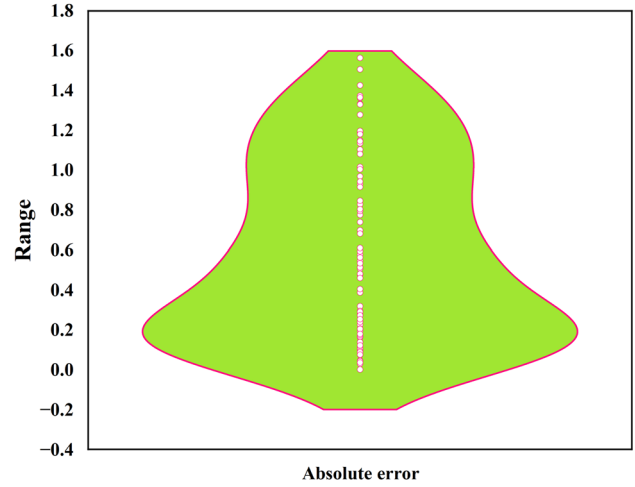


Figure 8: Violin plot for the GEP model's error distribution.

$$D = (((((CD + FA) \times FA) \times (8.161 \times DA)) - ((FA + S) + (-9.823 + FA))), \quad (13)$$

where WBR is the water binder ratio, B is the binder, FA is the fly ash, S is the GGBS, SP is the superplasticizer, AR is the aggregate ratio, and CD is the curing days.

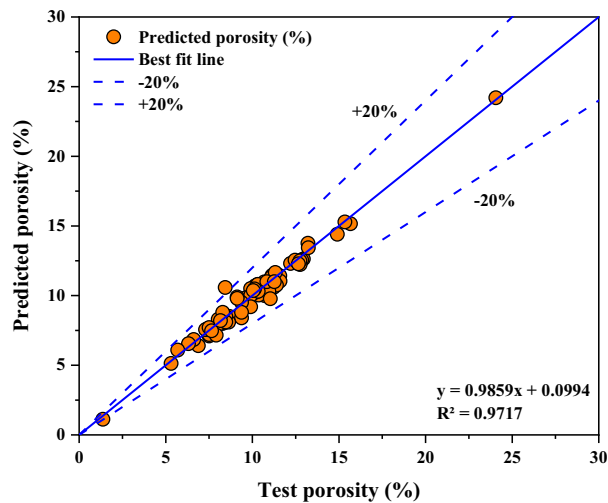
3.2 P-MEP model

To find the P of HPC, an empirical formula was established after examining the MEP findings to consider the effect of the seven autonomous constituents. The last set of sculpted calculated equations is shown in Eq. (14).

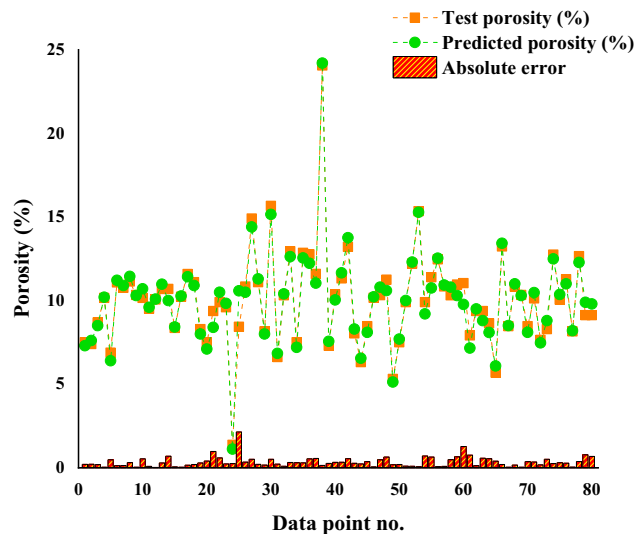
$$\begin{aligned} P(\%) = & (AR + WBR\sqrt{B} - SP - WBR \times S \times CD \\ & - SP((2\sqrt{B} - SP + 2\sqrt{CD} \\ & - B \times FA \times AR)(WBR\sqrt{B} - SP \\ & - SP \times B \times FA \times AR)) - S\sqrt{CD} - \frac{CD}{B + S \times CD} \\ & + S(WBR \times S \times CS + 2SP(2\sqrt{B} - SP + 2\sqrt{CD} \\ & - B \times AR \times CD)), \end{aligned} \quad (14)$$

where WBR is the water binder ratio, B is the binder, FA is the fly ash, S is the GGBS, SP is the superplasticizer, AR is the aggregate ratio, and CD is the curing days.

Figure 9(a) shows that the MEP model can handle oversimplification and is well-trained thanks to its R^2 value of 0.971. Furthermore, it performs adequately on new, untested data. The P-MEP model outperforms the P-GEP model in terms of accuracy, as evidenced by its higher R^2 value. A perfect fit to the data is depicted by the solid blue line in Figure 9(a), whereas the dotted lines show the percent



(a)



(b)

Figure 9: The P-MEP method entails: (a) a correlation between the anticipated and tested P -value and (b) a distribution of the expected and tested P -values, along with all errors.

deviation (20%) from that line. The measured values of P were quite close to the predictions of the MEP model. The MEP approach was used to efficiently identify the P of HPC. Its predictions were within the 20% threshold 99% of the time, indicating very good accuracy. The outcomes of a comparison of the target and actual values, as calculated in MEP simulations, are shown in Figure 9(b). The data showed that the MEP forecast had a marginal error of 0.348% on average and ranged from 0.021 to 2.145%. The overall error rates were below 1.5%, with 59 error values being less than 0.05%, 19 between 0.05 and 1.0%, and just 2 larger than 1.0%. When comparing the two models, the MEP model predicts extreme values better than the GEP model. The MEP

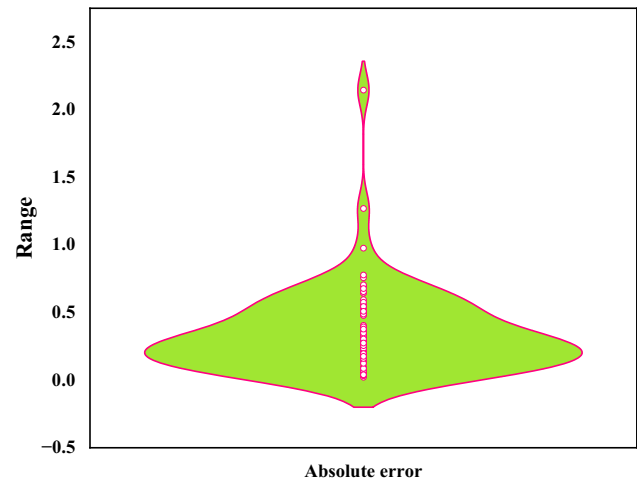


Figure 10: Violin plot for the MEP model's error distribution.

model decreases the correlation coefficient and standard deviations of the errors, as demonstrated in Figure 10's violin plot. Because of its simplicity and generalizability, the MEP equation sees extensive use. With a higher correlation coefficient and lower error levels, in comparison to the GEP model, the MEP model seems to be more effective.

3.3 Validation of the models

Using the previously mentioned Eqs. (2)–(8), the following efficiency and error metrics were calculated: RRMSE, NSE, R , MAE, RSE, and RMSE. The results of these computations are presented in Table 3. Improving the accuracy of predictions is demonstrated by decreasing the error values of the models. The equivalent P-MEP model has an MAE value of only 0.348%, a considerable decrease from the 0.591% found for the P-GEP model. In contrast, the similar P-MEP model saw a significant decrease in the MAPE value from 6.10% in the P-GEP model to 3.80%. The trend was also observed when looking at other error-based statistical

Table 3: Findings obtained through statistical analysis

Property	CrA-GEP	CrA-MEP
MAE (%)	0.591	0.348
MAPE (%)	6.10	3.80
RMSE (%)	0.745	0.469
R	0.962	0.986
RSE (%)	0.244	0.346
NSE	0.923	0.971
RRMSE (%)	0.626	0.512

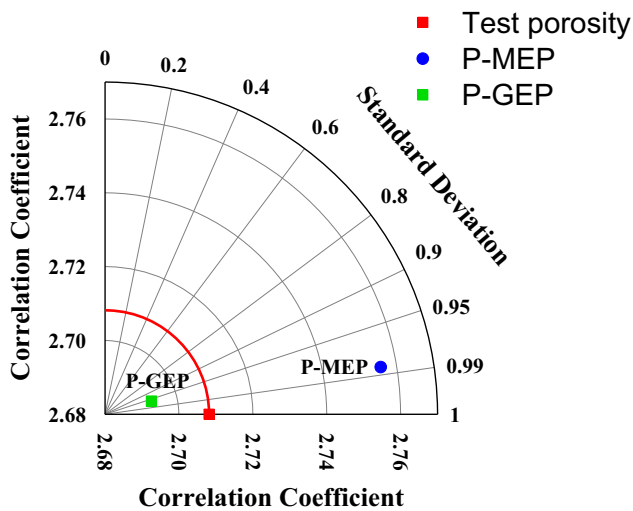


Figure 11: Taylor diagram for the models' validation.

measures including RRMSE, RSE, and RMSE. The produced models were not only validated based on errors but also assessed for efficiency using two metrics: Pearson's coefficient (R) and NSE. The accuracy of a model's predictions is directly proportional to its efficiency rating. The NSE value rose to 0.971 in the similar P-MEP model, compared to 0.923 in the P-GEP model. Applying Pearson's coefficient (R) to the produced models produced similar outcomes. A Taylor diagram comparing the built forecasting models GEP and MEP is shown in Figure 11. The MEP technique seems to be substantially more in line with the experimental line when predicting the P of HPC, in contrast to the GEP model. Due to its high efficiency, small standard deviation, low error, and high R^2 , the MEP technique has the best ML-centered strategy for predicting the P of HPC.

3.4 SHAP analysis findings

The influence of different raw constituents on the P of HPC 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. The results of the violin SHAP graph for each raw material and their effect on the P of HPC are displayed in Figure 12. The x-axis SHAP value shows the relative contribution of each raw ingredient, and the graph employs different hues to indicate the different factors. According to the SHAP analysis plot in Figure 12, there was a positive correlation between the input binder and WBR with the P of HPC. This is seen by the higher-intensity red dots on the positive side of the plot compared to the lower-intensity blue dots on the negative side. It clearly illustrates that increasing the binder and WBR values after a certain limit will result in the increase in P in HPC. The SHAP study confirms the same relationship between binder and WBR with porosity as previous research in the same field [88]. Figure 12's SHAP analysis plot further demonstrates that, as seen by the red dots on the positive side of the plot and the lower intensity of the blue dots on the negative side, the porosity was positively and directly correlated with the CDs and AR. Moreover, the link between the SP, FA, and slag with porosity is more indirect, as more dots can be observed on the negative side of the plot, which implies that increasing the content of these variables will result in a decrease in the P of HPC. It is important to note that these findings rely on inputs and size of the database utilized in this study. By changing the input parameters and data points, diverse results may be obtained.

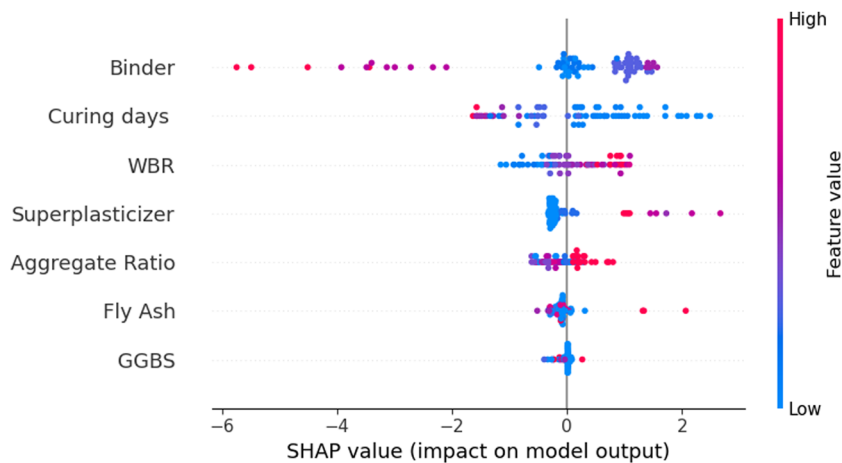


Figure 12: The significance and impact of input elements are suggested by the SHAP plot.

4 Discussion

To compare the findings of the current study, Table 4 has been constructed. Several previous studies that used GEP and MEP tools to estimate the various properties of building materials have been listed in the table. It was identified that the findings of the other researchers were also in line with the present study results, *i.e.*, MEP models exhibited superior predictability than GEP models. The GEP and MEP models used in this study have the benefit of being HPC-specific predictions because they can only be fed data from a small set of seven factors. All the models use the same unit measurements and testing procedure; therefore, their predictions for P are trustworthy. A deeper understanding of the mix design and the effect of each input parameter can be achieved by incorporating mathematical equations into the models. If more than seven parameters

are included in the composite analysis, it can render the predicted models useless. Models trained using data that are drastically different from their target application may fail to deliver expected outcomes. The models' predictive capabilities could be compromised if the units of the input parameters were changed or stored inconsistently. The models cannot function unless the unit sizes are constant. Forecasting material strength, ensuring quality, evaluating risk, performing predictive maintenance, and enhancing energy efficiency are just a few of the many applications of ML-based models in the construction industry. On the other hand, there are certain problems with these models. For instance, they aren't always accurate, they use inaccurate data, and they heavily depend on human input. Future research could look into ways to improve ML-based solutions and overcome these limitations. Some ideas include incorporating IoT devices, creating hybrid models, using

Table 4: List of previous studies that used GEP and MEP methods and comparison with present study results

Ref.	Type of material	Material property predicted	Value of R^2 obtained	
			GEP	MEP
Current study	HPC	Porosity	0.92	0.97
[89]	Eggshell and glass-based concrete	Water absorption	0.88	0.90
[90]	Metakaolin-based concrete	Compressive strength	0.91	0.96
[91]	Alkali-activated concrete	Compressive strength	0.89	0.93
[91]	Alkali-activated concrete	Slump	0.86	0.92
[92]	Plastic sand paver block	Compressive strength	0.87	0.91
[65]	Rice husk ash concrete	Compressive strength	0.83	0.89
[93]	Alkali-activated materials	Compressive strength	0.82	0.86

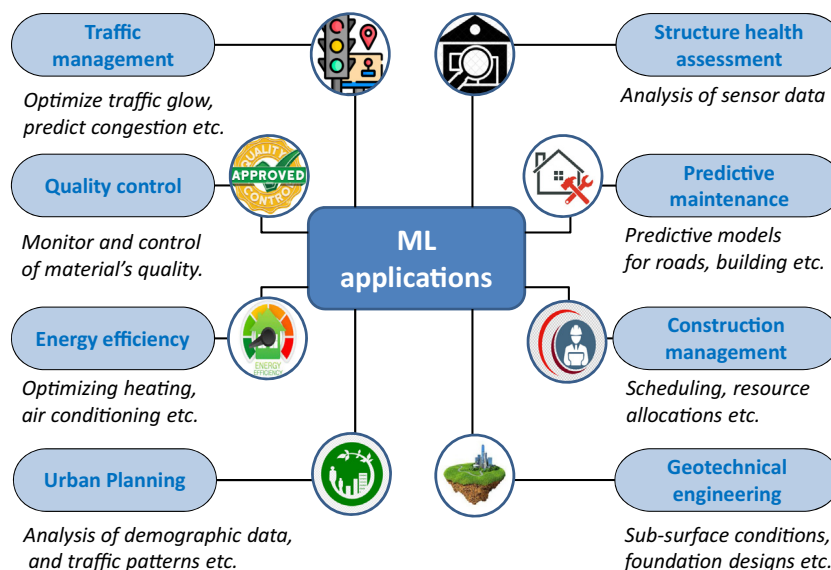


Figure 13: ML applications in civil engineering.

explainable AI techniques, considering sustainability, and customizing data generation and distribution for specific industries. The construction business stands to benefit momentarily from these technological advancements, which could lead to fewer project delays, higher levels of safety, and more sustainable practices by improving efficiency, interpretability, transparency, and informed decision-making. This study's results have the potential to encourage more sustainable building practices by increasing the use of HPC in the construction sector. Figure 13 displays the applications of ML in the field of engineering.

5 Conclusion

In this study, prediction models were built for HPC's porosity using MEP and GEP. A total of 240 data samples were obtained for HPC's porosity from the related experimental studies. The generated dataset was used for training, testing, and model validation. The primary findings of the study are as follows:

- The porosity of HPC was adequately estimated by the GEP approach ($R^2 = 0.925$), whereas the MEP method demonstrated greater precision ($R^2 = 0.971$).
- For GEP, the average discrepancy between actual and predicted porosity (errors) was 0.590%, whereas for MEP, it was 0.348%. The MEP method provided a more accurate prediction of HPC's porosity, and the error rates confirmed the reasonable accuracy of the GEP model.
- The models' efficacy has been validated statistically. ML models have decreased errors and enhanced R^2 . In contrast to the MEP model's 3.80% MAPE, the GEP model was 6.10%. Whereas the GEP model had an RMSE of 0.745%, the MEP model had 0.469%. Additional domains of model performance validation were bolstered by these decisions.
- SHAP analysis showed that the relation of binder, WBR, CDs, and aggregates ratio with the porosity of HPC is direct (positive), which means that increasing the content of these input variables would result in an increase in porosity. Whereas, the relation of SP, fly ash, and GGBS as per the SHAP analysis was more in-direct.

The factor that makes GEP and MEP so crucial for feature prediction in other databases is the unique mathematical expression 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. The built prediction models for the porosity of concrete may aid

academics and the building sector in making quick predictions and mix design optimization of concrete.

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, 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.

References

- [1] Ortega-López, V., J. A. Fuente-Alonso, A. Santamaría, J. T. San-José, and Á. Aragón. Durability studies on fiber-reinforced EAF slag concrete for pavements. *Construction and Building Materials*, Vol. 163, 2018, pp. 471–481.
- [2] 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.
- [3] Soja, W., F. Georget, H. Maraghechi, and K. Scrivener. Evolution of microstructural changes in cement paste during environmental drying. *Cement and Concrete Research*, Vol. 134, 2020, id. 106093.
- [4] 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.
- [5] Zingg, L., M. Briffaut, J. Baroth, and Y. Malecot. Influence of cement matrix porosity on the triaxial behaviour of concrete. *Cement and Concrete Research*, Vol. 80, 2016, pp. 52–59.
- [6] Linares-Alemparte, P., C. Andrade, and D. Baza. Porosity and electrical resistivity-based empirical calculation of the oxygen diffusion coefficient in concrete. *Construction and Building Materials*, Vol. 198, 2019, pp. 710–717.

- [7] Tan, Y., Y. Zhu, and H. Xiao. Evaluation of the hydraulic, physical, and mechanical properties of pervious concrete using iron tailings as coarse aggregates. *Applied Sciences*, Vol. 10, 2020, id. 2691.
- [8] 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.
- [9] Hansen, T. C. Physical structure of hardened cement paste. A classical approach. *Materials and Structures*, Vol. 19, 1986, pp. 423–436.
- [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, 2022, id. 772.
- [11] Basheer, L., P. A. M. Basheer, and A. E. Long. Influence of coarse aggregate on the permeation, durability and the microstructure characteristics of ordinary Portland cement concrete. *Construction and Building Materials*, Vol. 19, 2005, pp. 682–690.
- [12] Ahmad, S., A. K. Azad, and K. F. Loughlin. Effect of the key mixture parameters on tortuosity and permeability of concrete. *Journal of Advanced Concrete Technology*, Vol. 10, 2012, pp. 86–94.
- [13] 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.
- [14] Cao, C. *Machine learning-based prediction of porosity for concrete containing supplementary cementitious materials*. arXiv preprint arXiv:211207353, 2021.
- [15] Cao, C. Prediction of concrete porosity using machine learning. *Results in Engineering*, Vol. 17, 2023, id. 100794.
- [16] Miller, S. A., P. J. M. Monteiro, C. P. Ostertag, and A. Horvath. Concrete mixture proportioning for desired strength and reduced global warming potential. *Construction and Building Materials*, Vol. 128, 2016, pp. 410–421.
- [17] Song, H.-W. and V. Saraswathy. Studies on the corrosion resistance of reinforced steel in concrete with ground granulated blast-furnace slag – An overview. *Journal of Hazardous Materials*, Vol. 138, 2006, pp. 226–233.
- [18] Shi, X., S. Chen, Q. Wang, Y. Lu, S. Ren, and J. Huang. Mechanical framework for geopolymers construction: an optimized LSTM technique to predict compressive strength of fly ash-based geopolymers concrete. *Gels*, Vol. 10, 2024, id. 148.
- [19] Thomas, M. D. A. and J. D. Matthews. The permeability of fly ash concrete. *Materials and Structures*, Vol. 25, 1992, pp. 388–396.
- [20] Hassan, K. E., J. G. Cabrera, and R. S. Maliehe. The effect of mineral admixtures on the properties of high-performance concrete. *Cement and Concrete Composites*, Vol. 22, 2000, pp. 267–271.
- [21] 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.
- [22] 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, 2024, id. 179.
- [23] Papadakis, V. G. Effect of fly ash on Portland cement systems: part I. Low-calcium fly ash. *Cement and Concrete Research*, Vol. 29, 1999, pp. 1727–1736.
- [24] Papadakis, V. G. Effect of fly ash on Portland cement systems: part II. High-calcium fly ash. *Cement and Concrete Research*, Vol. 30, 2000, pp. 1647–1654.
- [25] Chidiac, S. E. and M. Shafikhani. Electrical resistivity model for quantifying concrete chloride diffusion coefficient. *Cement and Concrete Composites*, Vol. 113, 2020, id. 103707.
- [26] Shafikhani, M. and S. E. Chidiac. A holistic model for cement paste and concrete chloride diffusion coefficient. *Cement and Concrete Research*, Vol. 133, 2020, id. 106049.
- [27] Khan, M. I. Permeation of high performance concrete. *Journal of Materials in Civil Engineering*, Vol. 15, 2003, pp. 84–92.
- [28] Khan, M. I. Predicting properties of high performance concrete containing composite cementitious materials using artificial neural networks. *Automation in Construction*, Vol. 22, 2012, pp. 516–524.
- [29] Cai, R., T. Han, W. Liao, J. Huang, D. Li, A. Kumar, et al. Prediction of surface chloride concentration of marine concrete using ensemble machine learning. *Cement and Concrete Research*, Vol. 136, 2020, id. 106164.
- [30] Khan, S. A., M. A. Khan, M. N. Amin, M. Ali, F. Althoeay, 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. 107932.
- [31] Jin, C., Y. Qian, S. A. Khan, W. Ahmad, F. Althoeay, 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.
- [32] 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.
- [33] Huang, J., M. Zhou, J. Zhang, J. 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.
- [34] 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.
- [35] Huang, J. and J. Xue. Optimization of SVR functions for flyrock evaluation in mine blasting operations. *Environmental Earth Sciences*, Vol. 81, 2022, id. 434.
- [36] Boukhatem, B., R. Rebouh, A. Zidol, M. Chekired, and A. Tagnit-Hamou. An intelligent hybrid system for predicting the tortuosity of the pore system of fly ash concrete. *Construction and Building Materials*, Vol. 205, 2019, pp. 274–284.
- [37] 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 geopolymers concrete modified with nanoparticles. *Neural Computing and Applications*, Vol. 35, 2023, pp. 12453–12479.
- [38] 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.
- [39] 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.
- [40] 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.
- [41] 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.
- [42] Wang, R., J. Zhang, Y. Lu, S. Ren, and J. Huang. Towards a reliable design of geopolymers concrete for green landscapes: a comparative study of tree-based and regression-based models. *Buildings*, Vol. 14, 2024, id. 615.
- [43] 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 witzczak 1-40D model: a comparative study. *Materials*, Vol. 15, 2022, id. 1791.
- [44] 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.
- [45] Cheng, A.-S., T. Yen, Y.-W. Liu, and Y.-N. Sheen. Relation between porosity and compressive strength of slag concrete. In: *Structures Congress 2008: Crossing Borders*; 2008, pp. 1–8.
- [46] Al-Amoudi, O. S. B., I. M. Asi, and M. Maslehuddin. Performance and correlation of the properties of fly ash cement concrete. *Cement, Concrete, and Aggregates*, Vol. 18, 1996, pp. 71–77.
- [47] Shafiq, N., M. F. Nuruddin, and I. Kamaruddin. Comparison of engineering and durability properties of fly ash blended cement concrete made in UK and Malaysia. *Advances in Applied Ceramics*, Vol. 106, 2007, pp. 314–318.
- [48] Van den Heede, P., E. Gruyaert, and N. De Belie. Transport properties of high-volume fly ash concrete: capillary water sorption, water sorption under vacuum and gas permeability. *Cement and Concrete Composites*, Vol. 32, 2010, pp. 749–756.
- [49] Younsi, A., P. Turcyr, E. Rozière, A. Ait-Mokhtar, and A. Loukili. Performance-based design and carbonation of concrete with high fly ash content. *Cement and Concrete Composites*, Vol. 33, 2011, pp. 993–1000.
- [50] Ahmad, S. and A. K. Azad. An exploratory study on correlating the permeability of concrete with its porosity and tortuosity. *Advances in Cement Research*, Vol. 25, 2013, pp. 288–294.
- [51] 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.
- [52] Khan, K., W. Ahmad, M. N. Amin, and A. F. Deifalla. Investigating the feasibility of using waste eggshells in cement-based materials for sustainable construction. *Journal of Materials Research and Technology*, Vol. 23, 2023, pp. 4059–4074.
- [53] Khan, K., W. Ahmad, M. N. Amin, M. I. Rafiq, A. M. Abu Arab, I. A. Alabdullah, et al. Evaluating the effectiveness of waste glass powder for the compressive strength improvement of cement mortar using experimental and machine learning methods. *Heliyon*, Vol. 9, No. 5, 2023, id. e16288.
- [54] 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.
- [55] 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.
- [56] 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. 124377.
- [57] Koza, J. *On the programming of computers by means of natural selection. Genetic programming*, MIT Press, Cambridge, Massachusetts, 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. 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.
- [60] 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.
- [61] 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.
- [62] Ferreira, C. *Gene expression programming: mathematical modeling by an artificial intelligence*, Vol. 21, Springer, The Netherlands, 2006.
- [63] 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.
- [64] 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.
- [65] 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.
- [66] 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.
- [67] 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.
- [68] Oltean, M. and C. Grosan. A comparison of several linear genetic programming techniques. *Complex Systems*, Vol. 14, 2003, pp. 285–314.
- [69] 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.
- [70] 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.
- [71] 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.

- [72] Grosan, C. and A. Abraham. Stock market modeling using genetic programming ensembles. In *Genetic Systems Programming: Theory and Experiences*, Springer-Verlag Berlin Heidelberg, 2006, pp. 131–146.
- [73] Oltean, M. and D. Dumitrescu. Multi expression programming. *Journal of Genetic Programming and Evolvable Machines*, 2002.
- [74] 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.
- [75] Shahin, M. A. Genetic programming for modelling of geotechnical engineering systems. In: *Handbook of Genetic Programming Applications*, Springer, 2015, pp. 37–57.
- [76] 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.
- [77] 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.
- [78] 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.
- [79] Wang, Q., T. Cheng, Y. Lu, H. Liu, R. Zhang, and J. Huang. Underground mine safety and health: a hybrid MEREC–CoCoSo system for the selection of best sensor. *Sensors*, Vol. 24, 2024, id. 1285.
- [80] 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.
- [81] 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.
- [82] Kisi, O., J. Shiri, and M. Tombul. Modeling rainfall-runoff process using soft computing techniques. *Computers & Geosciences*, Vol. 51, 2013, pp. 108–117.
- [83] 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.
- [84] Shahin, M. A. Use of evolutionary computing for modelling some complex problems in geotechnical engineering. *Geomechanics and Geoengineering*, Vol. 10, 2015, pp. 109–125.
- [85] 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.
- [86] Taylor, K. E. Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, Vol. 106, 2001, pp. 7183–7192.
- [87] 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, No. 2, 2024, pp. 1557–1582.
- [88] Lee, J.-W., Y.-I. Jang, W.-S. Park, and S.-W. Kim. A study on mechanical properties of porous concrete using cementless binder. *International Journal of Concrete Structures and Materials*, Vol. 10, 2016, pp. 527–537.
- [89] Xia, X. Optimizing and hyper-tuning machine learning models for the water absorption of eggshell and glass-based cementitious composite. *Plos One*, Vol. 19, 2024, id. e0296494.
- [90] Iftikhar Faraz, M., S. Ul Arifeen, M. Nasir Amin, A. Nafees, F. Althoey, and A. Niaz. A comprehensive GEP and MEP analysis of a cement-based concrete containing metakaolin. *Structures*, Vol. 53, 2023, pp. 937–948.
- [91] Zou, B., Y. Wang, M. Nasir 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.
- [92] Iftikhar, B., S. C. Ali, M. Vafaei, M. F. Javed, M. F. Rehman, S. S. Abdullaev, et al. Predicting compressive strength of eco-friendly plastic sand paver blocks using gene expression and artificial intelligence programming. *Scientific Reports*, Vol. 13, 2023, id. 12149.
- [93] Zheng, X., Y. Xie, X. Yang, M. N. Amin, S. Nazar, S. A. Khan, et al. A data-driven approach to predict the compressive strength of alkali-activated materials and correlation of influencing parameters using Shapley Additive explanations (SHAP) analysis. *Journal of Materials Research and Technology*, Vol. 25, 2023, pp. 4074–4093.