Research Article

Hao Liu*, Suleman Ayub Khan*, Muhammad Nasir Amin, Fadi Althoey, and Muhammad Tahir Qadir

Evaluating the strength loss and the effectiveness of glass and eggshell powder for cement mortar under acidic conditions

https://doi.org/10.1515/rams-2024-0042
received April 08, 2024; accepted June 28, 2024

Abstract: The cementitious composite's resistance to the introduction of harmful ions is the primary criterion that is used to evaluate its durability. The efficacy of glass and eggshell powder in cement mortar exposed to 5% sulfuric acid solutions was investigated in this study using artificial intelligence (AI)-aided approaches. Prediction models based on AI were built using experimental datasets with multi-expression programming (MEP) and gene expression programming (GEP) to forecast the percentage decrease in compressive strength (CS) after acid exposure. Furthermore, SHapley Additive exPlanations (SHAP) analysis was used to examine the significance of prospective constituents. The results of the experiments substantiated these models. High coefficient of determination (R^2) values (MEP: 0.950 and GEP: 0.913) indicated statistical significance, meaning that test results and anticipated outcomes were consistent with each other and with the MEP and GEP models, respectively. According to SHAP analysis, the amount of eggshell and glass powder (GP) had the most significant link with CS loss after acid deterioration, showing a positive and negative correlation, respectively. In order to optimize efficiency and cost-effectiveness, the created models possess the capability to theoretically assess the decline in CS of GP-modified mortar across various input parameter values.

Keywords: cement mortar, acid attack, compressive strength

*** Corresponding author: Hao Liu,** College of Architecture, Zhejiang Tongji Vocational College of Science and Technology, Hangzhou, 311231, Zhejiang, China, e-mail: hzliuhao516@163.com

*** Corresponding author: Suleman Ayub Khan,** Department of Civil Engineering, COMSATS University Islamabad, Abbottabad, 22060, Pakistan, e-mail: sulemanayub@cuiatd.edu.pk

Muhammad Nasir Amin, Muhammad Tahir Qadir: Department of Civil and Environmental Engineering, College of Engineering, King Faisal University, Al-Ahsa, 31982, Saudi Arabia

Fadi Althoey: Department of Civil Engineering, College of Engineering, Najran University, Najran, Saudi Arabia

1 Introduction

The annual global consumption of cement-based materials (CBMs) is second only to that of water [1]. Derivatives of CBMs are widely employed in construction owing to their low cost and high durability [2,3]. Durable CBMs are characterized by their ability to maintain their functionality, mechanical performance, and quality even after being subjected to ecological conditions [4]. The resistance of CBMs to chemical assaults, weathering, abrasion, and other types of corrosion [5] is one indicator of their durability. CBMs decay when exposed to various harsh elements. The source of the attack could be internal or external, and the attacking method could be mechanical, physical, or chemical. Physical and chemical assaults can ruin the composite's paste and aggregate [6]. Several different types of harmful factors [7] reduce the strength and durability performance of CBMs; thus, their effectiveness in an aggressive setting is the main concern. Products derived from cement are frequently attacked by acid, sulfate, and other harmful elements because of the rapid expansion in the business sector. Currently, the building industry is primarily concerned with creating a material that can endure severe surroundings and maintain its expected lifespan [8].

Ion resistance affects CBM durability. Porosity can be estimated from CBM absorbency, void volume, and pore connectivity [9]. Items made of cement are highly vulnerable to corrosion when exposed to sulfuric acid (H_2SO_4) [10]. Hydroxide gas can hasten the deterioration and eventual collapse of cement-based concrete structures. Ions that dissolve in water due to reactions between cement paste's calcium-silicate-hydrate (CSH) gel and $Ca(OH)_2$ when subjected to weak or strong acids dissolve the components of CBMs [7]. H_2SO_4 is one of the most destructive acids for CBMs due to the presence of sulfate in its assault [10]. As a result of sulfates' chemical attack and salt crystallization's physical attack [11], the rapid degradation of CBMs can be accelerated by sulfate attack. CSH decomposition results in the production of new chemicals that exacerbate cement-

Open Access. © 2024 the author(s), published by De Gruyter. This work is licensed under the Creative Commons Attribution 4.0 International License.

based products' durability problems [11]. Therefore, it is crucial to investigate how acid attacks can reduce the durability of CBMs.

Energy-intensive and carbon dioxide release with cement production is a major contributor to climate change [12]. Recycling and reusing debris could help cement manufacturers lower production costs and carbon dioxide output [13–15]. Therefore, there is an urgent need for cement-based products that are beneficial to the environment in the building sector [16]. Furthermore, natural aggregate extraction results in significant carbon dioxide emissions and resource depletion [17]. Cement that has been treated with recycled glass powder (GP) and eggshell powder (EP) is a popular choice among construction materials due to its accessibility and affordability [18,19]. Replacing 10–20% of the cement or sand with GP can enhance the mechanical properties of the material, including its compressive and flexural strengths [20]. Also, using EP as cement or sand replacement in cementitious composites may enhance the strength properties [21]. Reduced cement demand and CO_2 emissions, conservation of natural resources, and simplification of waste management are all outcomes of utilizing GP and EP in place of cement [22,23].

Performance forecasting models for materials and structures are now being developed by professionals in order to cut costs and save time [24]. Attribute estimates are made using forecasting models, such as regression-based approaches [25–27]. Machine learning (ML) and other artificial intelligence (AI) techniques are currently at the forefront of model creation in this field [28,29]. Many fields use a variety of modeling tools, not just ML, to probe a wide range of questions [30]. The urge for ML techniques for estimating the functionality of construction materials has increased. Despite the fact that most current ML studies have focused on traditional CBM strengths [31,32], for CBMs modified with GP and EP, only a small number of research have focused on property predicting [12,21,33,34]. Table 1 summarizes past ML studies. However, empirical equation-based ML models using gene expression programming (GEP) and multi-expression programming (MEP) have not been developed for estimating the loss in compressive strength (CS) after an acid attack.

The objective of this work was to scrutinize the impact of acid assault on CS in GP-and EP-modified cement mortar using ML approaches. Models of ML estimation were developed using data collected during the experimental strategy. The research objectives were attained by employing ML techniques, such as MEP and GEP. The unique mathematical formulas provided by MEP and GEP allow for the estimation of features in a different database, which is why it is so important, contrary to the other ML techniques. The mathematical models used in this research can facilitate quick evaluation, improvement, and rationalization of mortar mixtures proportioning by scientists and engineers. Various methods were utilized to evaluate the performance of ML algorithms, including the R^2 coefficient, statistical analyses, and the variability of predicted results. The primary goal of this research was to evaluate how effectively ML techniques can predict material quality. These ML approaches require a dataset, which can be created through exploratory experiments or by examining existing data sources. Input from this data set could be used by ML models to approximate material qualities. The capacity of ML approaches to foretell changes in CS following an acid assault on cement mortar with GP and EP was evaluated using seven input parameters in conjunction with experimental data. To go even further into the importance of raw components, SHapley Additive exPlanations (SHAP) analysis was conducted. These models, using algorithms like GEP and MEP, enhance the accuracy of predictions for factors such as CS, durability, and performance under various conditions, thereby improving construction quality and safety.

2 Overview of ML methods

2.1 GEP

Holland was the first person to present the genetic algorithm (GA) [40]. Its foundation rests on Darwin's idea of natural selection. An accumulation of GAs is used to

Table 1: Literature-ML research

Ref.	Materials investigated	Attributes projected	ML technique utilized
[35]	RHA-based concrete	CS	AdaBoost, Extreme gradient boosting, and Gradient boosting
[36]	GP-mortar	Flexural strength	BR, and support vector machine (SVM)
[37]	Geopolymer concrete	CS	MEP and GEP
[38]	Mining waste-based cement	CS	SVM, decision tree, and RF
[39]	Metakaolin-centered concrete	Mechanical characteristics	MEP and GEP

illustrate the advancement of the genomic process and chromosomes of constant length are used to illustrate the resolution of the process. In his proposal, Koza suggested a variation of GA that he referred to as gene programming (GPg) [41]. The GPg approach is a generic method for solving issues that employ genetic evolution to automatically develop a model [42]. Parse trees and other nonlinear structures are utilized in place of binary strings of a fixed length, which is one of the reasons why GPg is such a versatile technique. To handle reproduction-associated challenges in line with Darwinian theory [43], a well-established machine intelligence software uses chromosomal parameters (such as mutation, crossover, and reproduction) that occur naturally. GPg eliminates poor-performing programs during reproduction. As was the case in the previous instance, the trees that are the least suitable for the region are cut down, and the ones that are left are used to repopulate the area. Protecting early model convergence is the function of the evolution process [44]. GPg eliminates poor-performing programs during reproduction. The process is the same as before: the worst trees are chopped down, and the best ones are planted back into the ground. The evolution process protects early model convergence [43]. A large number of the parse trees are generated by a crossover chromosomal processor despite the fact that GPg is capable of automatically generating a model [45]. The creation of nonlinear GPg forms necessitates dual roles as genotype and phenotype, resulting in convoluted representations for desirable features [46].

Ferreira is the one who initially presented the GEP, which is a version of the GPg [46]. Parse trees and fixed-length linear chromosomes are used in GEP modeling, which follows the population generation hypothesis. An improved GPg, GEP, uses chromosomes of set length to encrypt small software programs. GEP can predict complex and nonlinear issues with mathematical equations [47,48]. In accordance with the GPg standard, the fitness operation, the ultimate set, and the final criteria have all been set. The GEP method generates random chromosomes, which are identified and numbered using "Karva" language. GEP uses static-length lines. GPg contemplates parse trees of dissimilar lengths while programming with data. These unique strings are expressed as nonlinear expression/parse trees with branch forms that range in size to represent the chromosomes after being coded as fixed-length genomes [49]. Furthermore, these genotypes, in addition to certain forms of phenol, are encoded [46]. One benefit of genetic engineering is its ability to transfer genomes directly from one generation to the next without causing structural mutations or duplications. This is one of the reasons why genetic engineering is so beneficial.

Typically, a chromosome is made up of two parts: the "head" and the "tail." Consequently, the creation of creatures from a single gene that contains a large number of genes is yet another distinctive characteristic [43]. These genes encode a diverse array of operations, encompassing logical, mathematical, arithmetic, and Boolean logic functions. It is the responsibility of the operators of the genetic code to connect cells to the roles that have been allocated to them. In order to obtain and infer the information that is contained in these chromosomes, a whole new language known as Karva is utilized. The development of empirical formulas is facilitated by this language. Next, starting with Karva, a leading revolution is used to traverse the expression tree (ET). This then follows the previous step. In accordance with Eq. (1), the nodes that are located in the layer that is the lowest are moved to the bottom and recorded by ET [47]. It is conceivable that variations in the quantity of the GEP gene and the duration of K-expression could be influenced by an unequal distribution of ETs. This is something that deserves more investigation.

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

Due to the fact that its outcomes are not reliant on any established associations, GEP is considered to be an advanced ML technique. Figure 1 depicts in detail the various steps that were used to develop the GEP mathematical equation, which offers a comprehensive overview of these processes. During birth, the number of chromosomes that are present in each individual is predetermined. After that, these chromosomes are formally designated as ETs after evaluations have been conducted on the health of all individuals. One of the members of the population who is the healthiest is selected to reproduce. Iterative processes, which include the participation of the most appropriate individuals, are used to arrive at the best possible answer. There are three genetic processes that are finally utilized in order to get at the ultimate numerical expression. These processes are mutation, crossover, and breeding.

2.2 MEP

The MEP encrypts solutions using a novel linear-based GPg approach that has been proven effective. Similar to the MEP, the GEP is based on the system. A distinctive feature of MEP, a relatively recent advancement derived from the GPg method, is its ability to encode multiple software components (variants) on a single chromosome. The fitness values are then used to choose the best chromosome [50],

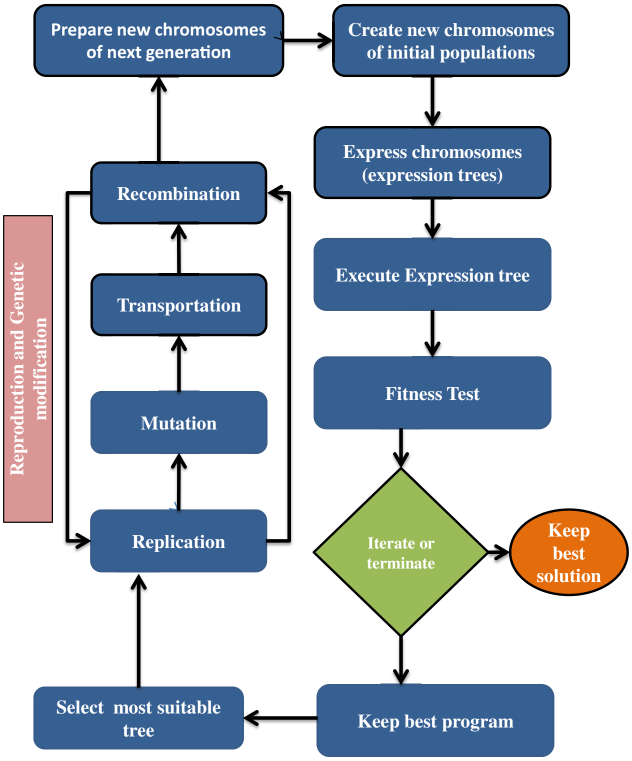


Figure 1: GEP technique flow diagram [35].

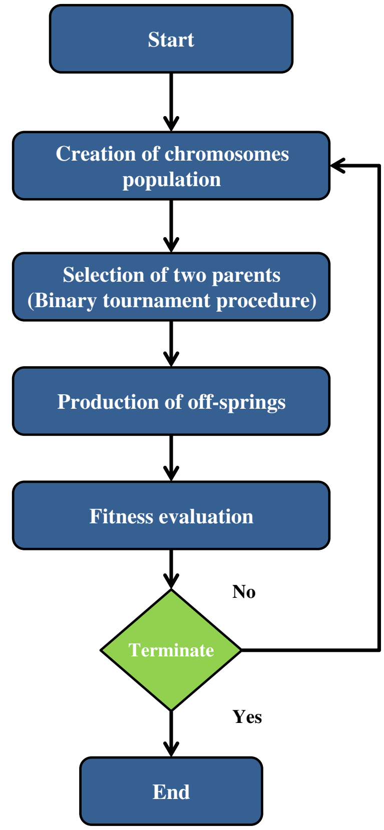


Figure 2: Step-by-step method of MEP [35].

yielding the ultimate solution. According to Grosan and Oltean [51], choosing between two parents is the consequence of a method in which a binary environment is recombined to produce two different children. As presented in Figure 2, the procedure remains until the most suitable platform is discovered, which occurs before the criteria for termination are satisfied. This is the location where mutations that occur in newborns take place. Analogous to how the GEP model permits the fitting of several factors, the MEP method also allows for this option. The key regulators of MEP are variables such as the range and size of the subpopulation, the chance of crossover, and the function set [52]. When the population size is all programs, it is harder to estimate and slower to account for. Similarly, the code length strongly affects the mathematical expression size. Table 3 shows all MEP parameters used to estimate the CS credibly.

The appraisal and simulation stages of both approaches commonly make use of data sets that are derived from the literature that is pertinent to the situation [53]. The methods of linear generalization, such as the GEP and MEP methods, are becoming increasingly popular and can be used to make more correct forecasts concerning the potential of sustainable concrete. Combining linear chromosomal programming with maximum likelihood estimation (MEP) was the most effective neural network-based technique, according to Grosan and Abraham. This held true when contrasted with alternative methods that relied on neural networks [54]. The mechanism of the GEP is comparatively more complicated than the mechanism of the MEP [52]. MEP recycles code does not require non-coding items to be exhibited at a static place in the chromosomes, and uses explicit encryption for function argument references. The density of MEP is lower than that of GEP [55]. It has been observed that its chromosome has the usual GEP head and tail, which include codes that neatly encode syntactically correct software [51]. Therefore, more study is needed to establish causality and assess the efficacy of these two chromosomal methods for solving certain engineering problems.

3 Research strategy

3.1 Data collection

With the utilization of 225 experimental results taken from the literature as a dataset [56,57], the MEP and GEP models were used in this investigation to calculate the CS of cement mortar that has been modified by GP and EP as a

Table 2: Descriptive statistics of dataset retrieved from previous studies [56,57]

Descriptive statistics	Cement ($kg{\cdot}m^{-3}$)	Sand ($kg{\cdot}m^{-3}$)	Water ($kg{\cdot}m^{-3}$)	SF ($kg{\cdot}m^{-3}$)	SP ($kg{\cdot}m^{-3}$)	GP ($kg{\cdot}m^{-3}$)	EP ($kg{\cdot}m^{-3}$)	%CS-loss
Mean	730.42	734.16	191.15	152.15	38.13	32.58	30.98	10.82
Standard error	2.50	2.55	0.44	1.10	0.09	1.91	1.90	0.21
Median	721.00	729.00	191.00	153.00	38.00	0.00	0.00	10.49
Mode	760.00	810.00	191.00	153.00	38.00	0.00	0.00	6.81
Standard deviation	53.05	53.99	9.29	23.43	1.82	40.58	40.29	4.46
Sample variance	2814.78	2915.01	86.26	548.88	3.31	1647.02	1623.49	19.88
Kurtosis	−0.51	−0.64	−1.46	−1.46	−1.45	−0.70	−0.62	0.23
Skewness	−0.26	−0.30	0.08	−0.12	0.17	0.88	0.93	0.59
Range	198.00	198.00	23.00	58.00	4.50	121.50	121.50	24.37
Minimum	612.00	612.00	180.00	122.00	36.00	0.00	0.00	1.11
Maximum	810.00	810.00	203.00	180.00	40.50	121.50	121.50	25.48
Sum	328687.00	330371.00	86016.00	68467.00	17157.50	14658.80	13939.80	4870.92
Count	450.00	450.00	450.00	450.00	450.00	450.00	450.00	450.00

consequence of an acid attack. Cement (C), water (W), recycled GP, sand (S), EP, superplasticizer (SP), and silica fume (SF) were the seven independent variables that were employed to make an estimation concerning the percentage loss in CS (%CS-loss) of mortar that occurred after an acid attack. The data were prepared in order to facilitate its collection and organization. The dataset was expanded from its initial 225 data points to 450 with the assistance of a Python code that adhered to a predetermined procedure. The code is initiated by allowing the user to select a database file from a Tkinter-based file dialog box. After importing the file into a Pandas DataFrame, the code verified the current point count. The enhanced dataset was subsequently stored in a newly generated file combining synthetic and original data. A similar procedure was also used in prior research to increase the data points of a database [58]. Data mining is a well-known technique that involves the finding of knowledge from data. One approach to overcome a big impediment in this process is to prepare the data for data mining. For the purpose of streamlining the data, many methods for filtering the background noise and other information that is not important are included in the data preparation process. Also, many different types of specialists have speculated that the data-to-input ratio is the most important factor in how well the proposed model works. The best model for analyzing data points in order to find the relationship between the mentioned variables calls for a ratio higher than 5 [59]. A ratio of 64.3 meets the criteria established by the researchers in this study, which employs seven inputs with 450 data points. A regression analysis and several strategies for error distribution were utilized in the analysis of the model. These data were subjected to descriptive statistics, and the findings are presented in Table 2. Additionally, the validation approach was performed in order to evaluate the accuracy of the models that were utilized. Figure 3 also shows the frequency distribution of all variables using histograms, which is crucial. The distribution of all of the variables that serve as input can be used to provide a description of the overall frequency of occurrence in a data collection. In order to acquire an understanding of the frequency with which certain values arise in a data collection, one can create a relative frequency distribution. This can be done in order to gain this information.

3.2 Modeling methods

There are a substantial number of input factors that are required for ML approaches in order to arrive at the desired result [60]. In order for the ML method to yield perfect results, the data sample's properties must be dynamic. There is a possibility that the results will be in the middle of the pack if you use a value that is either static or variable with restricted change [21,61]. To forecast the CS of GP and EP-modified cement mortar following an acid assault, experimental data were analyzed. Water, superplasticizer, sand, GP, EP, cement, and SF were used as inputs by ML approaches to predict the CS loss succeeding the acid attack. A combination of GEP and MEP methods was employed to achieve the ML study's aims. In keeping with previous research, the algorithms were trained using 67% of the data and tested using the remaining 33% [12]. The R^2 number is a measure that quantifies the degree to which the data, as observed, corresponds to the theoretical predictions. The R^2 number is a representation of the gap that exists between the model and the data that has been observed [62]. The divergence is higher when the value is near zero and

Table 3: MEP and GEP models with set parameters (parameters similar to ref. [35])

GEP		MEP	
Parameters	**Settings**	**Parameters**	**Settings**
General	%CS loss	Code length	25
Head size	8	Sub-population size	1,000
Linking function	Addition	Number of sub-populations	100
Chromosomes	200	Function set	+, −, ×, ÷, square root
IS transposition rate	0.00546	Replication number	15
Stumbling mutation	0.00141	Crossover probability	0.9
Upper bound	10	Mutation probability	
Function set	+, −, ×, ÷, square root	Number of runs	15
Gene transposition rate	0.00277	Operators/variables	0.5
Constant per gene	10	Number of generations	1,000
Two-point recombination rate	0.00277	Terminal set	Problem input
Mutation rate	0.00138	Number of treads	
One-point recombination rate	0.00277	Problem type	Regression
Data type	Floating number	Number of generations	1000
Genes	4	Error	MSE, MAE
RIS transposition rate	0.00546		
Leaf mutation	0.00546		
Inversion rate	0.00546		
Gene recombination rate	0.00277		
Lower bound	−10		
Random chromosomes	0.0026		

smaller when it is near one. To further assess the model's utility, statistical methods were also employed. Additionally, statistical validation and SHAP analysis were utilized to delve deeper into the significance of the raw components. Figure 4 shows the processes that go into ML-based modeling. What follows is a rundown of the study's validation procedures and ML methods.

3.3 Structure development of GEP and MEP models

Selecting suitable input parameters is the initial stage in building an AI model. For this study, seven input variables were chosen for their possible effect on the percentage of CS loss in GP and EP-modified cement mortar. GEP completed the construction of these models with the help of GeneXproTools version 5.0. Code is generated using GeneXproTools, a data generator after the variables are first classified, and then their missing values are randomly generated and processed. When compared to the prototype that came before it, the version that was developed is more efficient and, in general, of superior quality [63]. A variety of programming languages, including Visual Basic, C++, and MATLAB, facilitate both the creation of programs and models [64]. After a great deal of trial and error, as well as comparison to earlier research, the GEP parameters that were utilized in this investigation were established [65,66]. To test how changing GEP parameters affected the accuracy of predictions, the best starting combination was found *via* trial and error. Optimal ordering of GEP hyper-parameters was selected and included in the modeling process for the purpose of making accurate predictions and converting them into comprehensible mathematical expressions (refer to Table 3). A higher chromosome and gene count, together with a larger head size, is associated with an increase in complexity. It will take more time to run the application as these values increase. A greater number of genes and chromosomes, however, allows for a more precise model in general. Previous research on GP and EP mortar looked at two ensemble ML methods, BR and random forest (RF) [56]. Nonetheless, GEP and MEP are two distinct genetic ML approaches, and this study compares and analyzes their results. The results of the present study are compared in Table 3 according to the quality of their hyper-parameters tuning, the success of their statistical tests, and the accuracy of their mathematical expression for future prediction. Also, the previous efforts [56] failed to incorporate hyper-parameter tuning information into the GEP mathematical expression-building process. Moreover, independently verifying the results of the RF method is challenging without access to the Python code. Conversely, two mathematical models related to genomics

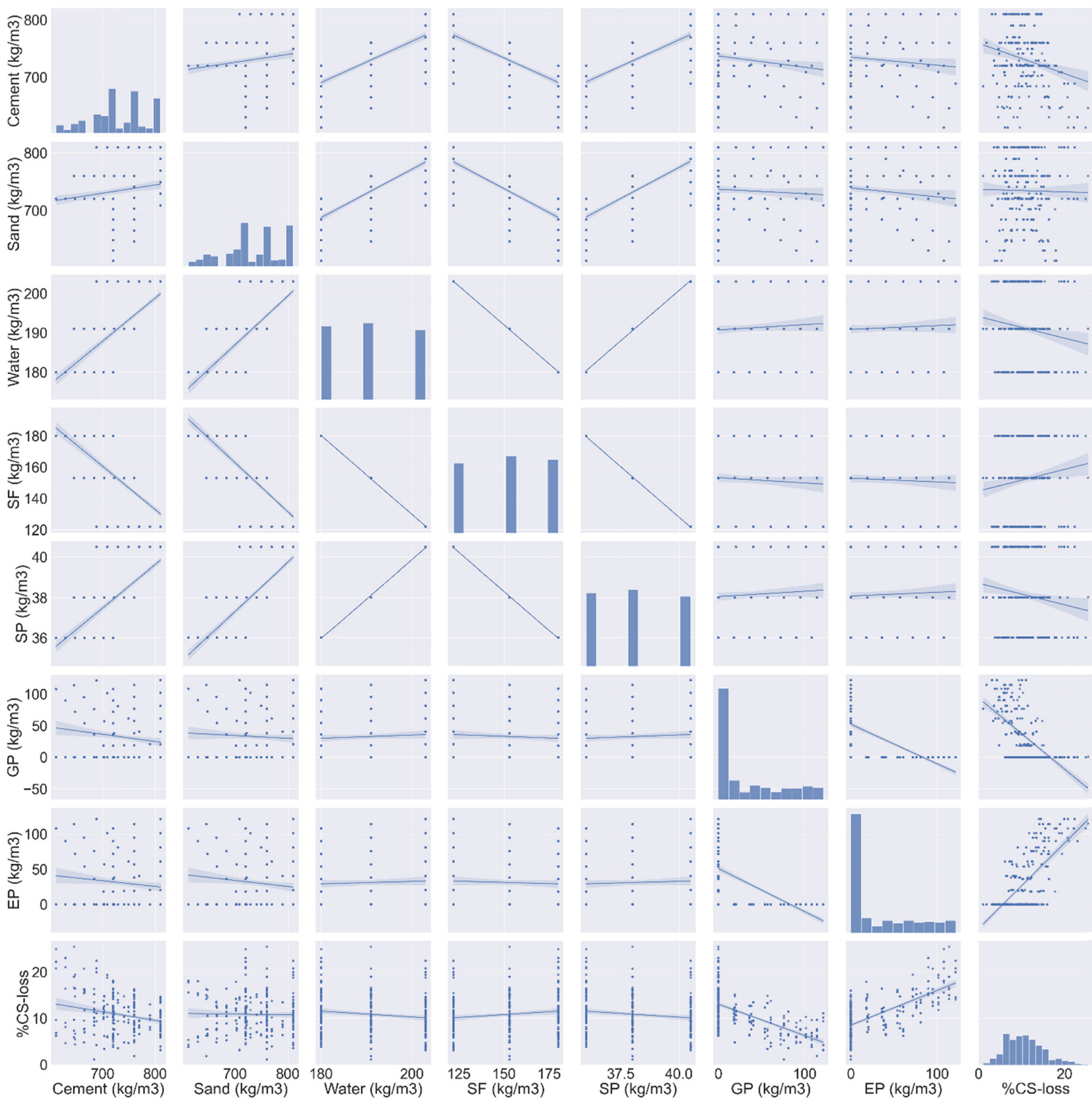


Figure 3: Database input and output frequency distribution.

were employed to present the findings of this study. A training set, consisting of 67% of the data, was used to develop the models, while a testing and validation set, comprising 33% of the data, was used to evaluate them. It is easy to make predictions about the future using the whole spectrum of data offered in this study.

It has already been mentioned that the linear variety of MEP is the most famous kind of genetic programming [67]. Since the MEP can also provide an equation centered on the outcomes of the construction cost prediction models [55], it was used to design the compaction constraints by employing a finished 450 datasets within the multi-expression programming X. Despite being variable-length substrings, MEP genes guarantee that the overall length of every chromosome is directly proportionate to the number of genes found there. An endpoint or abstract function representation is stored by each gene, and pointers to the arguments provided to a function are also stored by the genes that code for that function. Over the course of this study, the indicator values for the parameters of the function have consistently been found to be lower than the placement of the relevant function on that chromosome [51].

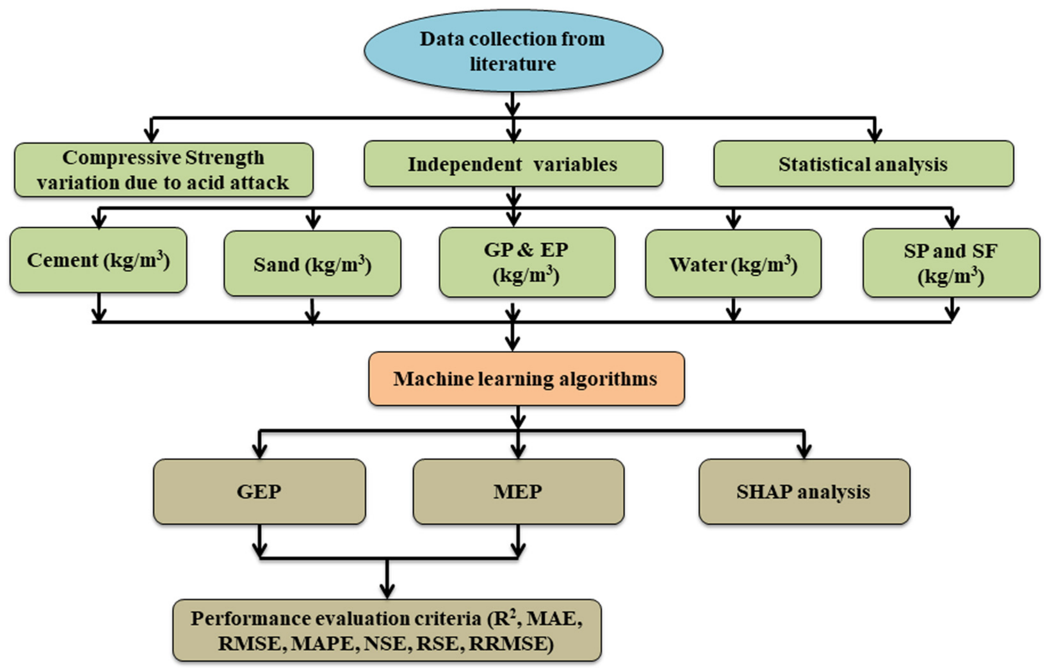


Figure 4: Methods of generating, modeling, and validating data.

A cursory examination of the formulation is performed, and additional reading material on the expanded GEP approach and MEP modeling can be found elsewhere [67]. Table 3 describes the hyperparameters used in constructing these models, which are crucial for tuning the algorithms to achieve optimal performance. Hyperparameters, such as population size, mutation rate, and number of generations, determine the learning process and significantly impact the models' accuracy and efficiency.

3.4 Model performance evaluation

The statistical effectiveness of the models generated by GEP and MEP was evaluated using training, testing, and validation datasets. Seven distinct arithmetic metrics were computed for each of the three groups: Pearson's correlation coefficient (R), mean absolute error (MAE), Nash-Sutcliffe efficiency (NSE), root mean square error (RMSE), relative squared error (RSE), relative RMSE, and mean absolute percentage error (MAPE) [53,65]. All of these statistical measures are as follows (Eqs (2)–(8)):

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

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

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

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

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

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

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

The numbers a_i and p_i, which represent the ith real and projected results, respectively, are the symbols that are used. These symbols denote the mean values of the actual and predicted results, respectively, with "n" representing the total number of observations in the dataset. There is a simultaneous representation of both of these symbols. The correlation between expected and observed outcomes (a_i and p_i) is known as R, and it is a prominent statistic for evaluating the usefulness of a model. A strong correlation between the actual and expected production amounts exists when $R > 0.8$ [68]. The division and multiplication operations do not trigger an output from R. The factor R^2 was calculated for both the actual and projected outcomes since the model is now more functional and delivers more balanced estimation results. More efficient

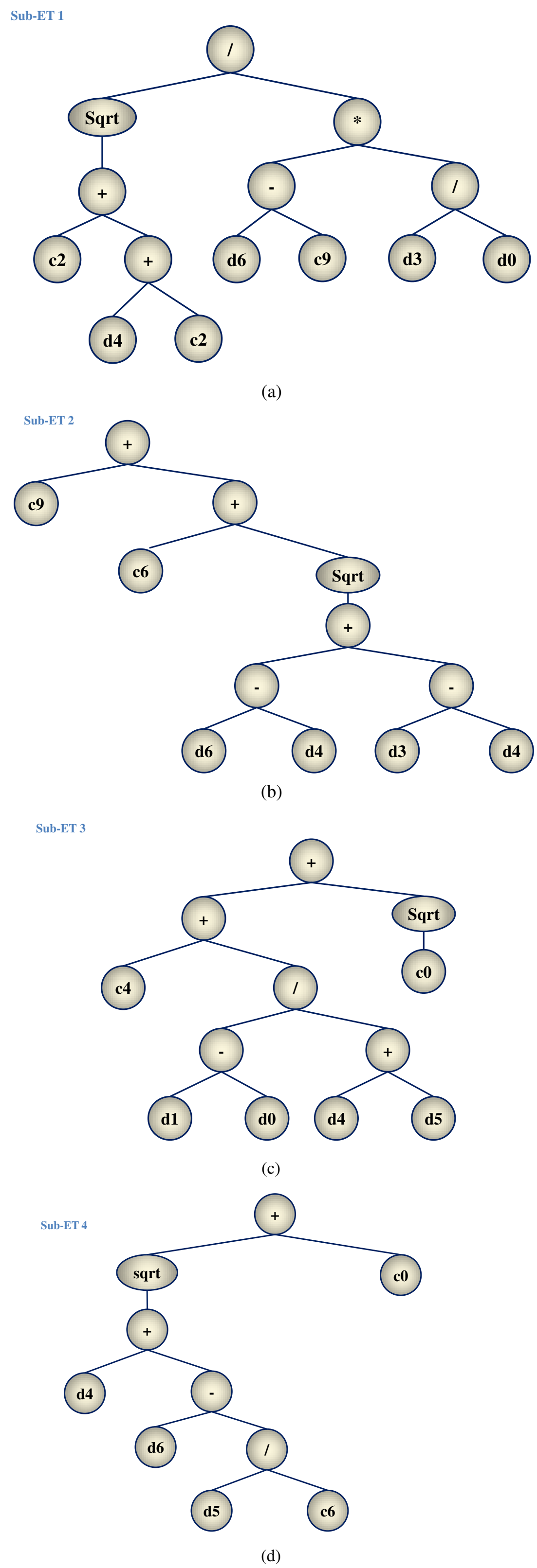


Figure 5: Expression tree representing the finalized model. (a) Sub-ET 1, (b) Sub-ET 2, (c) Sub-ET 3 and (d) Sub-ET 4.

model construction is indicated by a larger R^2 value, which is closer to 1 [69]. Like RMSE, MAE was able to manage higher error levels with ease. Errors have less of an effect on the enactment of the produced model if the MAE and RMSE values are smaller [70]. MAE, however, was discovered to be useful primarily for continuous and smooth data sets [71]. Typically, when the previously computed error values drop, the model 's performance increases.

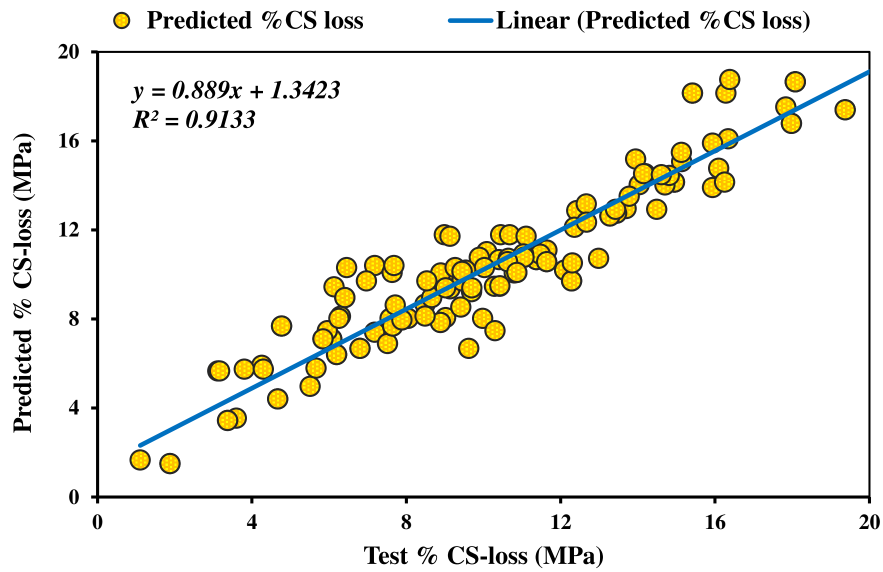
4 Results and analysis

4.1 GEP model development

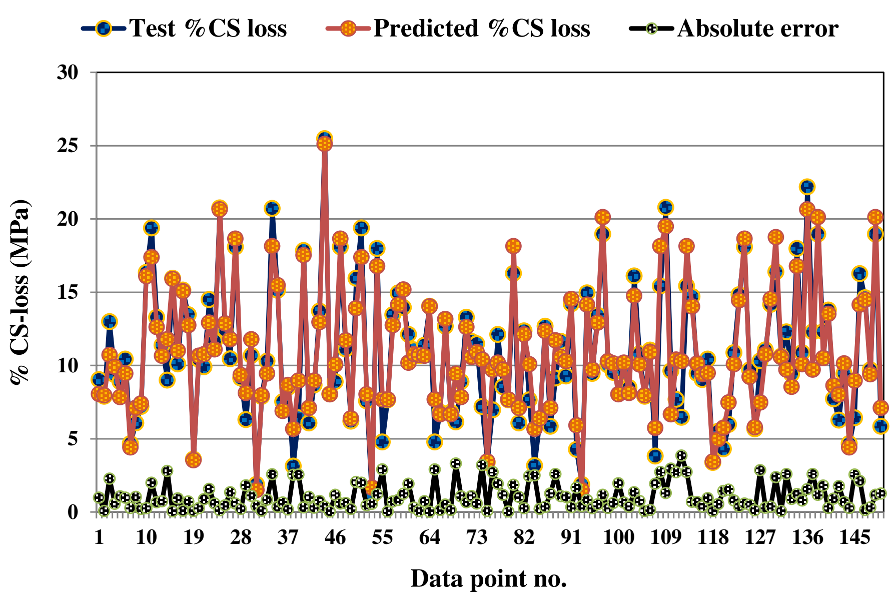
Using head size and chromosomal number to deduce mathematical linkages, the GEP method (as shown in Figure 5(a)–(d)) generated ET-centered models to predict the percentage of CS loss induced by the acid assault. The five arithmetic operators (square root, ÷, +, −, x) are used to create the majority of the %CS-loss following acid assault sub-ETs. The result is a mathematical formulation produced by encoding the sub-ETs of the GEP model. In order to predict the future %CS-loss, the formula in Eq. (9) with these input values can be utilized. The generated model for GP and EP-based mortar has enough data points to outperform an optimal model under perfect circumstances. Figure 6(a) displays regression lines illustrating the relationship between CS loss, comparing the model's predictions with experimental data from both the training and validation sets. The strong agreement (R^2 = 0.913) between the actual and predicted outcomes indicates that the GEP method was successful in accurately estimating the CS of mortar modified with GP and EP. Figure 6(b) shows the experimental data plotted against absolute error, which graphically shows the highest percentage of error for the proposed GEP equation. With a minimum of 0.016 and a maximum of 3.841, the experimental data and the GEP equation show an average absolute error of barely 1.044. Furthermore, 88 of the error readings are within the range of 1–2 MPa, while 88 is of 1 MPa. Note that maximal error frequencies really occur very seldom.

$$\begin{aligned}\%\text{CS-loss (MPa)} = {} & \left(\frac{\sqrt{-12.629 + (\mathrm{SP} - 12.629)}}{(\mathrm{EP})3.646\ )\times\left(\frac{\mathrm{SF}}{C}\right)}\right) \\ & + (10.379 + (-8.205 \\ & + (\sqrt{(\mathrm{EP} - \mathrm{SP}) + (\mathrm{SF} - \mathrm{SP})}))) \\ & + \left(\left(-8.082 + \frac{(\mathrm{S} - \mathrm{C})}{(\mathrm{SP} - \mathrm{C})}\right) + \sqrt{5.508}\right) \\ & + \left(\sqrt{\left(\mathrm{SP} + \left(\mathrm{EP} - \frac{\mathrm{GP}}{3.191}\right)\right)} - 6.598\right),\end{aligned} \tag{9}$$

where S represents sand, SP represents superplasticizer, W represents water, C represents cement, GP represents glass



(a)



(b)

Figure 6: The GEP model: (a) test-predicted %CS-loss correlation and (b) test-and-predicted %CS-loss error distributions.

powder, SF represents silica fume, and EP represents eggshell powder.

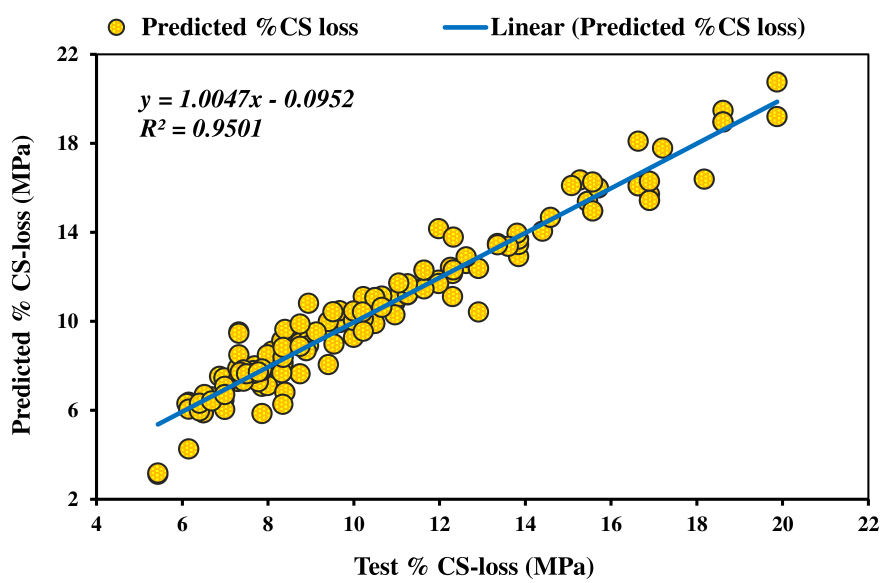
4.2 MEP model development

An empirical equation to determine the reduction in cementitious strength percentage in mortar modified with GP and EP was formulated based on an analysis of the MEP results, considering the impacts of the seven distinct components. The full set of mathematical equations used for the modeling are as follows (Eq. (10)):

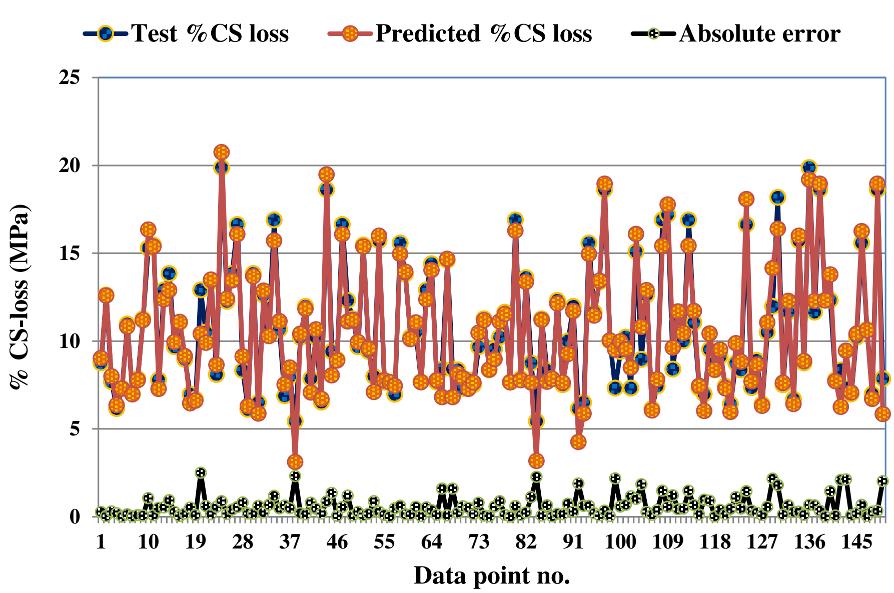
$$\%\text{CS-loss (MPa)} = \frac{\text{SP}(\sqrt{\text{S} + \text{W}} + \text{SF} + \text{EP})}{\text{C} - \text{SP} + 3\text{GP}}, \tag{10}$$

where S represents sand, SP represents superplasticizer, W represents water, C represents cement, GP represents glass powder, SF represents silica fume, and EP represents eggshell powder.

Figure 7 shows that the test results and the MEP prediction are correlated throughout the training and testing stages. An ideal regression line would have a slope close to 1. Figure 7(a) shows that the MEP model can handle oversimplification because it was well-trained; the testing data R^2 value was 0.950. Thus, the MEP model appears to be superior to the GEP model because of its higher R^2 value. The disparities between the actual and expected results of the MEP models are illustrated graphically in Figure 7(b). Evidence from the given data shows that the MEP prediction errors varied from a low of 0.010 MPa to a high of 2.505 MPa. An essential point to remember is that the maximum error of the GEP model occurs more often than the MEP-predicted outcome's maximum error. When it comes to making predictions, the MEP and GEP models are top-notch. The correlation coefficient and statistical error are both enhanced by the MEP equation. The simplicity and compactness of the MEP equation are the reasons for its many practical applications. Table 4 also shows the values of the respective statistical errors of the two models. The MEP model demonstrates



(a)



(b)

Figure 7: The MEP model: (a) association between the test and the predicted %CS-loss; (b) error distributions for both the test and the predicted %CS-loss.

Table 4: Statistics-based MEP and GEP model performance indicators

Parameters	MEP	GEP
MAE (MPa)	1.044	0.568
MAPE (%)	12.70	6.00
RMSE (MPa)	1.377	0.810
NSE	0.912	0.947
R	0.956	0.975
RSE	0.196	0.244
RRMSE (MPa)	0.226	0.434

superior performance compared to the GEP model, as it exhibits stronger correlation and lower error levels.

It is possible to further discuss, from different viewpoints, MEP outperforms GEP in every conceivable way in estimating the %CS-loss due to acid assault of GP and EP-based mortar. One positive feature is the MEP model's clarity and openness. For the purpose of calculating mortar's CS, MEP uses an equation that considers the additive impacts of each component. Due to the fact that it is simple to learn and interpret, this equation is useful in practical calculations. The GEP model, in contrast, relies on a complex nonlinear equation derived from human DNA. Because of its intricacy, the equation may be hard to understand and may not reveal any information about the relationships between the variables.

4.3 Statistical assessment of the models

The MEP model's statistical performance is also an important consideration. The MEP model successfully explains 95% of the detected variation in the %CS-loss of the mortar samples, as shown by the relatively high value of R^2 (coefficient of determination). The MEP model's improved R^2 value during validation further indicates its potential utility for forecasting data that have not yet been collected. When the R^2 value is high, the independent variables (sand, cement, water, silica fume, superplasticizer, GP, EP, and %CS-loss) exhibit strong correlations with the dependent variable (%CS-loss). In contrast to the GEP model, the MEP model provides more precise forecasts with a reduced RMSE. The MEP model's projected %CS-loss aligns more closely with the values observed in the mortar specimens due to lower values for RRMSE, MAPE, MAE, RSE, and RMSE. Table 4 demonstrates that the MEP model surpasses the GEP model in terms of prediction accuracy. This superiority is evident from considerably lower values of statistical parameters such as RMSE, MAE, MAPE, RRMSE, and RSE. Furthermore, Table 4 indicates that the MEP model achieves a higher NSE value compared to the GEP model, signifying its superior predictive accuracy. A high NSE indicates that the model is producing accurate predictions. The MEP model's correctness and usefulness can be evaluated using these statistical measures. Figure 8 uses violin plots to depict the distribution of errors (faults) in the MEP and GEP models. Violin plots combine box plots and density plots, showing the probability density of the data at different error levels, which helps in visualizing the spread and skewness of model errors. This provides a clear comparison of error distributions between the two models. The transition from GEP to MEP resulted in a considerable reduction in model errors.

For the purpose of predicting the percentage of CS loss in the GP and EP mortar samples, the MEP technique is an ideal modeling technique. This is due to the fact that it is user-friendly, it performs well in terms of mathematics, and it has the ability to incorporate the impacts of GP into a rectilinear equation. These discoveries may have applications in the real world, such as understanding how to change the components of GP and EP-modified mortar in the most effective manner in order to attain the required CS in construction projects. Furthermore, these findings open the door to the prospect of creating trustworthy prediction models for different kinds of modified mortar and concrete. These findings also make it possible to develop building methods that are more ecologically friendly and efficient.

4.4 SHAP results

Researchers in this study looked at how acid attacks affected GP and EP mortar and what parts of it were

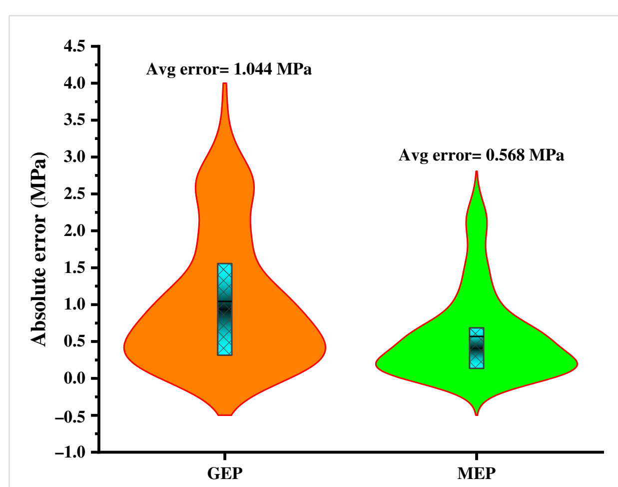


Figure 8: Violin plot as a means of displaying ML model errors.

responsible for those impacts. The SHAP tree explanation is used all across the world to help people better comprehend the local and global feature implications of SHAP. Different input features affect the acid attack %CS-loss of GP and EP mortar, as shown in Figure 9 of the SHAP diagram. The *x*-axis displays the proportion of the SHAP value attributable to each raw material, while the *y*-axis depicts the independent variables. The most important element, with a stronger positive correlation with the percentage of CS loss of mortar following an acid assault, was found to be the EP amount. Implying that incorporating EP increases the loss in CS with acid attack. Second, the GP amount was determined to be an important factor with a negative impact, suggesting that the incorporation of GP reduces the CS loss with acid. A stronger negative correlation between cement and the percentage of CS lost following an acid assault was observed, indicating that the control specimens (those lacking GP and EP) suffered a reduced loss of CS following the acid attack. The impact of sand was noted to be both positive and negative. Reduced data variability made it difficult to draw definitive findings regarding the effects of water, SF, and SP. The results might be more convincing if a bigger data set was used along with a wider range of input variables.

Figure 10 displays the various raw material contributions to the acid attack's weakening of GP and EP mortar. Figure 10(a) shows the EP impact and its interplay. At the reduced quantity of EP (up to 60 $kg \cdot m^{-3}$), the CS loss after the acid attack was less, while at the higher quantities of EP, the strength loss was higher. This might be due to the lower EP reactivity and dilution of cement at higher EP levels. The strength loss caused by acid assault was significantly decreased up to a GP level of 80 $kg \cdot m^{-3}$, as shown in Figure 10(b). The reduction in CS loss with GP usage might

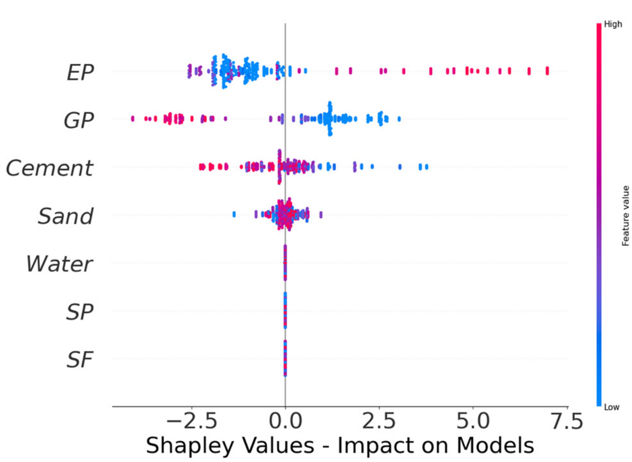


Figure 9: Impact of input parameters on CS-loss.

be attributed to the pozzolanic nature and finer particle size of GP, which made the matrix more dense and restricted the ingress of harmful ions. Figure 10(c)) implies the correlation of cement, exhibiting a decreasing trend with increasing cement quantity in the mix. The impact of sand was noted to be feasible at reduced quantities, as shown in Figure 10(d). Water, SP, and SF have a negligible impact on the loss of mortar strength, as shown in Figure 10(e)–(g), due to less variation in the provided parameters. The outcomes of the SHAP analysis were notably influenced by both the type of raw material and the size of the dataset under examination. The number of samples utilized and the input factors could affect the results.

5 Discussion

Worldwide, ordinary Portland cement is extensively used as the only binding material, diminishing raw materials [72] and emitting approximately 5–8% of global anthropogenic emissions [73]. In efforts to mitigate the release of CaO_2 by the OPC industry, identifying alternatives to OPC is paramount. Supplementary cementitious materials, including GP, EP, silica fume, fly ash, and rice husk ash, stand out as promising eco-friendly and energy-efficient construction materials. These materials have been partially employed to replace cement and sand in this respect [74]. Using ML and SHAP techniques, this research sought a deeper comprehension of the application of GP- and EP-modified cement mortar. To calculate the percentage of cement strength lost due to acid attack on GP and EP-modified mortar, this study employed GEP and MEP ML methods. By comparing their respective levels of accuracy, we were able to determine which strategy was the best predictor. The MEP approach yielded more accurate findings than the GEP technique, with an R^2 of 0.950 for % CS-loss as compared to the GEP-R^2 value of 0.913. The difference between the predicted and actual results (errors) is more evidence of the MEP method's superior accuracy. In order to determine how well the two datasets agree, error analysis is used to compare the MEP model's experimental and projected results with the GEP model. Table 5 displays the results of previous research that confirm the MEP technique outperforms the GEP method when it comes to estimating CBC strengths. The higher accuracy of the MEP model compared to the GEP model can be attributed to its enhanced ability to capture complex relationships within the data. MEP utilizes a multi-expression approach, allowing it to generate multiple solutions simultaneously and select the best-performing one. This flexibility results in a more robust and accurate model. Additionally, the MEP structure helps in avoiding overfitting

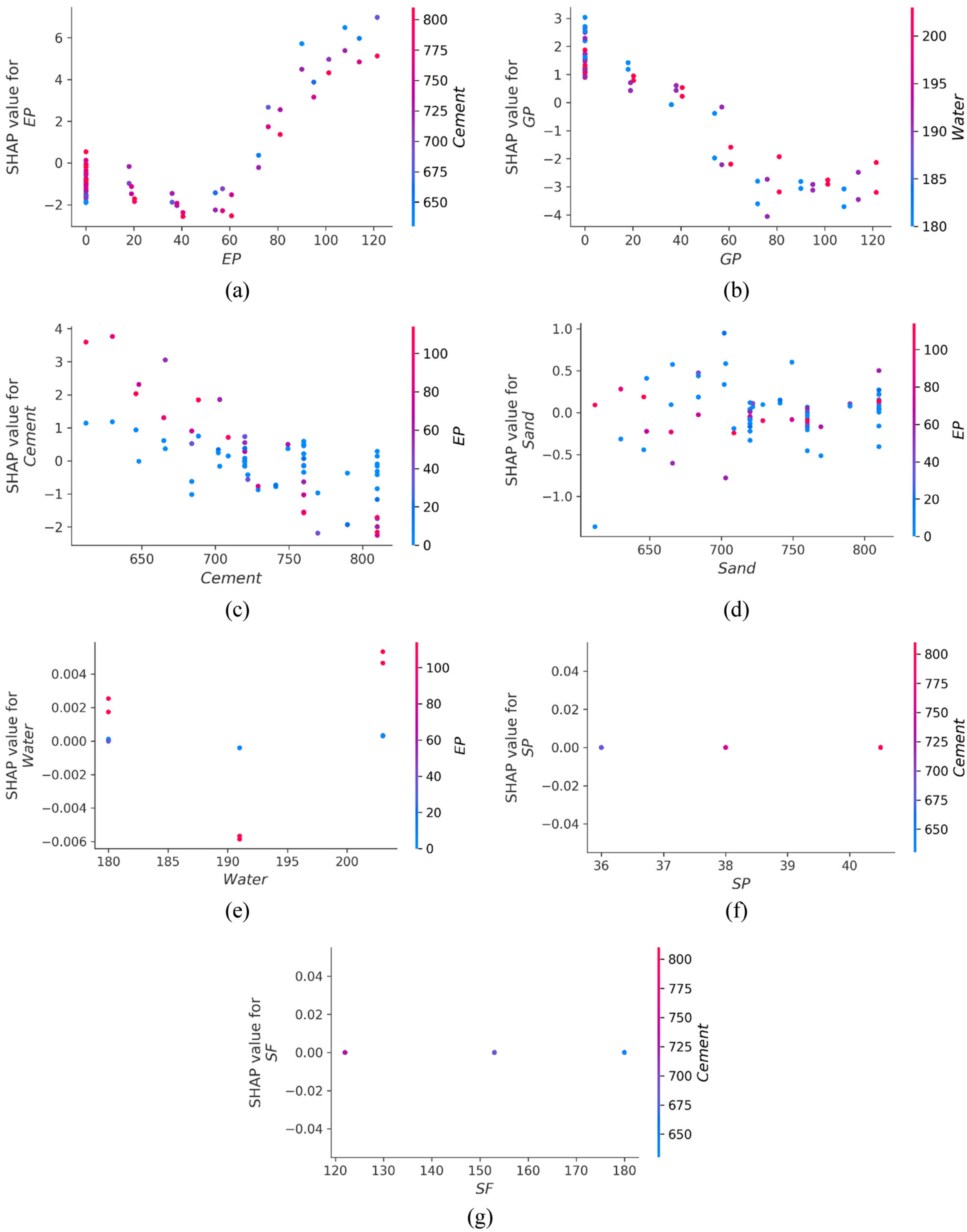


Figure 10: Interaction of input parameters for the %CS-loss of GP and EP mortar: (a) EP, (b) GP), (c) cement, (d) sand, (e) water, (f) SP, and (g) SF.

Table 5: Previous techniques used for modeling

Ref.	Method	Material studied	Attribute investigated	Most effective model (R^2-value)
Present study	GEP and MEP	GP and EP-based mortar	%CS loss after acid attack	MEP (R^2 = 0.950)
[37]	GEP and MEP	FA-based geopolymer concrete	CS	MEP (0.97)
[26]	MARS, MEP, and ANN	Mixture of sand and cement, including metakaolin clay	CS	MEP (0.96)
[75]	GEP, MEP	Eco-friendly sand paver bricks made of plastic	CS	MEP (0.91)
[76]	MLPNN, MEP, and ANFIS.	Viscose-based eco-friendly pavement	CS, STS (split tensile strength)	MEP (0.93)

by balancing model complexity and prediction accuracy more effectively than GEP.

Statistical approaches were also used to evaluate the accuracy of the ML methods. Model accuracy is directly proportional to the magnitude of the R^2 value and the size of the deviations (MAE, RMSE, MAPE, *etc.*). The performance of algorithms for attribute forecasting across various study topics is heavily influenced by the number of inputs and data samples utilized, making it challenging to determine the ideal ML technique. To compare the predictions of the two models, several statistical tests were employed, including root mean squared error, mean percentage error, root mean squared relative error, and MAE. When compared to GEP, the data demonstrated that MEP was far more accurate. To further investigate the interplay between the constituent materials and their impact on GP- and EP-modified mortar's CS, an SHAP analysis was conducted. Due to the high link between mortar strength loss from acid attack and EP and GP, these input characteristics were determined to be significant, implying their use in optimal limits.

The fact that the GEP and MEP models can be made to function with only seven inputs is the source of the value that both of these models possess. Because of this characteristic, it is guaranteed that the forecasts that are derived are unique to the application of GP and EP in CBCs to be accurate. Since all models utilize consistent unit measurements and rely on the same testing technique, the output forecasts they generate are considered reliable. The mathematical equations provided by the models facilitate a better understanding of mix design and the impact of each input parameter. On the other hand, once the initial seven inputs have been taken into consideration, the incorporation of additional parameters into the composite analysis may have an impact on the applicability of the anticipated models. It is likely that the models that were developed will not work successfully when they are confronted with unexpected inputs. This is due to the fact that the models were calibrated to cope with a specific collection of data as they were being generated. Deviation from consistent units or alterations in input parameters could lead to inaccuracies in the results of the predictive models. It is critical that the units used stay constant if we wish to think of the models as helpful.

6 Conclusions

A research investigation was undertaken to explore the impacts of recycled GP and EP on acid-affected cement mortar, employing ML models. Two ML models, GEP and MEP, used experimental data to predict the percentage of CS loss in acid-attacked GP- and EP-based cement mortar. The following are important conclusions of the research:

1) The GEP approach provided sufficient precision (R^2 = 0.931), whereas the MEP method had greater precision (R^2 = 0.950) for estimating %CS-loss.
2) On average, 1.044 and 0.568 MPa, respectively, separated the experimental test from the predicted CS (errors) in the GEP and MEP methods. These error statistics further demonstrated that the MEP technique was more precise than the GEP models in forecasting the %CS-loss of GP and EP-modified mortar.
3) Statistical validation confirmed that the models used were effective. The accuracy of ML models was demonstrated by lower errors and better R^2. The MAPE for % CS-loss prediction was 12.70% in the GEP model and 6.00% in the MEP model. Likewise, the RMSE values were 1.377 MPa for the GEP model and 0.810 MPa for the MEP model. Statistical results indicate that the MEP model outperformed the GEP model in predicting the percentage of CS loss in GP and EP mortar.
4) It was found from the SHAP results that EP and GP quantities were the most influential factors with positive and negative correlations, respectively, followed by cement with negative and sand with both positive and negative correlations.
5) The SHAP interaction plots showed that increasing the EP quantity up to 60 $kg{\cdot}m^{-3}$ and GP quantity up to 80 $kg{\cdot}m^{-3}$ exhibited a greater resistance to the acid attack.

The reason that GEP and MEP are so important is that they provide a novel mathematical expression for predicting outcomes by varying the values of input parameters. In order to facilitate rapid evaluation, improvement, and justification of mortar mixture proportioning, mathematical models that have been derived from this study can be applied by specialists in the fields of science and engineering.

Acknowledgments: The authors acknowledge the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia (Grant No. KFU241289). The authors are thankful to the Deanship of Graduate Studies and Scientific Research at Najran University for funding this work under the Growth Funding Program grant code NU/GP/SERC/13/109-2. The authors extend their appreciation for the financial support that made this study possible.

Funding information: This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia (Grant No. KFU241289) and the Deanship of Graduate Studies and Scientific Research at Najran University for funding this work under the Growth Funding Program grant code NU/GP/SERC/13/109-2.

Author contributions: H.L.: conceptualization, formal analysis, supervision, visualization, and writing-original draft. S.A.K.: supervision, data acquisition, software, methodology, writing-original draft, writing, reviewing, and editing. M.N.A.: funding acquisition, investigation, project administration, writing, reviewing, and editing. F.A.: formal analysis, validation, writing, reviewing, and editing. M.T.Q.: investigation, resources, visualization, 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 and/ or analyzed during the current study are available from the corresponding author upon reasonable request.

References

[1] Bai, W., J. Shen, J. Guan, J. Wang, and C. Yuan. Study on compressive mechanical properties of recycled aggregate concrete with silica fume at different strain rates. *Materials Today Communications*, Vol. 31, 2022, id. 103444.

[2] Cao, Y., J. Bao, P. Zhang, Y. Sun, and Y. Cui. A state-of-the-art review on the durability of seawater coral aggregate concrete exposed to marine environment. *Journal of Building Engineering*, Vol. 60, 2022, id. 105199.

[3] Jin, C., Y. Qian, K. Khan, A. Ahmad, M. N. Amin, F. Althoey, et al. Predicting the damage to cementitious composites due to acid attack and evaluating the effectiveness of eggshell powder using interpretable artificial intelligence models. *Materials Today Communications*, Vol. 37, 2023, id. 107333.

[4] Bahraq, A. A., J. Jose, M. Shameem, and M. Maslehuddin. A review on treatment techniques to improve the durability of recycled aggregate concrete: Enhancement mechanisms, performance and cost analysis. *Journal of Building Engineering*, Vol. 55, 2022, id. 104713.

[5] Baloch, W. L., H. Siad, M. Lachemi, and M. Sahmaran. A review on the durability of concrete-to-concrete bond in recent rehabilitated structures. *Journal of Building Engineering*, Vol. 44, 2021, id. 103315.

[6] Hill, R. L., A. J. Boyd, J. E. Dongell, G. Hichborn, R. E. Neal, C. K. Nmai, et al. *Guide to Durable Concrete Reported by ACI Committee*, Vol. 201, 2008.

[7] Sahoo, S., T. R. Mahapatra, N. Priyadarshini, S. Mahapatra, S. Naik, and S. Jaypuria. Influence of water binder ratio on strength and acid resistance of concrete made up of mineral admixture as supplementary cementitious material. *Materials Today: Proceedings*, Vol. 26, 2020, pp. 796–803.
[8] Nnaemeka, O. F. and N. B. Singh. Durability properties of geopolymer concrete made from fly ash in presence of Kaolin. *Materials Today: Proceedings*, Vol. 29, 2020, pp. 781–784.
[9] Afroughsabet, V. and T. Ozbakkaloglu. Mechanical and durability properties of high-strength concrete containing steel and polypropylene fibers. *Construction and Building Materials*, Vol. 94, 2015, pp. 73–82.
[10] Aliyu, I., T. A. Sulaiman, A. Mohammed, and J. M. Kaura. Effect of sulphuric acid on the compressive strength of concrete with quarry dust as partial replacement of fine aggregate. *FUDMA Journal of Sciences*, Vol. 4, 2020, pp. 553–559.
[11] Valencia-Saavedra, W. G. and R. Mejía de Gutiérrez. Resistance to chemical attack of hybrid fly ash-based alkali-activated concretes. *Molecules*, Vol. 25, 2020, id. 3389.
[12] Alsharari, F., K. Khan, M. N. Amin, W. Ahmad, U. Khan, M. Mutnbak, et al. Sustainable use of waste eggshells in cementitious materials: An experimental and modeling-based study. *Case Studies in Construction Materials*, Vol. 17, 2022, id. e01620.
[13] Khan, M. and M. Ali. Improvement in concrete behavior with fly ash, silica-fume and coconut fibres. *Construction and Building Materials*, Vol. 203, 2019, pp. 174–187.
[14] Abdalla, A. A. and A. Salih Mohammed. Theoretical models to evaluate the effect of SiO2 and CaO contents on the long-term compressive strength of cement mortar modified with cement kiln dust (CKD). *Archives of Civil and Mechanical Engineering*, Vol. 22, 2022, id. 105.
[15] Abdalla, A. A., A. S. Mohammed, S. Rafiq, R. Noaman, W. S. Qadir, K. Ghafor, et al. Microstructure, chemical compositions, and soft computing models to evaluate the influence of silicon dioxide and calcium oxide on the compressive strength of cement mortar modified with cement kiln dust. *Construction and Building Materials*, Vol. 341, 2022, id. 127668.
[16] Lao, J.-C., B.-T. Huang, L.-Y. Xu, M. Khan, Y. Fang, and J.-G. Dai. Seawater sea-sand Engineered Geopolymer Composites (EGC) with high strength and high ductility. *Cement and Concrete Composites*, Vol. 138, 2023, id. 104998.
[17] Hamada, H., A. Alattar, B. Tayeh, F. Yahaya, and A. Adesina. Sustainable application of coal bottom ash as fine aggregates in concrete: A comprehensive review. *Case Studies in Construction Materials*, Vol. 16, 2022, id. e01109.
[18] Almeshal, I., M. M. Al-Tayeb, S. M. A. Qaidi, B. H. A. Bakar, and B. A. Tayeh. Mechanical properties of eco-friendly cements-based glass powder in aggressive medium. *Materials Today: Proceedings*, Vol. 58, 2022, pp. 1582–1587.
[19] Yang, D., J. Zhao, W. Ahmad, M. Nasir Amin, F. Aslam, K. Khan, et al. Potential use of waste eggshells in cement-based materials: A bibliographic analysis and review of the material properties. *Construction and Building Materials*, Vol. 344, 2022, id. 128143.
[20] Qin, D., Y. Hu, and X. Li. Waste Glass utilization in cement-based materials for sustainable construction: a review. *Crystals*, Vol. 11, 2021, id. 710.
[21] Amin, M. N., W. Ahmad, K. Khan, M. N. Al-Hashem, A. F. Deifalla, and A. Ahmad. Testing and modeling methods to experiment the flexural performance of cement mortar modified with eggshell powder. *Case Studies in Construction Materials*, Vol. 18, 2023, id. e01759.
[22] Jiang, Y., T.-C. Ling, K. H. Mo, and C. Shi. A critical review of waste glass powder–Multiple roles of utilization in cement-based materials and construction products. *Journal of environmental management*, Vol. 242, 2019, pp. 440–449.
[23] Amin, M. N., W. Ahmad, K. Khan, S. Nazar, A. M. A. Arab, and A. F. Deifalla. Evaluating the relevance of eggshell and glass powder for cement-based materials using machine learning and SHapley Additive exPlanations (SHAP) analysis. *Case Studies in Construction Materials*, Vol. 19, 2023, id. e02278.
[24] Emad, W., A. S. Mohammed, A. Bras, P. G. Asteris, R. Kurda, Z. Muhammed, et al. Metamodel techniques to estimate the compressive strength of UHPFRC using various mix proportions and a high range of curing temperatures. *Construction and Building Materials*, Vol. 349, 2022, id. 128737.
[25] Deifalla, A. Refining the torsion design of fibered concrete beams reinforced with FRP using multi-variable non-linear regression analysis for experimental results. *Engineering Structures*, Vol. 226, 2021, id. 111394.
[26] 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.
[27] Abdalla, A. A. and A. S. Mohammed. The efficiency of hybrid intelligent models to evaluate the effect of the size of sand and clay metakaolin content on various compressive strength ranges of cement mortar. *Neural Computing and Applications*, Vol. 36, 2024, pp. 1–21.
[28] Güçlüer, K., A. Özbeyaz, S. Göymen, and O. Günaydın. A comparative investigation using machine learning methods for concrete compressive strength estimation. *Materials Today Communications*, Vol. 27, 2021, id. 102278.
[29] Jiao, H., Y. Wang, L. Li, K. Arif, F. Farooq, and A. Alaskar. A novel approach in forecasting compressive strength of concrete with carbon nanotubes as nanomaterials. *Materials Today Communications*, Vol. 35, 2023, id. 106335.
[30] Akhtar, M. and A. Khajuria. The synergistic effects of boron and impression creep testing during paced controlling of temperature for P91 Steels. *Advanced Engineering Materials*, Vol. 25, 2023, id. 2300053.
[31] Young, B. A., A. Hall, L. Pilon, P. Gupta, and G. Sant. Can the compressive strength of concrete be estimated from knowledge of the mixture proportions?: New insights from statistical analysis and machine learning methods. *Cement and Concrete Research*, Vol. 115, 2019, pp. 379–388.
[32] de Melo, V. V. and W. Banzhaf. Improving the prediction of material properties of concrete using Kaizen Programming with Simulated Annealing. *Neurocomputing*, Vol. 246, 2017, pp. 25–44.
[33] Sun, J., Y. Wang, X. Yao, Z. Ren, G. Zhang, C. Zhang, et al. Machine-learning-aided prediction of flexural strength and ASR expansion for waste glass cementitious composite. *Applied Sciences*, Vol. 11, 2021, id. 6686.
[34] Alkadhim, H. A., M. N. Amin, W. Ahmad, K. Khan, S. Nazar, M. I. Faraz, et al. Evaluating the strength and impact of raw ingredients of cement mortar incorporating waste glass powder using machine learning and shapley additive explanations (SHAP) methods. *Materials*, Vol. 15, 2022, id. 7344.
[35] 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.
[36] Amin, M. N., H. A. Alkadhim, W. Ahmad, K. Khan, H. Alabduljabbar, and A. Mohamed. Experimental and machine learning approaches to investigate the effect of waste glass powder on the flexural strength of cement mortar. *PloS One*, Vol. 18, 2023, id. e0280761.
[37] Chu, H.-H., M. A. Khan, M. Javed, A. Zafar, M. I. Khan, H. Alabduljabbar, et al. Sustainable use of fly-ash: Use of gene-expression programming (GEP) and multi-expression programming (MEP) for forecasting the compressive strength geopolymer concrete. *Ain Shams Engineering Journal*, Vol. 12, 2021, pp. 3603–3617.
[38] Ma, H., J. Liu, J. Zhang, and J. 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, pp. 1–10.
[39] Faraz, M. I., S. U. Arifeen, M. N. 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.
[40] Holland, J. H. *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence*, MIT press, Cambridge, Massachusetts, United States, 1992.
[41] Koza, J. R. Human-competitive results produced by genetic programming. *Genetic Programming and Evolvable Machines*, Vol. 11, 2010, pp. 251–284.
[42] Gholampour, A., T. Ozbakkaloglu, and R. Hassanli. Behavior of rubberized concrete under active confinement. *Construction and Building Materials*, Vol. 138, 2017, pp. 372–382.
[43] 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.
[44] Bagheri, A., A. Nazari, and J. Sanjayan. The use of machine learning in boron-based geopolymers: Function approximation of compressive strength by ANN and GP. *Measurement*, Vol. 141, 2019, pp. 241–249.
[45] Koza, J. R. Genetic programming as a means for programming computers by natural selection. *Statistics and computing*, Vol. 4, 1994, pp. 87–112.
[46] Ferreira, C. *Gene expression programming: Mathematical modeling by an artificial intelligence*, Vol. 21, Springer, Berlin, Germany, 2006.
[47] 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.
[48] 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.
[49] Nazar, S., J. Yang, M. F. Javed, K. Khan, L. Li, and Q. F. Liu. An evolutionary machine learning-based model to estimate the rheological parameters of fresh concrete, *Structures*, Vol. 48, 2023, pp. 1670–1683.
[50] 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.
[51] Oltean, M. and C. Grosan. A comparison of several linear genetic programming techniques. *Complex Systems*, Vol. 14, 2003, pp. 285–314.
[52] 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.
[53] 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.
[54] Grosan, C. and A. Abraham. Stock market modeling using genetic programming ensembles. *Genetic Systems Programming: Theory and Experiences*, Vol. 13, 2006, pp. 131–146.
[55] Oltean, M. and D. Dumitrescu. Multi expression programming. *Journal of Genetic Programming and Evolvable Machines*, 2002.
[56] Alfaiad, M. A., K. Khan, W. Ahmad, M. N. Amin, A. F. Deifalla, and N. A. Ghamry. Evaluating the compressive strength of glass powder-based cement mortar subjected to the acidic environment using testing and modeling approaches. *PloS One*, Vol. 18, 2023, id. e0284761.
[57] Wang, N., Z. Xia, M. N. Amin, W. Ahmad, K. Khan, F. Althoey, et al. Sustainable strategy of eggshell waste usage in cementitious composites: An integral testing and computational study for compressive behavior in aggressive environment. *Construction and Building Materials*, Vol. 386, 2023, id. 131536.
[58] Chen, Z. Application of machine learning boosting and bagging methods to predict compressive and flexural strength of marble cement mortar. *Materials Today Communications*, Vol. 39, 2024, id. 108600.
[59] Frank, I. E. and R. Todeschini. *The Data Analysis Handbook*, Elsevier, The Netherlands, 1994.
[60] Sufian, M., S. Ullah, K. A. Ostrowski, A. Ahmad, A. Zia, K. Śliwa-Wieczorek, et al. An experimental and empirical study on the use of waste marble powder in construction material. *Materials*, Vol. 14, 2021, id. 3829.
[61] Khan, M., J. Lao, and J.-G. Dai. Comparative study of advanced computational techniques for estimating the compressive strength of UHPC. *Journal of Asian Concrete Federation*, Vol. 8, 2022, pp. 51–68.
[62] 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.
[63] Ferreira, C. Genetic representation and genetic neutrality in gene expression programming. *Advances in Complex Systems*, Vol. 5, 2002, pp. 389–408.
[64] Hanandeh, S., A. Ardah, and M. Abu-Farsakh. Using artificial neural network and genetics algorithm to estimate the resilient modulus for stabilized subgrade and propose new empirical formula. *Transportation Geotechnics*, Vol. 24, 2020, id. 100358.
[65] 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.
[66] Soleimani, S., S. Rajaei, P. Jiao, A. Sabz, and S. Soheilinia. New prediction models for unconfined compressive strength of geopolymer stabilized soil using multi-gen genetic programming. *Measurement*, Vol. 113, 2018, pp. 99–107.

[67] 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.

[68] 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.

[69] 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.

[70] 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 and Nano-Objects*, Vol. 17, 2019, pp. 103–111.

[71] Shahin, M. A. Use of evolutionary computing for modelling some complex problems in geotechnical engineering. *Geomechanics and Geoengineering*, Vol. 10, 2015, pp. 109–125.

[72] Bond, J. E., R. Coursaux, and R. L. Worthington. Blending systems and control technologies for cement raw materials. *IEEE Industry Applications Magazine*, Vol. 6, 2000, pp. 49–59.

[73] Worrell, E., L. Price, N. Martin, C. Hendriks, and L. O. Meida. Carbon dioxide emissions from the global cement industry. *Annual Review of Energy and the Environment*, Vol. 26, 2001, pp. 303–329.

[74] Juenger, M. C. G., R. Snellings, and S. A. Bernal. Supplementary cementitious materials: New sources, characterization, and performance insights. *Cement and Concrete Research*, Vol. 122, 2019, pp. 257–273.

[75] Iftikhar, B., S. C. Alih, 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.

[76] Nafees, A., M. F. Javed, S. Khan, K. Nazir, F. Farooq, F. Aslam, et al. Predictive modeling of mechanical properties of silica fume-based green concrete using artificial intelligence approaches: MLPNN, ANFIS, and GEP. *Materials* 2021, Vol. 14, id. 7531.