Case Studies in Construction Materials 19 (2023) e02172

Contents lists available at ScienceDirect

Case Studies in Construction Materials

journal homepage: www.elsevier.com/locate/cscm

Unveiling the potential of an evolutionary approach for accurate compressive strength prediction of engineered cementitious composites

Amirhossein Davarpanah T.Q. [a], Amir R. Masoodi [a,*], Amir H. Gandomi [b,c,**]

[a] Department of Civil Engineering, Ferdowsi University of Mashhad, Iran
[b] Faculty of Engineering and IT, University of Technology Sydney, NSW 2007, Australia
[c] University Research and Innovation Center (EKIK), Óbuda University, 1034, Budapest, Hungary

ARTICLE INFO

Keywords:
Engineered cementitious composite (ECC)
Fly ash
Polyvinyl alcohol
Compressive strength
Gene expression programming

ABSTRACT

The different human activities in numerous fields of civil engineering have become possible due to recent development in soft computing. As many researchers have widely extended the use of evolutionary numerical methods to predict the mechanical properties of construction materials, it has become necessary to investigate the performance, accuracy, and robustness of these approaches. Gene Expression Programming (GEP) is a method that stands out among these methods as it can generate highly accurate formulas. In this study, two models of GEP are used to anticipate the compressive strength of engineered cementitious composite (ECC) containing fly ash (FA) and polyvinyl alcohol (PVA) fiber at 28 days. The experimental results for 76 specimens, which are made with ten different mixture properties, are taken from the literature to build the models. Considering the experimental results, four different input variables in the GEP approach are used to arrange the models in two modes: sorted data distribution (SDD) and random data distribution (RDD). Prognosticating the compressive strength values based on the mechanical properties of ECC containing FA and PVA will be possible for the models of the GEP method by using these input variables. The comparison between the experimental results and the results of training, testing, and validation sets of two models (GEP-I and GEP-II), each of which has two distinct distribution modes, is done. It is observed that both modes of RDD and SDD lead to responses with the same accuracy (R-square more than 0.9). Nevertheless, the GEP-I (SDD) model was chosen as the best model in this study based on its performance with the validation data set.

1. Introduction

In the last couple of decades, there has been a lot of achievements in constructing various concretes for different technical applications. In this regard, scientists have done different experiments on the mechanical properties, durability, and fracture parameters of these concretes. One of the concrete materials made in 1992 was engineered cementitious composite (ECC) [1]. ECC was made to increase the ductility and toughness of the concrete. Therefore, it is a suitable material for structural components that undergo repetitive and cyclic loading, such as beam-column joints. It is also an appropriate option for structures such as bridges that are exposed

* Corresponding author.
** Correspondence to: Faculty of Engineering & Information Technology, University of Technology Sydney, Australia.
E-mail addresses: ar.masoodi@um.ac.ir (A.R. Masoodi), gandomi@uts.edu.au (A.H. Gandomi).

https://doi.org/10.1016/j.cscm.2023.e02172
Received 5 March 2023; Received in revised form 22 May 2023; Accepted 27 May 2023
Available online 29 May 2023

to axial and bending stresses, weather effects, and potential damage from cracks, unreasonable expansion, or reinforcement corrosion. Failing to consider these factors could lead to costly repairs, the replacement of components, or even the entire bridge structure [2,3]. It should be noted that although high or ultra-high-performance concrete may be a better choice rather than ECC, the application of ECC and High-Performance Concrete (HPC) or Ultra-High-Performance Concrete (UHPC) differs from each other. ECC is primarily intended for structural elements that are subject to repetitive and cyclic loading. Hence, its ductility and toughness are crucial factors [4]. Although tensile capacity and ductility could be concerning for ECC materials, it is worth mentioning that the compressive strength of concrete is considered the most crucial factor in evaluating its quality. Therefore, researchers conduct this test on all their concrete samples, regardless of their type. Most studies have focused on increasing the compressive strength of the concrete mixture, as this parameter is seen as a measure of the quality of the concrete. However, this can lead to an increase in brittle failure in critical regions of structures, for example, in composite steel structural joints which are susceptible to concrete fracture [5–7]. Fundamentally, ECC can drastically cover both important factors of tensile strength and compressive strength in basic cementitious materials, because of the fibers in its structure [8]. Moreover, ECC can be used in the regions of the structure that are subjected to extremely repetitive, cyclic, axial, and bending loading [9].

Various mining materials like FA, silica fume (SF), metakaolin (MK), natural pozzolan (NP), and ground granulated blast furnace slag (GGBFS) have been studied by researchers on ECC [5,8,10–14]. Using these substitution materials in the mixtures of ECC is aimed at decreasing the amount of cement used in these compositions. Arivusudar and Suresh Babu conducted research that demonstrated that adding up to 40% SF to ECC mixtures containing FA instead of cement significantly improved the material's mechanical properties [15]. Boya et al. found that substituting Class F fly ash for partial or complete amounts of slag in ECC mixtures resulted in significant improvements in both the mechanical and durability properties of the mixture [2]. In contrast, most experimental studies on the effect of FA have focused on Class F's impact on the mechanical properties of ECC. Furthermore, it improves the durability, mixture workability, and materials' sustainability [16]. Additionally, over-substitution of FA causes a reduction in concrete strength in the early ages. On the other hand, most of the mixtures used in experimental studies have usually been composed of PVA fibers, at a volume percentage of 2 [17–21], due to their superior performance in ECC than the other fibers [22].

Nowadays, many researchers are utilizing a variety of numerical methods to reduce the use of building materials and the costs of experimental work. Using several software programs, they have conducted a numerical study of fracture parameters, mechanical properties, and durability characteristics of concrete. In the meantime, optimization methods based on artificial intelligence (AI), such as artificial neural networks (ANNs) [23–30], adaptive neuro-fuzzy inference systems (ANFIS) [31–33], and multiple linear regression (MLR) [34–37] have attracted research attention. Asteris and Mokos used neural networks to build models that could predict the compressive strength of concrete with an R-square (R^2) value higher than 0.9 [38]. They based their models' input parameters on concrete's non-destructive testing results, such as ultrasonic pulse velocity and Schmidt rebound hammer. Khademi et al. studied the prediction of the 28-day compressive strength of recycled aggregate concrete using different methods, including ANN, ANFIS, and MLR [39]. The researchers found that the models created using the ANN method were better than those created using the ANFIS method. However, the MLR method was not effective and produced unsatisfactory results with an R^2 value of 0.6. This indicated that linear relationships were not suitable for predicting the research outcomes. The report's authors attribute this to the non-linear correlations between the variables, which the MLR method cannot model. Altun et al. showed in their study on predicting the compressive strength of lightweight concrete containing steel fibers that the ANN method was more consistent with the experimental output variables than the MLR method [40].

Nevertheless, one of the drawbacks of these optimization methods has been the failure to provide an application relationship as an output of the analysis of the experimental results. Consequently, Scientists have begun to develop non-gradient methods in the field of intelligence research (IR). Many researchers in the field of IR have started to use the branch of meta-heuristic algorithm methods, including evolutionary computation [41–48], ant colony optimization algorithms (ACOA) [49–51], and artificial bees colony algorithm (ABCA) [52–56]. It should be noted that the choice of either method is dependent on the type of existing data and the condition of the issue. Meanwhile, most researchers have utilized the subsets of evolutionary computing called gene expression programming (GEP). The reason for this is its ability to provide high-accuracy formulas [57]. In this regard, a formulation for the viscoelastic behaviour of modified asphalt binder by utilization of GEP was presented by Gandomi et al. [58]. Algaifi et al. conducted research to predict the compressive strength of bacterial concrete using GEP modeling [59]. They compared the output data of the models they created with experimental results and concluded that the formulas produced by the GEP method were reliable and accurate. Alabduljabbar et al., in a separate study, used the GEP method to model the compressive strength of high-performance concrete (HPC)[60]. They analyzed 810 experimental data points and 15 input variables, and the results showed that the constructed GEP models were robust and reliable. Furthermore, a lot has been carried out in terms of predicted models according to GEP in the context of Modulus of Elasticity, Shear strength, Tensile strength, Compressive strength, and Flexural strength in various concretes [59,61–74]. In general, GEP is an algorithm that can be used for many different problems in various fields like engineering, finance, medicine, and biology. It can create complex solutions, which makes it useful for solving real-world problems.

According to existing literature and the lack of a direct relationship to anticipate the compressive strength of Engineered Cementitious Composite (ECC), authors have started to develop an experimental model using Gene Expression Programming (GEP) to predict the compressive strength of ECC, including FA and PVA fiber. For this purpose, utilize the compiled database to develop a model that is efficient, reliable, and robust, and that best matches the target parameter of the experimental data (f_c). After conducting experimental studies on the measurement of the mechanical properties of ECC, containing FA of F class and PVA fiber, the results of 28-day samples were collected. 76 data were collected from 10 different mix designs. These data were partitioned into the training and testing sets of the models for the sorted data distribution (SDD) mode, with a specific fit. On the other hand, this division was completely random for the mode of random data distribution (RDD) of the models. The set of models consisted of two numbers with

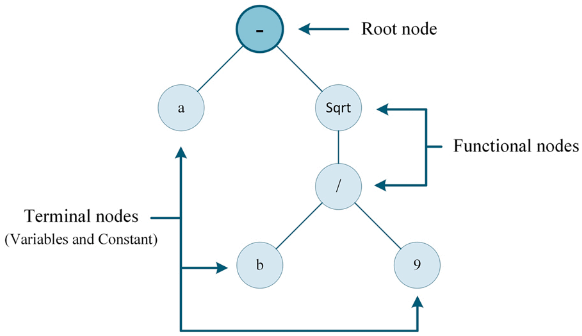

Fig. 1. The Parse Tree representation of a GP model [-a (sqrt (/b 9))].

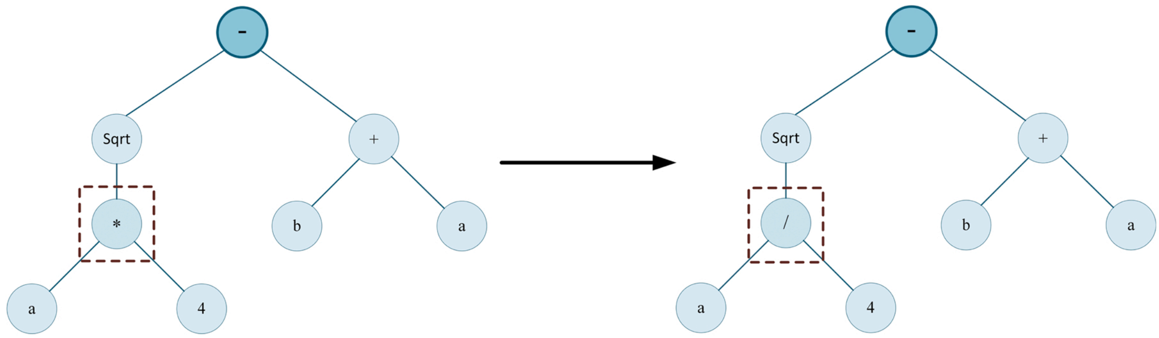

Fig. 2. Example of genetic programming cross-over.

different genes and distinct data distribution. The volume fraction of fiber (V_f), the length of fiber (L_f), the ratio of water to binder powder (W/B), and the ratio of binder powder to aggregates (B/A) are included in the intended variables for both models. The value of compressive strength (f_c) was considered as the output. All models were trained using 49 experimental results. The remaining findings were then used as experimental inputs for testing. Moreover, for validating the models, 7 experimental data results were utilized. The findings of experimental results were the same as the models' findings in three sets of training, testing, and validation.

2. Mathematical procedure

2.1. Definition of GP

Genetic programming (GP) approach was introduced by Koza [75]. In fact, this method has been suggested as the development of Genetic Algorithm (GA). Structural language of chromosomes shows the significant difference between GA and GP. This means that GA shows a linear string of 0 and 1 with a fixed length as the chromosome structure. Though the development proposed by Holland [76] was discussed in the subsequent writings, and real numbers were used in the chromosome structure. GP structures of computer programs are based on the LISP language. The components of the structural language of this program include terminals and functions. It should be noted that the selection of functions and constituents of terminals, including a set of input variables and constants, is performed indiscriminately. Then, they combine and a computer model is made into a tree-like structure. Representation of this format is called Parse Trees (PTs). It starts with a root node at the beginning. Thereafter, it is made of branches that are extended from each function and are ended to a variable or constant. For example, this is how the LISP language defines a two-variable program (-a (sqrt (/b 9)))). Fig. 1 indicates the tree structure of this example. In general, the structural language of chromosomes in GA is simpler than GP and has evolutionary limitations. For this reason, functional diversity in GP, which can be considered all kinds of variables and functions in the formation of the chromosome structure, cannot be expected in GA.

GP acts with the population of chromosomes (computer programs) that are accidentally generated. After the initial population generation, GP evolves through 3 main genetic operators of reproduction, cross-over, and mutation. Then, the cross-over genetic operator randomly chooses parts of two parents and moves them together, Fig. 2. Developing fitness for the next generation is the reason. One node is chosen by accident and then mutated during the mutation operator phase, Fig. 3 [77]. Furthermore, the mutation protects the model from the achievement of premature convergence. It involves expanding the other non-local features of the search algorithm.

Although the issue of evolutionary limitation and simple language structure in GA was resolved in GP, the dual role limitation of chromosomes has remained. This means that parse trees concurrently play the role of genotype and phenotype. Therefore, this feature

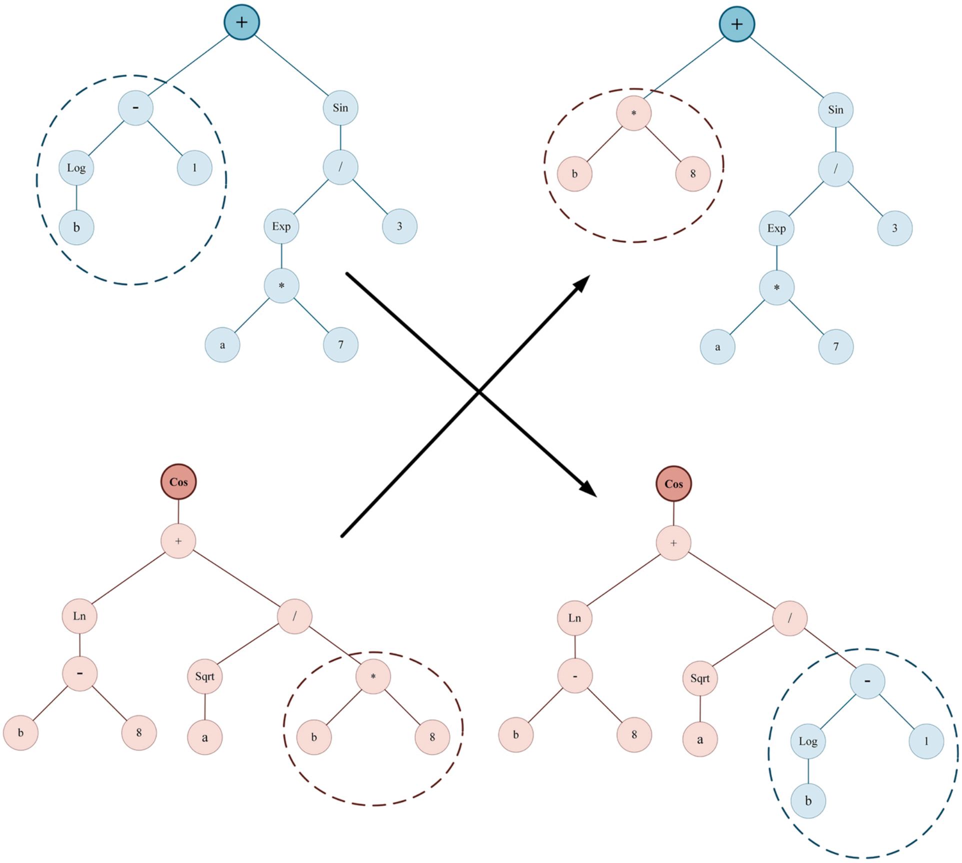

Fig. 3. Example of genetic programming mutation.

limits the operator implementation so that PTs do not provide correct results. In this respect, GEP offers a solution to remove these limitations.

2.2. Defining the GEP

Ferreira created the GEP algorithm in 1991 and introduced it formally in 2001 [78]. In order to eliminate the shortcomings, two important evolutionary viewpoints, GA and GP are used to reach GEP. The same as the method of GA, the chromosome genotype has been coded as linear strings of fixed size (genome). The different length and size of a tree structure of chromosome phenotype is similar to the GP algorithm. The structures of these trees in GEP are called expression trees (ETs). Thus, the dual role limitation of chromosomes in the last 2 perspectives is solved by this method in GEP. Using multiple genetic operators, GEP checks the permanent health of offspring chromosomes faster than GP. Hence, Ferreira reported that the GEP algorithm has successfully passed the natural evolutionary processes' first and second thresholds (replicator and phenotype thresholds).

Generally, one or more genes are included in the structure of chromosomes in the GEP algorithm. A head and a tail are two important parts of each gene. Additionally, these parts are placed in a certain form and length behind each other. For this reason, a gene is always valid and terminates in a correct ET. Subsequently, this method covers the weakness that exists in GP. The head includes both terminals and functions and the tail contains only terminals. Basic arithmetic operations (+ / * -), Boolean logic operators (e.g. AND, OR, IF), and algebraic functions, such as trigonometric and exponential functions, include the functions that may be used for this position. It should be noted that the user chooses the functions, and a logical relationship between input and output variables is implied. As well, terminals include a set of input variables and constants as the GP method. Ferreira invented Karva language to express a chromosome [78]. For instance, a linear GEP chromosome with 2 genes and head 4 can be as follows:

$$* - a / b a 3 b b + b a * b b 7 a b \quad (1)$$

The above phrase is in the Karva language, which is given as a gene expression. As can be seen, this expression includes a combination of functions, variables, and constants together. Furthermore, Karva language can express a chromosome as a tree structure. Moreover, one ET can be expressed as a simple mathematical relation. In this regard, Ferreira invented quadruple rules for this conversion [79]. ETs of this example can be drawn by considering a linking function of multiplication for it, Fig. 4. Additionally, each

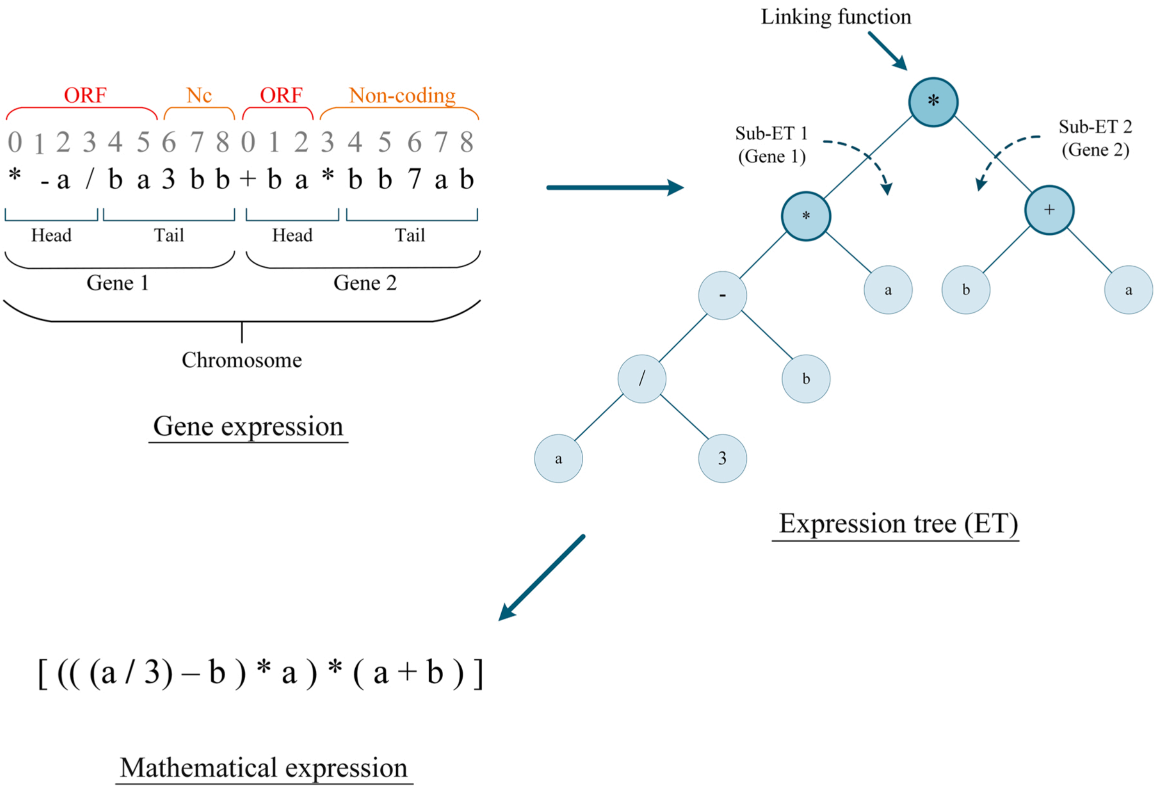

Fig. 4. The decoding of chromosomes with two genes in GEP.

gene in the structure of the chromosome is called a subset of ETs (Sub-ETs). Therefore, a chromosome combining these Sub-ETs is expressed to the GEP method.

As shown in Fig. 4, not all variables are placed in ETs, due to ETs are developed by functions, not by terminals. Each gene may have active and inactive parts. Furthermore, active encoded parts of any gene in ETs are called Open Reading Frames (ORF). Moreover, its inactive part is called non-coding. In the procedure of GEP, ETs are involved in the selection process. Then, they are led by their fitness to create new chromosomes. Instead of their related ETs, the genetic operator modifies the chromosomes during the reproduction phase. Fig. 5 is a general representation of how the GEP algorithm is implemented [80].

2.3. *Effective variables*

Thanks to the nature of ECC, the volume of cement used in the compositions of this mixture is more than normal concrete. Therefore, many experimental works have been carried out by researchers to replace chemical and mineral additives instead of used cement in the mixtures of ECC. Apart from the goal of reducing cement consumption, the difference in price between cement and other supplementary cementitious material is an economic benefit. As mentioned before, results of most experiment works that have been done to improve the mechanical properties of ECC have shown a better response of FA than the other additives. Known factors affecting mechanical properties are the volume of aggregate used, the volume of binder powder, and the quantity of water consumed. On the other hand, in order to make the mixture of ECC, fibers must be added. The dominant experiment works done in the field of ECC reviews have also focused on the utilization of PVA in the mixtures. Investigations of Yang indicate that PVA fiber had a major effect on improving the compressive strength peak in comparison with steel fiber. (steel fiber) [22]. Thereafter, two important factors affecting the fiber-reinforced concrete are mentioned fiber volume fraction consumed and aspect ratio (length and diameter of the fiber) [22]. Furthermore, it must be acknowledged that modeling the compressive strength behavior of the concrete is fundamentally difficult since it is affected by many different parameters. In this review, GEP approach is used to establish significant relations between the compressive strength of the ECC mixture (f'_c) and the effective variables as follows;

$$f'_c = f(\mathrm{V_f}, \mathrm{L}, \mathrm{WB}, \mathrm{BA}) \tag{2}$$

Where V_f (%), L (mm), WB, and BA are fiber volume fraction, fiber length, the ratio of water to binder powder (accordance with the Abrams W/B rule), and the ratio of binder powder to aggregates volume used in the compositions, respectively. According to a large-scale pilot and literature reviews that have been done, the above variables which are mentioned are used as input variables. Clearly, the output variable of this model is f'_c (MPa). It must be noticed that the values f'_c of the total data are inverted into cubes f'_c 100 × 100

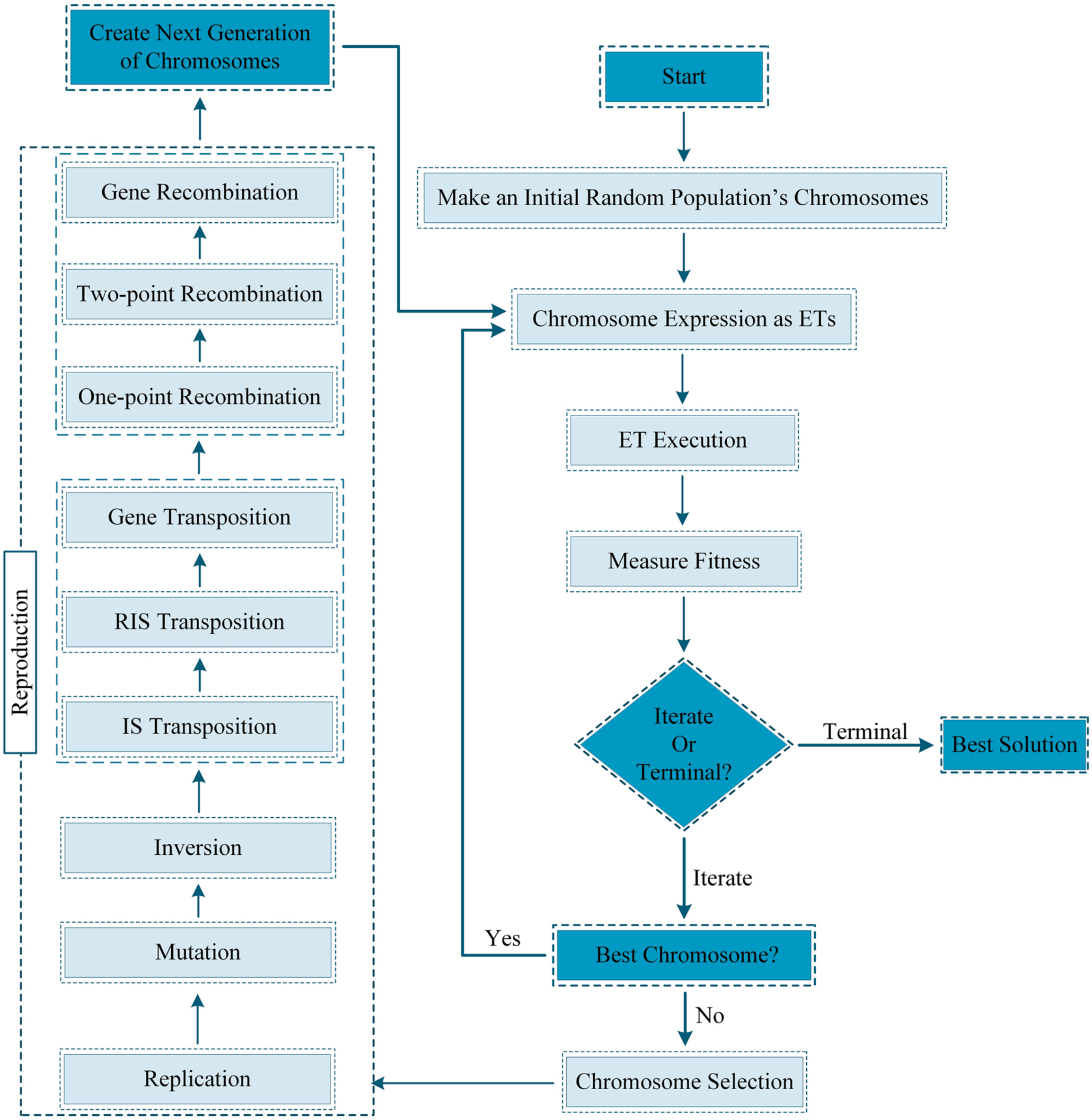

Fig. 5. The flowchart of a gene expression programming.

Table 1
The amount of input and output used in GEP approach models.

	Notation	Minimum	Mean	Maximum
Input variables				
Volume fraction of PVA fiber (%)	V_f	0.50	2.10	5.00
Length of fiber (mm)	L	6.00	8.84	12.00
Water to Binder ratio	WB	0.15	0.26	0.35
Binder to Aggregate ratio	BA	1.42	2.63	5.00
Output variable				
Compressive strength (MPa)	f'_c	5.38	27.08	108.29

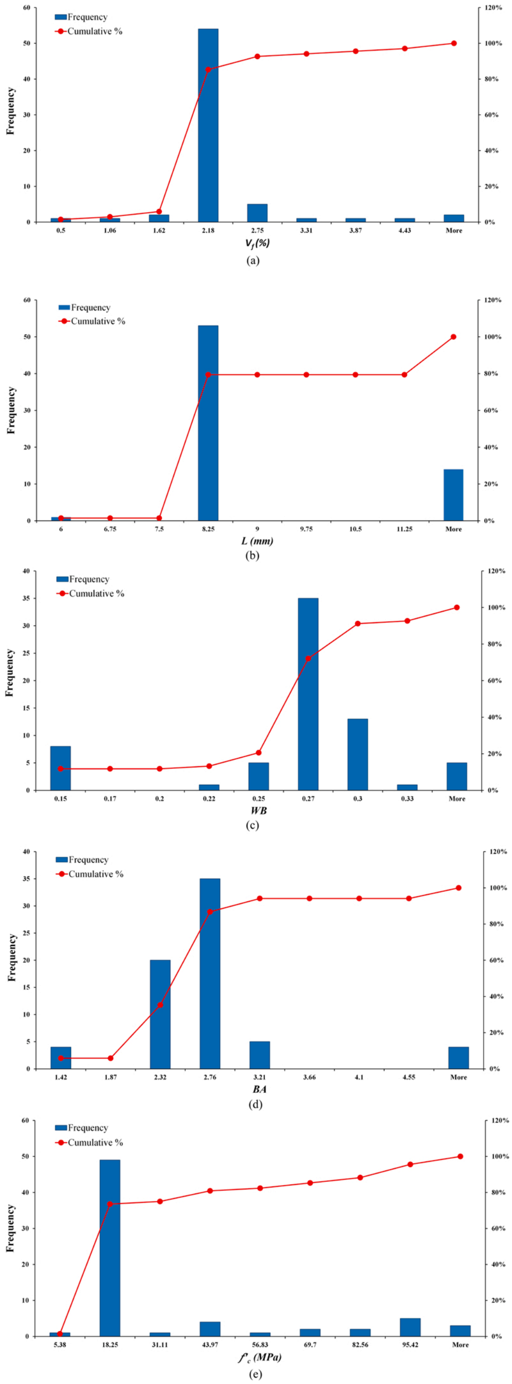

Fig. 6. Histograms of the variables, (a) V_f (%), (b) L(mm), (c) WB, (d) BA, (e) f'_c (MPa).

(mm). Table 1 shows the limit values of the desired input and output variables in the predicting modeling.

2.4. Experimental data

As is clear, the models derived from ANNs, GEP and other similar methods can, in most cases, predict within the data range used to develop their model. Due to this, the dataset used in the manufacturing process of the GEP model severely affects the reliability of the end result. Hence, a comprehensive set is the only possible solution to overcome this limitation. Therefore, a dataset should include a wide range of mixtures with various compositions. In this review, data collection is exploited which only natural aggregate, PVA fiber

Table 2
The prediction of statistics information of used variables in the model construction.

Parameter		V_f (%)	L (mm)	WB	BA	f'_c (MPa)
Standard error (SE)		0.0759	0.2039	0.0061	0.0828	3.6854
Median		2.0000	8.0000	0.2701	2.7473	12.0400
Standard deviation (SD)		0.6304	1.6944	0.0513	0.6883	30.6132
Mode		2.0000	8.0000	0.1500	2.7473	0.0000
Variance		0.3980	2.8710	0.0030	0.4740	937.1730
Kurtosis		10.182	-0.0470	1.0960	6.3530	0.8120
Skewness		2.4950	1.3030	- 1.0210	2.0270	1.5460
Range		4.5000	6.0000	0.2071	3.5715	102.9028
Sum		144.9700	610.0000	18.1779	182.1454	1910.8871
95% Confidence Interval for Mean	Lower Bound	1.9495	8.4335	0.2511	2.4744	20.3399
	Upper Bound	2.2524	9.2476	0.2757	2.8051	35.0481

Table 3
Correlation coefficients between all pairs of the explanatory variables.

Variable	V_f (%)	*L (mm)*	*WB*	*BA*	f'_c *(MPa)*
V_f (%)	1.000	0.415	- 0.579	0.160	- 0.302
L (mm)	0.415	1.000	- 0.735	0.518	0.760
WB	- 0.579	- 0.735	1.000	- 0.330	- 0.633
BA	0.160	0.518	- 0.330	1.000	0.038
f'_c*(MPa)*	- 0.302	0.760	- 0.633	0.038	1.000

Table 4
The settings of parameter for the GEP algorithm.

Parameter settings		GEP-I		GEP-II	
		SDD	RDD	SDD	RDD
General	Generation number	5000	5000	7000	7000
	Number of chromosomes	40	30	50	30
	Number of head sizes	8	8	10	8
	Number of genes	3	3	4	3
	Number of tail sizes	9	9	11	9
	Gene size	17	17	21	17
	Linking function	Addition	Addition	Multiplication	Multiplication
	Fitness function	RMSE	RMSE	RMSE	RMSE
	Function set	+ , -, * , /, Sqrt, Nop, X^3, 3Rt	+ , -, * , /, Sqrt, 3Rt, 4Rt, 5Rt, Nop, X^2, X^3, X^4, X^5	+ , -, * , /, Sqrt, 3Rt, 4Rt, 5Rt, Nop, X^2, X^3, X^4, X^5	+ , -, * , /, Sqrt, 3Rt, 4Rt, 5Rt, Nop, X^2, X^3, X^4, X^5
Genetic operators	Mutation rate	0.035	0.039	0.035	0.039
	Inversion rate	0.05	0.05	0.1	0.05
	IS transposition rate	0.1	0.1	0.1	0.1
	RIS transposition rate	0.1	0.1	0.1	0.1
	Gene transposition rate	0.1	0.1	0.1	0.1
	One-point recombination rate	0.2333	0.1	0.2333	0.1
	Two-point recombination rate	0.2333	0.3	0.2333	0.3
	Gene recombination rate	0.2333	0.3	0.2333	0.3
RNC	RNC mutation	0.1	0.1	0.1	0.1
	DC mutation	0.1	0.1	0.1	0.1
	DC IS transposition	0.1	0.1	0.1	0.1
Numerical Constant	Constants per gene	5	8	8	8
	Data type	Integer	Integer	Integer	Integer
	Lower bound	-9	-10	-15	-10
	Upper bound	9	10	15	10

of REC-15 type, and additives, containing FA of F class, SF, slag (SL), and limestone powder (LP) are used in their mix plans of ECC. It should be noted that the dominant percentage of these additives in the desired compositions was only FA. Datasets for this model comprise variables of ECC compositions, such as V_f, l, WB, and BA. The data are presented by frequency histogram to show the

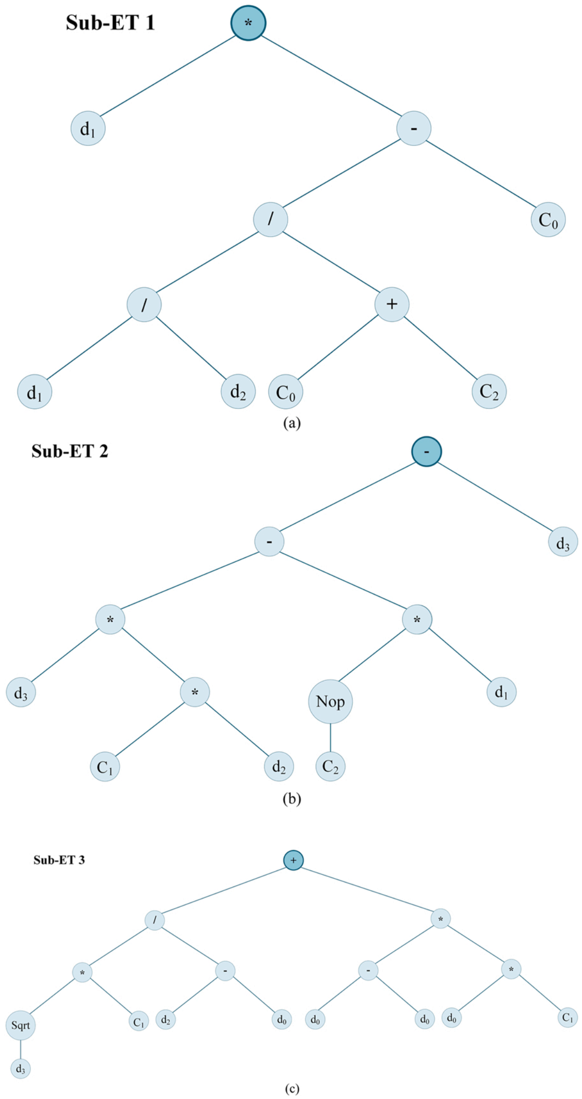

Fig. 7. Expression tree of GEP-I approach model in state of sorted data distribution (SDD) that includes (a) Sub-ET 1, (b) Sub-ET 2, (c) Sub-ET 3.

distribution of these parameters, Fig. 6. Furthermore, their descriptive statistics are listed in Table 2.

Recognition of the dependence of variables on one another is one of the most important issues in the quick access to an efficient and effective model. The first step in data interdependence analysis is the detailed study of what these variables are measuring are done. In this study, attention also needs to be given to highly correlated pairs. The range of variation for the correlation coefficient differs between -1 to $+1$. The high negative or positive correlation coefficients between pairs make it possible to result in poor model performance and make it difficult to interpret the effects of explanatory variables on the response. Furthermore, problems in analysis can be caused by this interdependence, thanks to it tends to overstate the strength of relationships between variables. This is a reason for a common occurrence which is known as the multiple collinearity problem [74,81]. Table 3 illustrates the correlation coefficients between all determined pairs in this model. As illustrated in the table, no strong correlation exists between the variables specified for this prediction.

For GEP-based modeling and the other non-linear regression-based analysis, the dataset is accidentally separated into two sub-categories of training and testing. Training data are utilized to construct the (genetic evolution) model. The testing data, which had no role in model development, is used to measure the model performance. On the other hand, some of the researchers use another subset called validation for better measurement of their predictions. In general, this work is aimed at achieving a coherent and consistent data partition. In this review, the data value of training, testing, and validation comprises 49 (about %64), 20 (almost %26), and 7 (nearly % 10), respectively. It refers to each set's share as a percentage (training, testing, and validation) of the total number of data being considered (76 data).

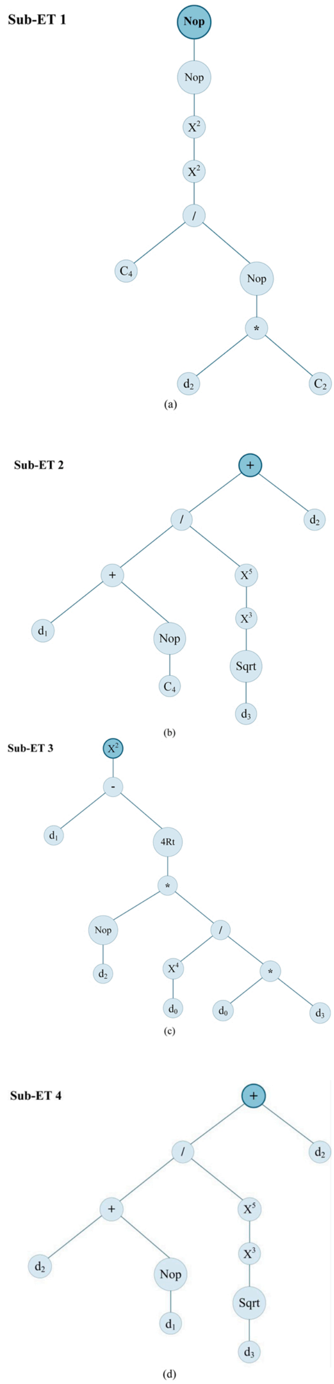

Fig. 8. Expression tree of GEP-II approach model in state of sorted data distribution (SDD) that includes (a) Sub-ET 1, (b) Sub-ET 2, (c) Sub-ET 3, (d) Sub-ET 4.

3. Application of GEP

GEP Model-making proceeds after the determination of input variables. First, the desired set of functions is selected by the user to exploit in modelling. In this review, GEP-I and GEP-II, which are two different models of GEP are developed. Table 4 shows the important factors for the configuration setting of these models. The distribution of input data of each model was considered in two ways, sorted (SDD) and random (RDD).

Determining the population size (number of chromosomes) is the next step which indicates the number of programs in the

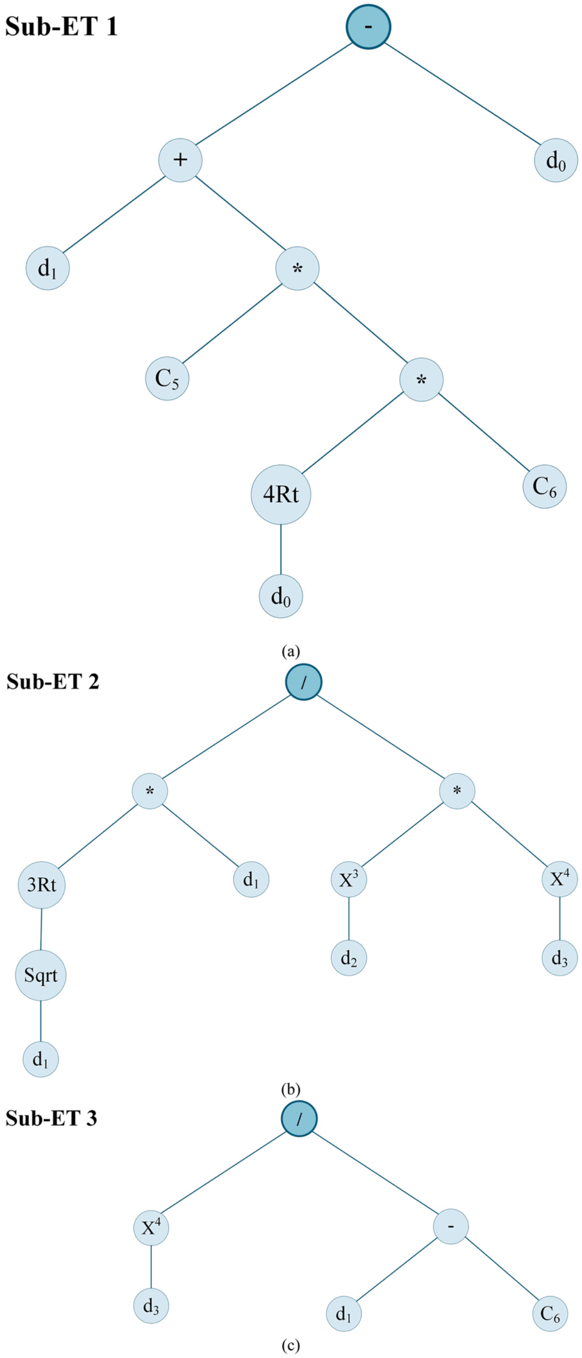

Fig. 9. Expression tree of GEP-I approach model in state of Random data distribution (RDD) that includes (a) Sub-ET 1, (b) Sub-ET 2, (c) Sub-ET 3.

population. The time of an implementation depends on the population size. As a result, the greater the population, the more complex and precise the model presented. It may take longer to achieve the best fitness for the model. It should also be noted that overpopulation causes an overfitting problem. However, getting a proper population size depends on how many possible solutions are and how complex a problem is. The program implementation is completed by considering a defined population for each model. After this step, it is time to design the chromosomal tree structure. In this way, the frames' heads and the number of genes that build these frames. The complexity of each term in the GEP model is shown by the first factor, and on the other hand, the number of terms is described by the number of genes in the chromosomal structure. As mentioned above, any gene is characterized by the subset of ETs. Thereafter, ETs (chromosome) is generated by combining sub-ETs with an assumed linking function. In order to link mathematical terms encoded in each gene when the desired number of genes in the modelling is greater than 1, the multiplication linking function is utilized. The ETs

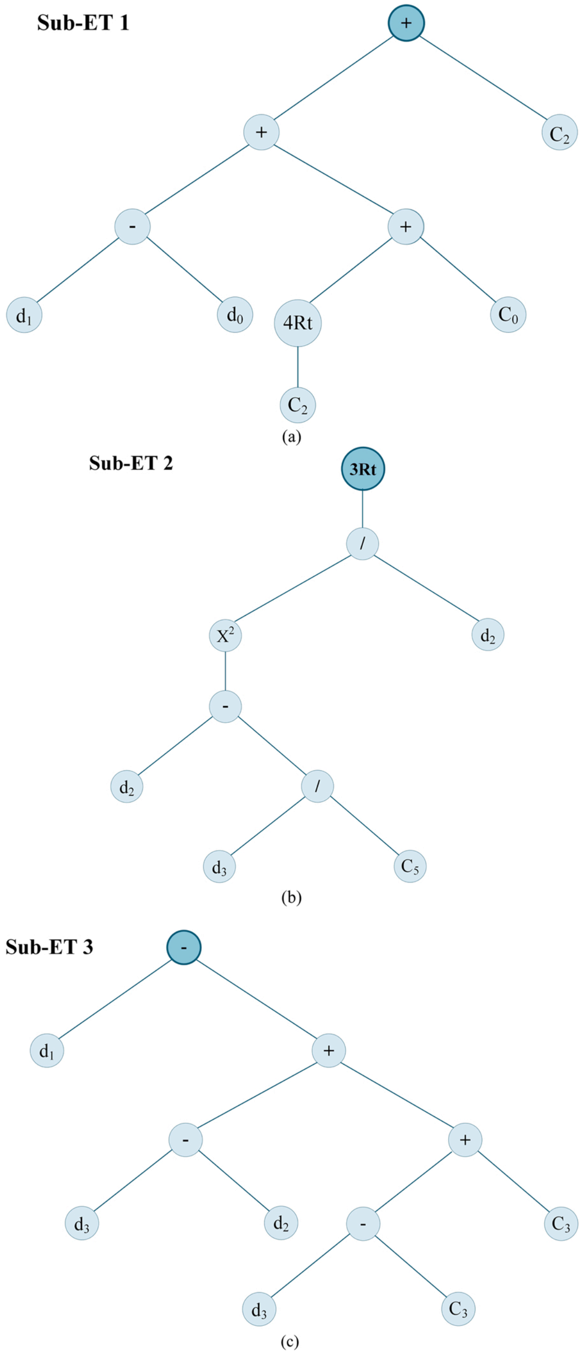

Fig. 10. Expression tree of GEP-II approach model in state of Random data distribution (RDD) that includes (a) Sub-ET 1, (b) Sub-ET 2, (c) Sub-ET 3.

of these two models show as sorted data distribution (SDD) in Fig. 7 and Fig. 8 and as random data distribution (RDD) in Fig. 9 and Fig. 10.

Finally, in order to calculate the fitness of evolved program for the models, RMSE is utilized. Trial and error is used to determine the whole values for the elements included in modelling [74]. GeneXpro Tools software 5.0 is used to develop experimental models based on the GEP method [82].

4. Evaluation of models

4.1. Validation study

One of the known relations concerning the evaluation of effectiveness and efficiency for the models is the correlation coefficient. This criterion is not recommended as using only this method to verify the model accuracy due to its insensitivity to the basic calculations (multiplication and division) of the outputs to the constant [83]. Therefore, five rules have been used in this review to compare the approach performance for the GEP-I and GEP-II models. Root mean squared error (RMSE), mean absolute percentage error (MAPE), R-square (R^2), root relative square error (RRSE), performance index (PI) [84], and relative absolute error (RAE) are the desired criteria. Terms of expressed error functions were provided as Eqs. (3)- (8) below.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(e_i - o_i)^2}{n}} \tag{3}$$

$$\mathrm{MAPE} = \frac{1}{n}\left[\frac{\sum_{i=1}^{n}|e_i - o_i|}{\sum_{i=1}^{n} e_i} \times 100\right] \tag{4}$$

$$R^2 = \frac{(n\sum e_i o_i - \sum e_i \sum o_i)^2}{(n\sum e_i{}^2 - (\sum e_i)^2)(n\sum o_i{}^2 - (\sum o_i)^2)} \tag{5}$$

$$\mathrm{RRSE} = \sqrt{\frac{\sum_i (e_i - o_i)^2}{\sum_i \left(e_i - \left(\frac{\sum_i e_i}{n}\right)\right)^2}} \tag{6}$$

$$\mathrm{RAE} = \frac{\sum_i |e_i - o_i|}{\sum_i \left|e_i - \left(\frac{\sum_i e_i}{n}\right)\right|} \tag{7}$$

$$\mathrm{PI} = \frac{\mathrm{RRMSE}}{1 + R} \tag{8}$$

In the above statistical calculation relations, e_i is the quantity of experimental samples, the output value of predicted samples is o_i and the total amount of data is n. In this regard, both functions of RRMSE and R in PI criteria, respectively, mean the relative root mean square error (RRMSE) and the correlation coefficient (R) [85]. Additionally, having an R^2 of more than 0.8 and statistical error values less such as RMSE, MAPE, RRSE, PI, and RAE for machine learning, indicates a reliable and efficient model.

4.2. Validity of the models and comparison of their results

Noted: modelling has been done 43 times. 50 runs have been taken in each model. Totally, 2150 formulas were made, of which the best 4 formulas were selected.

According to the ETs presented in Fig. 7 and Fig. 8 for two GEP-I and GEP-II models in the mode of sorted data distribution (SDD), their extracted formulas are given below, respectively.

$$f_c = \left[L \times \left(\frac{\left(\frac{L}{WB}\right)}{(C_0 + C_2)} - C_0\right)\right] + [((BA \times (C_1 \times WB)) - (WB \times L)) - BA] \times \left[\left(\frac{\sqrt{BA} \times C_1}{WB - V_f}\right)\right] \tag{9}$$

$$f_c = \left[\left(\frac{5}{WB \times 14}\right)^4\right] \times \left[\frac{L + 5}{(\sqrt{BA})^{15}} + WB\right] \times \left[\left(L - \sqrt[4]{WB \times \left(\frac{V_f{}^4}{V_f \times BA}\right)}\right)^2\right] \times \left[\frac{WB + L}{(\sqrt{BA})^{15}} + WB\right] \tag{10}$$

It is worth noting that all defined parameters presented in ETs for GEP-I and GEP-II models in both SDD and RDD modes, including d_0, d_1, d_2 and d_3 indicate V_f, L, WB and BA, respectively. For the method of the GEP-I model, the constant values in the formula are equal to; in Sub-ET1$C_0 = 2$ and $C_2 = 6$, in Sub-ET2$C_1 = -8$ and in Sub-ET3$C_1 = -9$. Additionally, for the method of the GEP-II model, the constant values in the formula are equal to; in Sub-ET1$C_2 = 14$ and $C_4 = 5$ and in Sub-ET2$C_4 = 5$.

According to the above process, the following formulas are extracted from Figs. 9 and 10 for both models in random data

Table 5
GEP models results in the SDD mode compared with experimental results are used as testing sets.

Data used in models construction				Compressive strength (MPa)			Ref.
V_f (%)	L (mm)	WB	BA	Exp.	GEP-I	GEP-II	
2	12	0.150	2.7473	95.12	101.57	99.69	[20]
2	8	0.250	2.6667	11.13	17.56	13.72	[91]
2	8	0.250	2.7532	10.18	17.25	13.59	[21]
2	8	0.250	2.7532	9.10	17.25	13.59	[21]
2	8	0.250	2.7532	7.54	17.25	13.59	[21]
2	8	0.278	2.4779	8.06	12.91	11.31	[21]
2	8	0.312	2.2026	8.63	8.99	9.63	[21]
2	8	0.263	2.6155	9.20	15.03	12.39	[21]
2	8	0.278	2.4779	9.70	12.91	11.31	[21]
2	8	0.269	2.2409	14.22	15.37	13.12	[17]
2	8	0.269	2.7516	11.69	13.27	11.57	[17]
2	8	0.270	2.7494	11.94	13.24	11.55	[17]
2	8	0.270	2.7556	13.93	13.22	11.55	[17]
2	8	0.270	2.2385	11.52	15.26	13.07	[17]
2	8	0.270	2.2428	12.14	15.24	13.04	[17]
2	8	0.269	2.2428	12.97	15.40	13.13	[17]
2	8	0.269	2.2409	12.51	15.37	13.12	[17]
2	8	0.301	2.7494	11.23	8.35	9.22	[17]
2	8	0.300	2.2371	10.14	10.45	10.33	[17]
2	12	0.232	2.7473	30.19	47.94	39.53	[18]

Table 6
GEP models results compared with experimental results are used as validation sets.

Data used in models construction					GEP-I		GEP-II		Ref.
V_f (%)	L (mm)	WB	BA	Exp.	SDD	RDD	SDD	RDD	
1	8	0.35	1.4285	67.88	69.72	58.81	66.55	60.97	[90]
1.5	8	0.35	1.4285	72.97	69.50	57.03	62.07	59.67	[90]
2	8	0.35	1.4285	74.65	69.19	55.54	58.09	58.38	[90]
1.75	12	0.273	2.7659	27.30	17.49	16.18	28.54	29.38	[88]
1.75	12	0.271	2.7586	21.79	17.77	16.50	29.01	29.70	[88]
1.75	12	0.272	2.7586	16.44	17.70	16.39	28.79	29.33	[88]
2	8	0.25	2.6667	10.39	11.50	11.66	13.72	13.47	[91]

Table 7
The output and statistical results of external validation.

Exp.	Ghafor K.	Khandaker M.	***Statistical parameters***	Ghafor K.	Khandaker M.
67.88	151.81	107.49			
72.97	149.55	107.49	RMSE	65.1819	63.7524
74.65	147.31	107.49	MAPE	21.5819	19.8036
27.30	97.29	84.50	R^2	0.9300	0.0851
21.79	70.59	81.51	RRSE	2.3788	2.3266
16.44	46.98	78.69	RAE	2.5181	2.3106
10.39	68.21	128.29	PI	33.1832	49.3516

distribution (RDD) mode. The formulas of the GEP-I and GEP-II models are presented, respectively.

$$f_c = \left[\left(L + \left(C_5 \times \left(\sqrt{4V_f} \times C_6\right)\right)\right) - V_f\right] + \left[\frac{\left(\sqrt{6L} \times L\right)}{(WB)^3 \times (BA)^4}\right] + \left[\frac{(BA)^4}{L - C_6}\right] \quad (11)$$

$$f_c = \left[\left(\left(L - V_f\right) + \left(\sqrt{4C_2 + C_0}\right)\right) + C_2\right] \times \left[\sqrt{3\frac{\left(WB - \left(\frac{BA}{C_5}\right)\right)^2}{WB}}\right] \times \left[L - \left((BA - WB) + BA\right)\right] \quad (12)$$

The numerical constants in formula 12 (GEP-I) based on its respective Sub-ETs are as follows; in Sub-ET1$C_5 = -3$ and $C_6 = 4$ and in Sub-ET3 C6 $= -1$. In this regard, the constant values in formula 13 (GEP-II) conforming to its Sub-ETs include $C_2 = 6$ and $C_0 = 9$ in Sub-ET1, $C_5 = 9$ in Sub-ET2 and $C_3 = -10$ in Sub-ET3.

In general, all 76 collected data for making these models are obtained data from the experimental works that have been done [17,

Table 8
Statistical parameters of the GEP models.

Statistical parameters		GEP-I			GEP-II		
		Training	Testing	Validation	Training	Testing	Validation
SDD	RMSE	3.8799	5.4556	4.7874	3.8137	3.3345	9.3520
	MAPE	0.2143	0.9686	1.3224	0.1985	0.7616	2.5946
	R^2	0.9853	0.9351	0.9776	0.9856	0.9820	0.9392
	RRSE	0.1224	0.2916	0.1796	0.1203	0.1782	0.3509
	RAE	0.1181	0.3338	0.1488	0.1094	0.2625	0.2920
	PI	0.0681	0.1726	0.0577	0.0669	0.1042	0.1140
RDD	RMSE	5.0522	7.4607	11.0478	6.5820	9.5602	10.2249
	MAPE	0.3145	0.3839	3.0310	0.3145	0.4296	3.0608
	R^2	0.9708	0.9567	0.9704	0.9493	0.9202	0.9523
	RRSE	0.1728	0.2265	0.4145	0.2251	0.2903	0.3836
	RAE	0.1760	0.2162	0.3412	0.2246	0.2419	0.3445
	PI	0.0953	0.1249	0.1336	0.1249	0.1616	0.1242

Table 9
Correlation of predicted output results of the GEP models.

Predicted output	GEP-I (RDD)	GEP-I (SDD)	GEP-II (RDD)	GEP-II (SDD)
GEP-I (RDD)	1	-0.0053	0.9754	0.0172
GEP-I (SDD)	-0.0053	1	0.0278	0.9944
GEP-II (RDD)	0.9754	0.0278	1	0.0510
GEP-II (SDD)	0.0172	0.9944	0.0510	1

18,20,21,86–91]. The collected data are used as the sets of training, testing, and validation for GEP-I and GEP-II models. 49 records from the experimental results were applied for training in each model. Moreover, another 20 data also are used for testing. 7 remained data is considered for the validation of the models. All data have 28-day f_c. The natural aggregate has been used in all combinations of the gathered samples. In the following, the input values and experimental results with the testing results obtained from two GEP-I and GEP-II models in sorted data distribution (SDD) mode are given in Table 5. On the other hand, input values and experimental results with validation results of all models in both sorted data distributed (SDD) and random data distributed (RDD) modes are listed in Table 6. In this regard, two other models about the predicted compressive strength of ECC have been proposed by Ghafor et al. [92] and Khandaker et al. [93] in order to compare these models with the ones presented in this study. Table 7 illustrates the output and statistical results of these two models. It should be noted that the output of these models has been converted to cubic f'_c 100 × 100 mm, just like the output of the models of this study. Based on the results from Table 7, it was observed that Khandaker's model had low accuracy, as its R^2 value was only 0.08. On the other hand, Ghafor's model had an R^2 value of 0.93, but a comparison of its output values with experimental data revealed a significant difference. This implies that the R^2 value alone should not be used as the sole criterion for evaluation, and therefore, this model also demonstrated poor performance.

Table 8 lists the results of statistical parameters for the training, testing and validation categories of GEP-I and GEP-II in both SDD and RDD modes. According to the results of the gained evaluation criteria, RMSE, MAPE, R^2, RRSE, RAE and PI of the training results for the GEP-I model in SDD mode are obtained 3.87, 0.21%, 0.98, 0.12, 0.11, and 0.06 respectively. Furthermore, the statistical results of the training process for the GEP-I model in RDD mode are 5.05, 0.31%, 0.97, 0.17, 0.17 and 0.09, respectively. Moreover, the statistical values of RMSE, MAPE, R^2, RRSE, RAE and PI for testing the GEP-I model in SDD mode are, respectively, 5.45, 0.96%, 0.93, 0.29, 0.33 and 0.17 and the evaluation criteria results of the testing set of this model in RDD state are 7.46, 0.38%, 0.95, 0.22, 0.21 and 0.12 respectively. The statistical values of RMSE, MAPE, R^2, RRSE, RAE and PI for training the GEP-II model in SDD mode are found 3.81, 0.19%, 0.98, 0,12, 0.10 and 0.06, respectively. In accordance with the previous state that was noticed, the statistical values of GEP-II model training results in RDD mode are as follows: 6.58, 0.31%, 0.94, 0.22, 0.22 and 0.12, respectively. The statistical indicators containing RMSE, MAPE, R^2, RRSE, RAE and PI for testing results of the GEP-II model in SDD mode are 3.33, 0.76%, 0.98, 0.17, 0.26 and 0.10, respectively. Subsequently, the results of this model in RDD mode are obtained 9.56, 0.42%, 0.92, 0.29, 0.24 and 0.16, respectively. The values of statistical parameters for validation results comprising RMSE, MAPE, R^2, RRSE, RAE and PI for the GEP-I model in SDD mode are revealed 4.78, 1.32%, 0.97, 0.17, 0.14 and 0.05, and these results for this model in RDD mode are 11.04, 3.03%, 0.97, 0.41, 0.34 and 0.13, respectively. Meanwhile, these statistical values, which contain RMSE, MAPE, R^2, RRSE, RAE, and PI for the GEP-II model validation results in SDD mode, are found 9.35, 2.59%, 0.93, 0.35, 0.29, and 0.11, and also for the RDD mode of this model 10.22, 3.06%, 0.95, 0.38, 0.34 and 0.12, respectively. As seen in Table 8, the value of R^2 for all three sets of training, testing and validation of the two suggested models in both SDD and RDD modes is presented above 0.9. Based on the comparison between the two models in both their modes, all the proposed models are admissible and accepted. It is because all these models predict a close response to the results of the experimental studies for the f_c results of 28-day samples of ECC for three sets of training, testing, and validation.

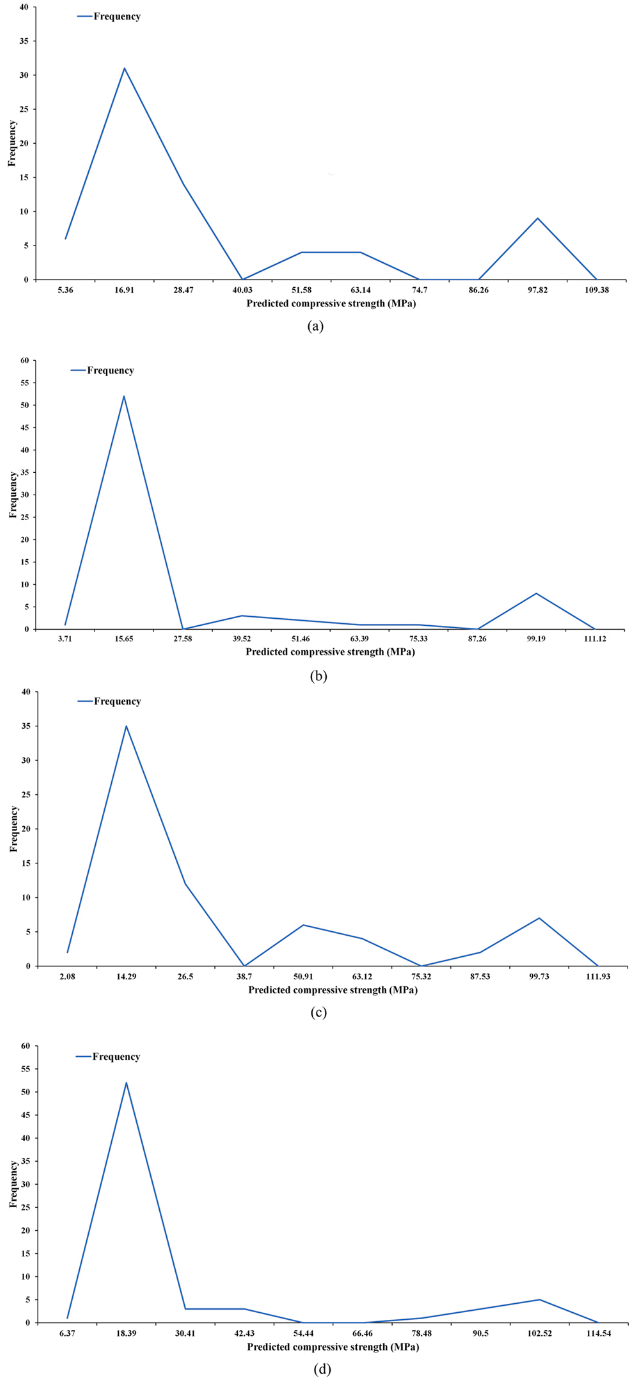

Fig. 11. Frequency distribution of f'_c in the models (a) GEP-I (RDD), (b) GEP-I (SDD), (c) GEP-II (RDD), (d) GEP-II (SDD).

4.3. Parametric and sensitivity analysis

First, to perform an overview of the results of the constructed models, the correlation of the predicted outputs of the models has been analyzed. According to Table 9, the results of GEP-I and GEP-II models have the least dependence on each other in two modes of sorted data distribution (SDD) and random data distribution (RDD). Despite the formulas between the two models being different in RDD and SDD modes, both models showed high dependence in each of the two modes. For instance, the correlation between GEP-I and GEP-II models in RDD mode is equal to 0.9754. Further, the distribution of f'_c of all models is presented in Fig. 11. The frequency diagrams of f_c are used to determine whether the output responses of the models created match the frequency values of the

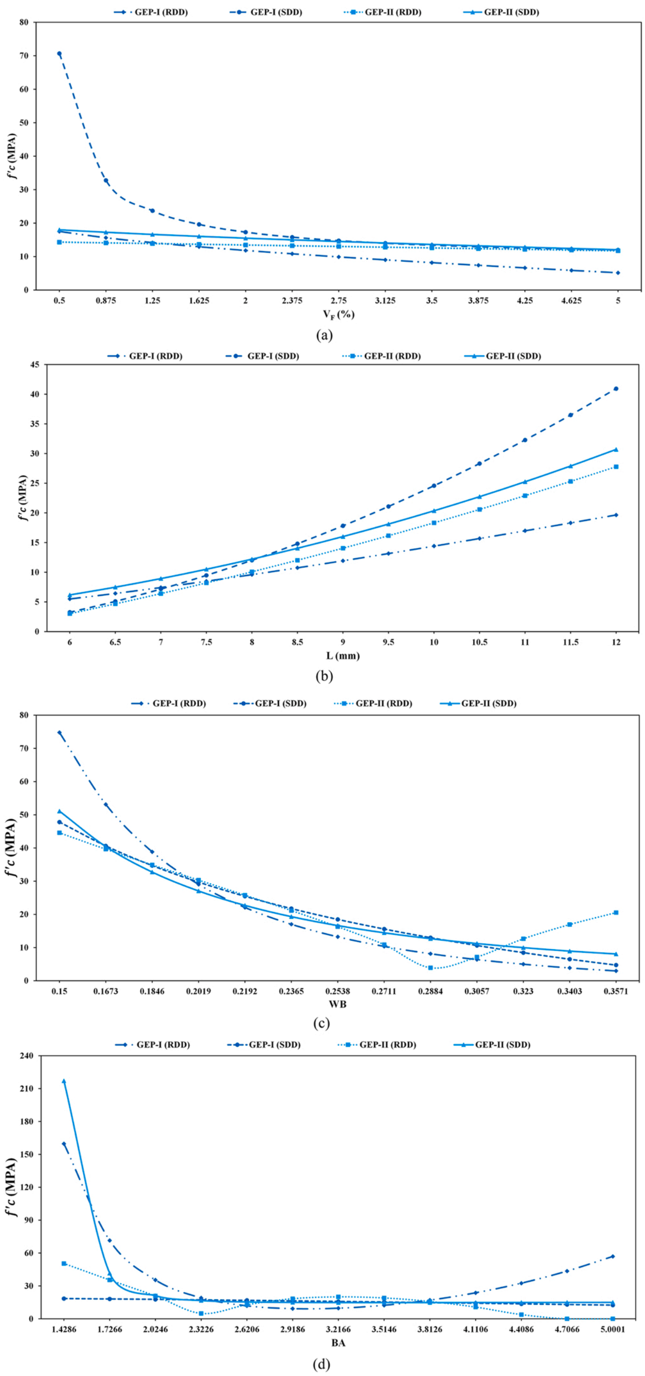

Fig. 12. Analyses of compression strength parametrically in GEP-based models.

experimental data of fc. As the results indicate, the trend of the GEP-I model in SDD mode more closely matches the frequency graph of the experimental data in Fig. 6(e) compared to other models.

On the other hand, the created models can be evaluated by conducting parametric studies. In this study, the compressive strength response of the predicted models to a set of input variables (V_f, L, WB, BA) is investigated. In this analysis, the range of an input variable is placed with the average of other input variables in each formula of the models. Fig. 12 indicates The parametric analysis performed on the models. As can be seen, the compressive strength of the models grows with the increase of the input variable L. Moreover, as the values of V_f and WB variables rise, the f'_c results decrease in all models except the GEP-II (RDD) model. The f'_c of the models go down with the growth of the value of the BA variable except for the GEP-I (RDD) model.

Table 10
GEP model sensitivity results (in percentage) for input parameters.

Models	Input parameter			
	V_f	L	WB	BA
GEP-I (RDD)	4.96	5.68	28.86	60.48
GEP-I (SDD)	40.36	25.89	29.61	4.12
GEP-II (RDD)	1.94	18.64	30.61	48.79
GEP-II (SDD)	2.15	8.89	15.60	73.34

In this regard, to measure the impact of the input variables in each model on the predicted results, the study of sensitivity analysis can be used. The following equation can be utilized as a percentage of sensitivity:

$$I_i = f'_{\max}(N_i) - f'_{\min}(N_i) \tag{13}$$

$$\mathrm{Si} = \frac{I_i}{\sum_{i=1}^{n} I_i} \times 100$$

In the above formula, $f'_{\max}(N_i)$ and $f'_{\min}(N_i)$, respectively, are the maximum and minimum values of the predicted output in the range of the i-th input variable. Additionally, the values of other input variables are equal to their average. The sensitivity results of each input parameter on each model's output results are presented in Table 10. As demonstrated, the variable of the ratio of water to binder (WB) has an effective role in all models. Additionally, the input parameter of binder to aggregates ratio (BA) has a much more effective role than other parameters in the final response of all models except the GEP-I (SDD) model. However, the variable effect of V_f on the predicted compressive strength in all models except the GEP-I (SDD) model is extremely insignificant.

5. Conclusions

By utilizing the algorithm GEP, this paper attempts to present a novel and efficient approach to the formulation of engineered cementitious composite (ECC) that includes FA and PVA fibre. Therefore, for predicting the compressive strength of 28-day ECC samples, two different proposed models of GEP-I and GEP-II in two modes (SDD and RDD) are used. Moreover, it should be paid attention that all desired input variables to make these models, including V_f, L, WB, and BA were collected from the results of experimental works that have been done before. The developed model of GEP-I in two modes of SDD and RDD in this study with the same number of genes was equal to 3. Nonetheless, the number of GEP-II model genes in SDD and RDD modes was 4 and 3, respectively. Additionally, two different linking functions (addition and multiplication) were used to compare the performance of the presented formulas. Furthermore, in order to evaluate the efficiency of the suggested formulas, some statistical criteria, including RMSE, MAPE, RRSE, RAE and PI were used. The study conducted produced the following results:

1. The sub-ET density in RDD mode was lower than in SDD mode because RDD mode models are created with fewer chromosomes.
2. The accuracy of all models produced by both modes was the same ($R^2 > 0.9$), indicating that they all meet the necessary requirements for an efficient and reliable model.
3. Out of all the models in the validation set, the GEP-I model in SDD mode performed the best with an R^2 value of approximately 0.98, and the lowest statistical criterion values (RMSE, MAPE, RRSE, RAE, and PI) compared to other models.
4. Based on the correlation analysis among the models, GEP-I and GEP-II in SDD mode exhibited the strongest correlation, which was 0.99, due to their sorted distribution data.
5. The GEP-I model in SDD mode demonstrated a better distribution and weighting of sensitivity results across the input parameters compared to other models.
6. According to the frequency diagrams of f'_c (MPa), it was observed that the GEP-I model in SDD mode had a more accurate correspondence with the experimental data of f'_c (MPa) when compared to other models.
7. The results of the parametric analysis showed that most models performed well in generating the expected output response based on the input variables, except for the GEP-II model in RDD mode with respect to the V_f and WB variables, and the GEP-I model in RDD mode in relation to the BA variable.

According to work done, the GEP algorithm can be implemented and executed to predict the properties of cementitious composites and various concretes. Thus, GEP may work as an effective and powerful method. Furthermore, using GEP has the ability to save time and money compared to traditional testing methods. GEP can produce precise predictions quickly without requiring extensive physical testing. This makes it an appealing choice for engineers who want to improve their designs and shorten project timelines. Moreover, predicting the properties of materials may lead to creating new and better materials with superior performance characteristics. In general, applying GEP in civil engineering has the potential to transform the field and encourage creativity in the design of construction and infrastructure.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Data will be made available on request.

References

[1] Z.Q. Zhao, R.J. Sun, Z.Q. Feng, S.S. Wei, D.W. Huang, Mechanical Properties And Applications Of Engineered Cementitious Composites (ECC). Applied Mechanics and Materials, Trans Tech Publ, 2013.
[2] E. Booya, K. Gorospe, S. Das, P. Loh, The influence of utilizing slag in lieu of fly ash on the performance of engineered cementitious composites, Constr. Build. Mater. 256 (2020), 119412.
[3] M. Khan, M. Cao, M. Ali, Cracking behaviour and constitutive modelling of hybrid fibre reinforced concrete, J. Build. Eng. 30 (2020), 101272.
[4] M. Khan, J. Lao, J.-G. Dai, Comparative study of advanced computational techniques for estimating the compressive strength of UHPC, J. Asian Concr. Fed. 8 (2022) 51–68.
[5] Z. Chen, Y. Yang, Y. Yao, Quasi-static and dynamic compressive mechanical properties of engineered cementitious composite incorporating ground granulated blast furnace slag, Mater. Des. 44 (2013) 500–508.
[6] S.H. Said, H.A. Razak, The effect of synthetic polyethylene fiber on the strain hardening behavior of engineered cementitious composite (ECC), Mater. Des. 86 (2015) 447–457.
[7] M. Ali, A. Soliman, M. Nehdi, Hybrid-fiber reinforced engineered cementitious composite under tensile and impact loading. Mater. Des. 117 (2017) 139–149.
[8] S. Choucha, A. Benyahia, M. Ghrici, M.S. Mansour, Effect of natural pozzolan content on the properties of engineered cementitious composites as repair material, Front. Struct. Civ. Eng. 12 (3) (2018) 261–269.
[9] W.S. Alaloul, M.A. Musarat, B.A. Tayeh, S. Sivalingam, M.F.B. Rosli, S. Haruna, M.I. Khan, Mechanical and deformation properties of rubberized engineered cementitious composite (ECC), Case Stud. Constr. Mater. 13 (2020), e00385.
[10] Z. Yu, Y. Yingzi, Mechanical properties of engineered cementitious composites with high volume fly ash, J. Wuhan. Univ. Technol.: Mater. Sci. Engl. Ed. (S1) (2009) 166–170.
[11] Y. Zhu, Y. Yang, Y. Yao, Use of slag to improve mechanical properties of engineered cementitious composites (ECCs) with high volumes of fly ash, Constr. Build. Mater. 36 (2012) 1076–1081.
[12] E.-H. Yang, Y. Yang, V.C. Li, Use of high volumes of fly ash to improve ECC mechanical properties and material greenness, Acids Mater. J. 104 (6) (2007) 620.
[13] S. Wang, V.C. Li, Engineered cementitious composites with high-volume fly ash, Acids Mater. J. 104 (3) (2007) 233.
[14] Sahmaran, M., H.E. Yucel, M. Al-Emam, I.O. Yaman, and M. Guler, Bond characteristics of engineered cementitious composite overlays. 2013.
[15] N. Arivusudar, S. Babu, Mechanical properties of engineered cementitious composites developed with silica fume, Cem. -Wapno-Beton= Cem. Lime Concr. 25 (4) (2020) 282–291.
[16] A. Wardhono, Comparison study of class F and class C fly ashes as cement replacement material on strength development of non-cement mortar. IOP Conference Series: Materials Science and Engineering, IOP Publishing, 2018.
[17] M. Şahmaran, Z. Bilici, E. Ozbay, T.K. Erdem, H.E. Yucel, M. Lachemi, Improving the workability and rheological properties of Engineered Cementitious Composites using factorial experimental design, Compos. Part B: Eng. 45 (1) (2013) 356–368.
[18] S. Wang, V.C. Li, Polyvinyl alcohol fiber reinforced engineered cementitious composites: material design and performances. Proc., Int'l Workshop on HPFRCC Structural Applications, Citeseer, Hawaii, 2005.
[19] S. Wang, V.C. Li, High-early-strength engineered cementitious composites, Acids Mater. J. 103 (2) (2006) 97.
[20] B.S. Mohammed, M.H. Baharun, M.F. Nuruddin, O.P.D. Erikol, N.A. Murshed, Mechanical properties of engineered cementitious composites mixture. Applied Mechanics and Materials, Trans Tech Publ, 2014.
[21] Y. Zhu, Z. Zhang, Y. Yang, Y. Yao, Measurement and correlation of ductility and compressive strength for engineered cementitious composites (ECC) produced by binary and ternary systems of binder materials: Fly ash, slag, silica fume and cement, Constr. Build. Mater. 68 (2014) 192–198.
[22] H. Yang, J. Li, Y. Huang, Study on mechanical properties and constitutive equation of hybrid fiber reinforced cementitious composites under static loading, Open Constr. Build. Technol. J. 10 (1) (2016).
[23] C. Bilim, C.D. Atiş, H. Tanyildizi, O. Karahan, Predicting the compressive strength of ground granulated blast furnace slag concrete using artificial neural network, Adv. Eng. Softw. 40 (5) (2009) 334–340.
[24] H. Naderpour, A. Kheyroddin, G.G. Amiri, Prediction of FRP-confined compressive strength of concrete using artificial neural networks, Compos. Struct. 92 (12) (2010) 2817–2829.
[25] S. Gupta, Using artificial neural network to predict the compressive strength of concrete containing nano-silica, Civil Eng. Archit. 1 (3) (2013) 96–102.
[26] L. Shi, S. Lin, Y. Lu, L. Ye, Y. Zhang, Artificial neural network based mechanical and electrical property prediction of engineered cementitious composites, Constr. Build. Mater. 174 (2018) 667–674.
[27] A.A. Shahmansouri, M. Yazdani, S. Ghanbari, H.A. Bengar, A. Jafari, H.F. Ghatte, Artificial neural network model to predict the compressive strength of eco-friendly geopolymer concrete incorporating silica fume and natural zeolite, J. Clean. Prod. 279 (2021), 123697.
[28] A.S. Hosseini, P. Hajikarimi, M. Gandomi, F.M. Nejad, A.H. Gandomi, Optimized machine learning approaches for the prediction of viscoelastic behavior of modified asphalt binders, Constr. Build. Mater. 299 (2021), 124264.
[29] A.H. Gandomi, A.H. Alavi, Applications of computational intelligence in behavior simulation of concrete materials. Computational optimization and applications in engineering and industry, Springer, 2011, pp. 221–243.
[30] A. Ahmad, K. Chaiyasarn, F. Farooq, W. Ahmad, S. Suparp, F. Aslam, Compressive strength prediction via Gene Expression Programming (GEP) and Artificial Neural Network (ANN) for Concrete Containing RCA, Buildings 11 (8) (2021) 324.
[31] B. Vakhshouri, S. Nejadi, Prediction of compressive strength of self-compacting concrete by ANFIS models, Neurocomputing 280 (2018) 13–22.
[32] B. Vakhshouri, S. Nejadi, Predicition of compressive strength in light-weight self-compacting concrete by ANFIS analytical model, Arch. Civ. Eng. (2)) (2015).
[33] D.J. Armaghani, P.G. Asteris, A comparative study of ANN and ANFIS models for the prediction of cement-based mortar materials compressive strength, Neural Comput. Appl. 33 (9) (2021) 4501–4532.
[34] Y. El Asri, M.B. Aicha, M. Zaher, A.H. Alaoui, Prediction of compressive strength of self-compacting concrete using four machine learning technics, Mater. Today.: Proc. 57 (2022) 859–866.
[35] R.H. Faraj, A.A. Mohammed, A. Mohammed, K.M. Omer, H.U. Ahmed, Systematic multiscale models to predict the compressive strength of self-compacting concretes modified with nanosilica at different curing ages, Eng. Comput. (2021) 1–24.
[36] K. Ghafor, Multifunctional models, including an artificial neural network, to predict the compressive strength of self-compacting concrete, Appl. Sci. 12 (16) (2022) 8161.
[37] M.J. Taheri Amiri, A. Ashrafian, F.R. Haghighi, M. Javaheri Barforooshi, Prediction of the compressive strength of self-compacting concrete containing rice husk ash using data driven models, Modares Civ. Eng. J. 19 (1) (2019) 209–221.

[38] P.G. Asteris, V.G. Mokos, Concrete compressive strength using artificial neural networks, Neural Comput. Appl. 32 (15) (2020) 11807–11826.
[39] F. Khademi, S.M. Jamal, N. Deshpande, S. Londhe, Predicting strength of recycled aggregate concrete using artificial neural network, adaptive neuro-fuzzy inference system and multiple linear regression, Int. J. Sustain. Built Environ. 5 (2) (2016) 355–369.
[40] F. Altun, Ö. Kişi, K. Aydin, Predicting the compressive strength of steel fiber added lightweight concrete using neural network, Comput. Mater. Sci. 42 (2) (2008) 259–265.
[41] G. Du, L. Bu, Q. Hou, J. Zhou, B. Lu, Prediction of the compressive strength of high-performance self-compacting concrete by an ultrasonic-rebound method based on a GA-BP neural network, PLoS One 16 (5) (2021), e0250795.
[42] J.P.M. Rinchon, Strength durability-based design mix of self-compacting concrete with cementitious blend using hybrid neural network-genetic algorithm, IPTEK J. Proc. Ser. 3 (6) (2017).
[43] I. Ranjbar, V. Toufigh, M. Boroushaki, A combination of deep learning and genetic algorithm for predicting the compressive strength of high-performance concrete, Struct. Concr. (2022).
[44] G. Abdollahzadeh, E. Jahani, Z. Kashir, Predicting of compressive strength of recycled aggregate concrete by genetic programming, Comput. Concr. 18 (2) (2016) 155–163.
[45] M. Sarıdemir, Genetic programming approach for prediction of compressive strength of concretes containing rice husk ash, Constr. Build. Mater. 24 (10) (2010) 1911–1919.
[46] A. Baykasoğlu, H. Güllü, H. Çanakçı, L. Özbakır, Prediction of compressive and tensile strength of limestone via genetic programming, Expert Syst. Appl. 35 (1–2) (2008) 111–123.
[47] M. Sarıdemir, Effect of specimen size and shape on compressive strength of concrete containing fly ash: application of genetic programming for design, Mater. Des. 56 (1980–2015) (2014) 297–304.
[48] A. Nazari, Compressive strength of geopolymers produced by ordinary Portland cement: application of genetic programming for design, Mater. Des. 43 (2013) 356–366.
[49] D. Tien Bui, S. M.a.M. Abdullahi, H. Ghareh, H. Moayedi, Nguyen, Fine-tuning of neural computing using whale optimization algorithm for predicting compressive strength of concrete, Eng. Comput. 37 (1) (2021) 701–712.
[50] A. Nazari, J.G. Sanjayan, Modelling of compressive strength of geopolymer paste, mortar and concrete by optimized support vector machine, Ceram. Int. 41 (9) (2015) 12164–12177.
[51] L. Sun, M. Koopialipoor, D. Jahed Armaghani, R. Tarinejad, M. Tahir, Applying a meta-heuristic algorithm to predict and optimize compressive strength of concrete samples, Eng. Comput. 37 (2) (2021) 1133–1145.
[52] M.K. KELEŞ, A.E. KELEŞ, Ü. KILIÇ, Prediction of concrete strength with data mining methods using artificial bee colony as feature selector. 2018 International Conference on Artificial Intelligence and Data Processing (IDAP), IEEE, 2018.
[53] M. Najimi, N. Ghafoori, M. Nikoo, Modeling chloride penetration in self-consolidating concrete using artificial neural network combined with artificial bee colony algorithm, J. Build. Eng. 22 (2019) 216–226.
[54] H. Jahangir, D.R. Eidgahee, A new and robust hybrid artificial bee colony algorithm–ANN model for FRP-concrete bond strength evaluation, Compos. Struct. 257 (2021), 113160.
[55] C. Qi, H.-B. Ly, L.M. Le, X. Yang, L. Guo, B.T. Pham, Improved strength prediction of cemented paste backfill using a novel model based on adaptive neuro fuzzy inference system and artificial bee colony, Constr. Build. Mater. 284 (2021), 122857.
[56] S.A. Moghaddas, M. Nekoei, E.M. Golafshani, A. Behnood, M. Arashpour, Application of artificial bee colony programming techniques for predicting the compressive strength of recycled aggregate concrete, Appl. Soft Comput. (2022), 109641.
[57] M. Komasi, S.A. Hassanzadeh, Evaluation of compressive strength and rapid chloride permeability test of concretes containing metakaolin using Bayesian inference and GEP methods, Modares Civ. Eng. J. 21 (1) (2021) 203–217.
[58] A.S. Hosseini, P. Hajikarimi, M. Gandomi, F.M. Nejad, A.H. Gandomi, Genetic programming to formulate viscoelastic behavior of modified asphalt binder, Constr. Build. Mater. 286 (2021), 122954.
[59] H.A. Algaifi, A.S. Alqarni, R. Alyousef, S.A. Bakar, M.W. Ibrahim, S. Shahidan, M. Ibrahim, B.A. Salami, Mathematical prediction of the compressive strength of bacterial concrete using gene expression programming, Ain Shams Eng. J. 12 (4) (2021) 3629–3639.
[60] H. Alabduljabbar, M. Khan, H.H. Awan, S.M. Eldin, R. Alyousef, A.M. Mohamed, Predicting ultra-high-performance concrete compressive strength using gene expression programming method, Case Stud. Constr. Mater. (2023), e02074.
[61] P. Thamma, S. Barai, Prediction of compressive strength of cement using gene expression programming. Applications of Soft Computing, Springer, 2009, pp. 203–212.
[62] A. Nazari, F.P. Torgal, Modeling the compressive strength of geopolymeric binders by gene expression programming-GEP, Expert Syst. Appl. 40 (14) (2013) 5427–5438.
[63] S. Mahdinia, H. Eskandari-Naddaf, R. Shadnia, Effect of cement strength class on the prediction of compressive strength of cement mortar using GEP method, Constr. Build. Mater. 198 (2019) 27–41.
[64] H.A. Shah, S.K.U. Rehman, M.F. Javed, Y. Iftikhar, Prediction of compressive and splitting tensile strength of concrete with fly ash by using gene expression programming, Struct. Concr. 23 (4) (2022) 2435–2449.
[65] M.F. Javed, M.N. Amin, M.I. Shah, K. Khan, B. Iftikhar, F. Farooq, F. Aslam, R. Alyousef, H. Alabduljabbar, Applications of gene expression programming and regression techniques for estimating compressive strength of bagasse ash based concrete, Crystals 10 (9) (2020) 737.
[66] F. Özcan, Gene expression programming based formulations for splitting tensile strength of concrete, Constr. Build. Mater. 26 (1) (2012) 404–410.
[67] Y. Murad, A. Tarawneh, F. Arar, A. Al-Zu'bi, A. Al-Ghwairi, A. Al-Jaafreh, M. Tarawneh, Flexural strength prediction for concrete beams reinforced with FRP bars using gene expression programming, Structure (2021).
[68] A. Gholampour, A.H. Gandomi, T. Ozbakkaloglu, New formulations for mechanical properties of recycled aggregate concrete using gene expression programming, Constr. Build. Mater. 130 (2017) 122–145.
[69] H.A. Shah, Q. Yuan, U. Akmal, S.A. Shah, A. Salmi, Y.A. Awad, L.A. Shah, Y. Iftikhar, M.H. Javed, M.I. Khan, Application of machine learning techniques for predicting compressive, splitting tensile, and flexural strengths of concrete with metakaolin, Materials 15 (15) (2022) 5435.
[70] H. Sabetifar, M. Nematzadeh, An evolutionary approach for formulation of ultimate shear strength of steel fiber-reinforced concrete beams using gene expression programming, Structure (2021).
[71] A.H. Gandomi, A.H. Alavi, T. Ting, X.-S. Yang, Intelligent modeling and prediction of elastic modulus of concrete strength via gene expression programming. International conference in swarm intelligence, Springer, 2013.
[72] A.H. Gandomi, S.K. Babanajad, A.H. Alavi, Y. Farnam, Novel approach to strength modeling of concrete under triaxial compression, J. Mater. Civ. Eng. 24 (9) (2012) 1132–1143.
[73] A.H. Gandomi, G.J. Yun, A.H. Alavi, An evolutionary approach for modeling of shear strength of RC deep beams, Mater. Struct. 46 (12) (2013) 2109–2119.
[74] S.M. Mousavi, P. Aminian, A.H. Gandomi, A.H. Alavi, H. Bolandi, A new predictive model for compressive strength of HPC using gene expression programming, Adv. Eng. Softw. 45 (1) (2012) 105–114.
[75] J.R. Koza, Genetic programming as a means for programming computers by natural selection, Stat. Comput. 4 (2) (1994) 87–112.
[76] H. Jh, Adaptation in natural and artificial systems, Ann. Arbor. (1975).
[77] A. Ashour, L. Alvarez, V. Toropov, Empirical modelling of shear strength of RC deep beams by genetic programming, Comput. Struct. 81 (5) (2003) 331–338.
[78] Ferreira, C., 2001. Gene expression programming: a new adaptive algorithm for solving problems. arXiv preprint cs/0102027, 2001.
[79] C. Ferreira, Gene expression programming in problem solving. Soft computing and industry, Springer, 2002, pp. 635–653.
[80] C. Ferreira, Automatically defined functions in gene expression programming. Genetic systems programming, Springer, 2006, pp. 21–56.
[81] P. Dunlop, S. Smith, Estimating key characteristics of the concrete delivery and placement process using linear regression analysis, Civ. Eng. Environ. Syst. 20 (4) (2003) 273–290.

[82] GEPSOFT, 2013. GeneXproTools 5.0. 2013.
[83] H.-H. Chu, M.A. Khan, M. Javed, A. Zafar, M.I. Khan, H. Alabduljabbar, S. Qayyum, Sustainable use of fly-ash: Use of gene-expression programming (GEP) and multi-expression programming (MEP) for forecasting the compressive strength geopolymer concrete, Ain Shams Eng. J. 12 (4) (2021) 3603–3617.
[84] A.H. Gandomi, A.H. Alavi, M. Mousavi, S.M. Tabatabaei, A hybrid computational approach to derive new ground-motion prediction equations, Eng. Appl. Artif. Intell. 24 (4) (2011) 717–732.
[85] A.H. Gandomi, D.A. Roke, Assessment of artificial neural network and genetic programming as predictive tools, Adv. Eng. Softw. 88 (2015) 63–72.
[86] K. Turk, S. Demirhan, The mechanical properties of engineered cementitious composites containing limestone powder replaced by microsilica sand, Can. J. Civ. Eng. 40 (2) (2013) 151–157.
[87] J. Zhou, J. Pan, C.K. Leung, Mechanical behavior of fiber-reinforced engineered cementitious composites in uniaxial compression, J. Mater. Civ. Eng. 27 (1) (2015), 04014111.
[88] H. Noorvand, G. Arce, M. Hassan, T. Rupnow, L.N. Mohammad, Investigation of the mechanical properties of engineered cementitious composites with low fiber content and with crumb rubber and high fly ash content, Transp. Res. Rec. 2673 (5) (2019) 418–428.
[89] H. Siad, A. Alyousif, O.K. Keskin, S.B. Keskin, M. Lachemi, M. Sahmaran, K.M.A. Hossain, Influence of limestone powder on mechanical, physical and self-healing behavior of engineered cementitious composites, Constr. Build. Mater. 99 (2015) 1–10.
[90] M. Kamal, S.W. Khan, K. Shahzada, M. Alam, Experimental investigation of the mechanical properties of Engineered Cementitious Composites (ECC), Int. J. Adv. Struct. Geotech. Eng. 5 (2016) 40–45.
[91] H.-L. Wu, J. Yu, D. Zhang, J.-X. Zheng, V.C. Li, Effect of morphological parameters of natural sand on mechanical properties of engineered cementitious composites. Cement and Concrete, Cement Concrete Compos. 100 (2019) 108–119.
[92] K. Ghafor, H.U. Ahmed, R.H. Faraj, A.S. Mohammed, R. Kurda, W.S. Qadir, W. Mahmood, A.A. Abdalla, Computing models to predict the compressive strength of engineered cementitious composites (ECC) at various mix proportions, Sustainability 14 (19) (2022) 12876.
[93] K.M. Hossain, M.S. Anwar, S.G. Samani, Regression and artificial neural network models for strength properties of engineered cementitious composites, Neural Comput. Appl. 29 (2018) 631–645.