Civil Engineering Infrastructures Journal, 50(2): 207 – 219, December 2017
Print ISSN: 2322-2093; Online ISSN: 2423-6691
DOI: 10.7508/ceij.2017.02.001

Genetic Programming Based Formulation to Predict Compressive Strength of High Strength Concrete

Abdollahzadeh, G.R.[1*], Jahani, E.[2] and Kashir, Z.[3]

[1] Associate Professor, Faculty of Civil Engineering, Babol Noshirvani University of Technology, Babol, Iran.
[2] Assistant Professor, Faculty of Civil Engineering, University of Mazandaran, Babolsar, Iran.
[3] M.Sc. Student, Faculty of Civil Engineering, Tabari University of Babol, Babol, Iran.

Received: 05 Mar. 2016; Revised: 09 Mar. 2017; Accepted: 15 Mar. 2017

ABSTRACT: This study introduces, two models based on Gene Expression Programming (GEP) to predict compressive strength of high strength concrete (HSC). Composition of HSC was assumed simplified, as a mixture of six components (cement, silica fume, super-plastisizer, water, fine aggregate and coarse aggregate). The 28-day compressive strength value was considered the target of the prediction. Data on 159 mixes were taken from various publications. The system was trained based on 80% training pairs chosen randomly from the data set and then tested using remaining 20% samples. Therefore it can be proven and illustrated that the GEP is a strong technique for the prediction of compressive strength amounts of HSC concerning to the outcomes of the training and testing phases compared with experimental outcomes illustrate that the.

Keywords: Compressive Strength, Gene Expression Programming, HSC, Silica Fume.

INTRODUCTION

Achieving to high compressive strength in concrete, has been one of the main purposes in civil engineering. According to ACI, high strength concrete (HSC) has compressive strength higher than 42 MPa. However, most of the concrete regulations limited compressive strength of structural concrete to 60 MPa (BS 8110-1, 1997). Therefor the scope of HSC using, is not very broad. The basic ingredients in HSC mixtures are similar to conventional concrete, however, minerals and chemicals addition added to mixtures to raise compressive strength.

Compressive strength as a critical property of HSC quality, depend on various factors such as concrete mix design, the kind of materials forming the concrete, person's skills for testing, laboratory errors, and so on. Since many of these factors are unknown, it is difficult to obtain accurate formulation for concrete strength, so, using the method except mathematical formula to predict the strength of concrete to an acceptable level, will be important (Samayiinejaad, 2001). To predict the manner of HPC is relatively difficult compared to predicting conventional concrete manner. Therefore, traditional model of concrete properties is not

* Corresponding author E-mail: abdollahzadeh@nit.ac.ir

appropriate for analyzing HPC compressive strength (Castelli et al., 2013).

Predicting HPC behavior is relatively difficult compared to predicting conventional concrete behavior. Chou and Tsai (2012) explained the relationship between ingredients and concrete properties is highly nonlinear so it can be said certain properties of HPC are not completely found. Therefore, traditional model of concrete properties is not appropriate for analyzing HPC compressive strength. Totally, regression analysis can be used for empiric modeling of experimental outcomes of concrete parameters. Recently, addition to classical regression techniques, soft computing applications such as neural networks and Gene Expression Programming (GEP) can predict the explicit formulations of the characteristics and the performances of concrete (Abdollahzadeh et al., 2016; Gandomi et al., 2014; Saridemir, 2014). The neural networks based formulation are often too complex to be used as can be seen in empiric formularization of experimental investigations. On the other hand, genetic programming have many advantages in this respect as compared to classical regression techniques. Firstly, some functions define for regression technique and then analyses of these functions are performed. On the other hand, in GEP approach, there is no predefined function to be considered, i.e. GEP adds or deletes various combinations of parameters for considering the formulation that fits the experimental outcomes. In this sense, GEP can be accepted to be superior to regression techniques and neural networks. Where no analytical models exist, to model and obtain clear formularization of experimental investigations containing multivariate parameters, GEP has been confirmed to be a successful system (Gandomi et al., 2014; Saridemir, 2011).

A research about using of artificial neural network (ANN) and genetic programming (GP) to predict split tensile strength and water permeability of HSC containing TiO_2 nanoparticles were performed by Nazari and Riahi (2011). A collection of 144 samples produced with 16 varied mixtures for purpose of making ANN and GP modeling, were used by them to set training and testing phases. Eight input parameters that include the cement content (C), nanoparticle content (N), aggregate type (AG), water content (W), the amount of superplasticizer (S), the type of curing medium (CM), Age of curing (AC) and number of testing try (NT) have been used as data in the multilayer feed forward neural networks models and input variables of genetic programming models . ANN and GP models have been discovered to be reliable in the scope of variables. Also, ANN and GP are efficient for predicting the split tensile strength of TiO_2 nanoparticles concrete (Nazari and Riahi, 2011).

Castelli et al. (2013) predicted high performance concrete strength using Genetic Programming with geometric semantic genetic operators. The system they proposed was based on recently defined geometric semantic genetic operators for Genetic Programming. They tested the proposed implementation of GP with geometric semantic operators (GS-GP from now on) against a standard GP system (ST-GP). Experimental data from 17 different sources was used to check the reliability of the strength model. A total of 50 runs were performed with each technique. In each run a different partition between training and test data has been considered. 70% of the samples have been applied as training data, while the remainings have been applied as test data. Experimental outcomes show the suitability of the proposed system for the prediction of concrete strength. In particular, the new method provides significantly better outcomes than the ones produced by standard Genetic Programming and other machine learning methods, both on training and on out-of-sample data (Castelli et al., 2013).

Although there are many modeling works about properties of HSC, but the main purpose of this paper is to model and formulate the main mechanical properties of HSC, compressive strength at 28 days that is most used for quality control, by genetic programming.

MODELING PHASE

Data Set

To develop GEP models, 157 samples with different mixtures were collected from a M.S. thesis (Samayiinejaad, 2001). 80% of samples randomly used in training phase and 20% used in testing phase. This information includes the weight of concrete's components (cement, water, fine aggregates, coarse aggregates, silica-fume and super-plasticizer) and compressive strength of 28 days. Quantities about inputs and output amounts was presented in Table 1.

Gene Expression Programming Models and Parameters

Ferreira invented a new algorithm called Gene Expression Programming (GEP) in 1999. This algorithm (GEP) is incorporation of Genetic Algorithms (simple, linear chromosomes of fixed length) and Genetic Programming (non-linear entities of different sizes and shapes). Therefore, with GEP, the second evolutionary threshold – the phenotype threshold – is crossed, creating a new range of possibilities in evolutionary computation. This is corresponding to say that, in GEP, the genotype and phenotype are finally separated from one another, since the non-linear entities of different sizes and shapes are completely encoded in the linear chromosomes of fixed length and the system can now benefit from all the evolutionary advantages this produce (Ferreira, 2001).

Thus, the phenotype of GEP consists of the same kind of parse trees used in Genetic Programming. But the parse trees evolved by GEP (called expression trees) are the expression of a totally independent genome. Consequently, with GEP, a notable thing occurred: the second evolutionary threshold – the phenotype threshold – was crossed (Dawkins, 1995). And this means that merely the genome (slightly modified) is passed on to the next generation (Ferreira, 2001).

The fundamental steps of Gene Expression Programming are schematically represented in Figure 1 (Ferreira, 2006). To construct a GEP model, five components; the function set, terminal set, fitness function, control parameters and stop condition are needed. After encoding the problem for candidate solution and specifying the fitness function, the algorithm randomly creates an initiative population of viable individuals (chromosomes) and then transforms the each chromosome into an expression tree corresponding to a mathematical expression. Thereafter the predicted target is compared with the actual one and the fitness score for each chromosome is determined. If it is sufficiently good, the algorithm stops.

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

Input Variables	**Data Used in the Models**	
	Minimum	**Maximum**
Cement (kg/m^3)	245	610
Water (kg/m^3)	106	246
Fine aggregates (kg/m^3)	378.6	1204
Coarse aggregates (kg/m^3)	421	1239
Silica-fume (kg/m^3)	0	84.6
Superplasticizer (kg/m^3)	0	27.8
Output variable		
Compressive strength (MPa)	40	113.5

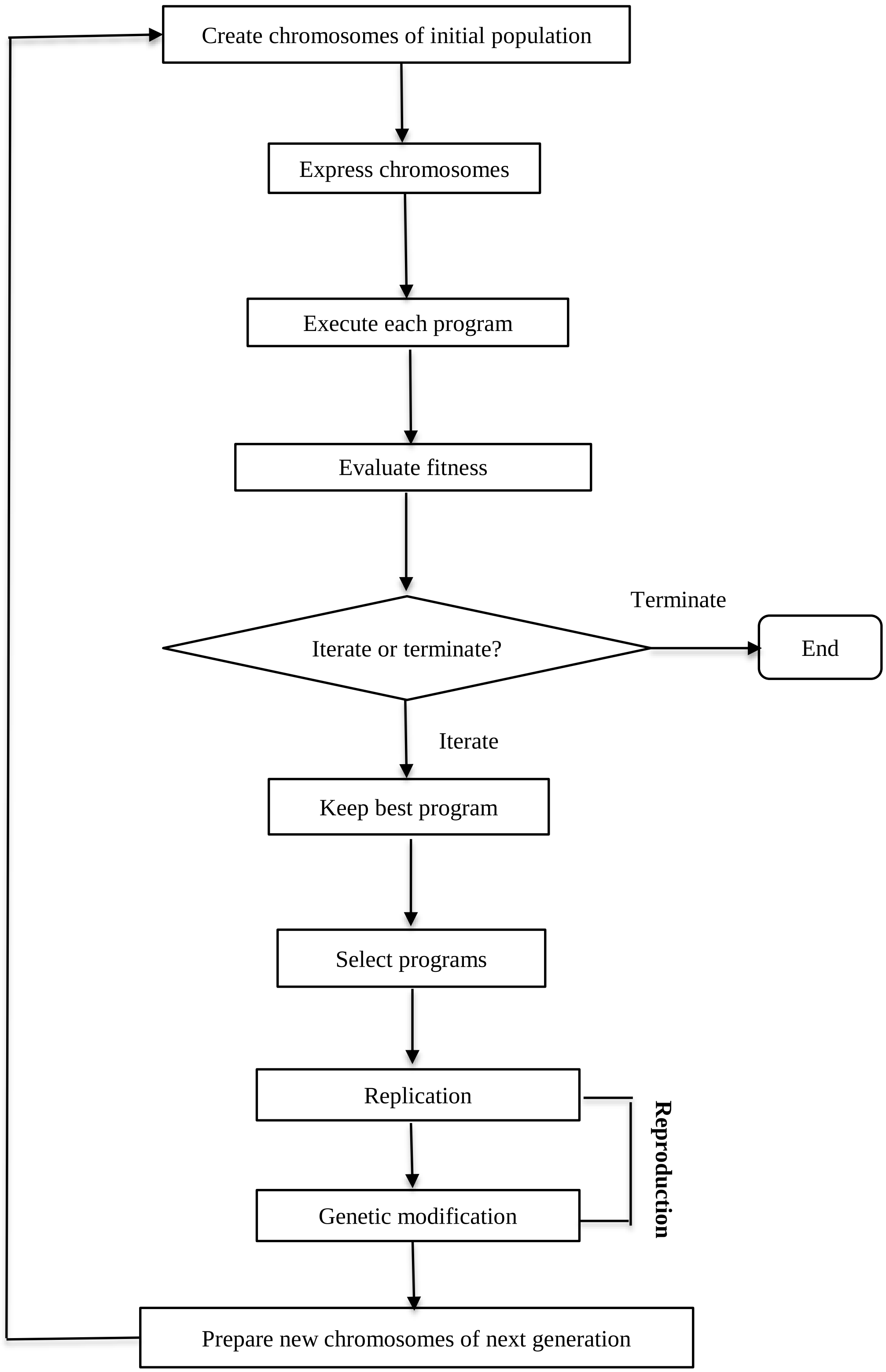


Fig. 1. Flowchart of Gene expression programming (Ferreira, 2006)

Otherwise, some of the chromosomes are selected using roulette wheel sampling and then mutated to obtain the new generations. This closed loop is continued until desired fitness score is achieved and then the chromosomes are decoded for the best solution of the problem (Kayadelen et al., 2009; Teodorescu and Sherwood, 2008).

Explicit formulations obtained from GEP compared to classical regression techniques, causes using of this approach in many areas of engineering field, recently. It is believed and proven for modeling and obtaining clear formulations of experimental studies, like multivariate problems, GEP is more powerful than regression techniques and neural networks (Milani and Nazari, 2012; Nazari et al., 2012; Bhargava et al., 2011; Podgornik et al., 2011; Ganguly et al., 2009).

Every matters in GEP algorithm are displayed by expression trees (ETs) that consist of operators, functions, constants and variables. An algebraic expression can be represented by two genes chromosome or an ET, as shown in Figure 2. This figure shows a chromosome with two genes is encoded as a linear string and also how it is expressed as an ET.

In this paper, we predict compressive strength of HSC by GEP modeling. So two models of GEP, namely GEP-I and GEP-II constructed to predict compressive strength of HSC and at end, modeling outcomes compared with experimental outcomes. In the GEP-I and GEP-II, as the number of genes used 3 and 4 genes (Sub-ETs), and as linking function used multiplication and addition, respectively. For using GEP there are five major steps.

First of all, is choosing fitness function, which in this problem, we measured the fitness f_i by using two following expressions:

$$f_i = \sum_{j=1}^{C_t} \left(M - |C_{(i,j)} - T_j|\right) \quad (1)$$

where M: is the range of selection, $C_{(i,j)}$: the value returned by the individual chromosome i for fitness case j (out of C_t fitness cases) and T_j: is the target value for fitness case j. if $\left|C_{(i,j)} - T_j\right|$ (the precision) less or equal to 0.01, then the precision is equal to zero, and $f_i = f_{max} = C_t M$. In this case, M = 100 was used, therefore f_{max} = 1000. The advantage of this kind of fitness functions is that the system can find the optimal solution by itself (Teodorescu and Sherwood, 2008).

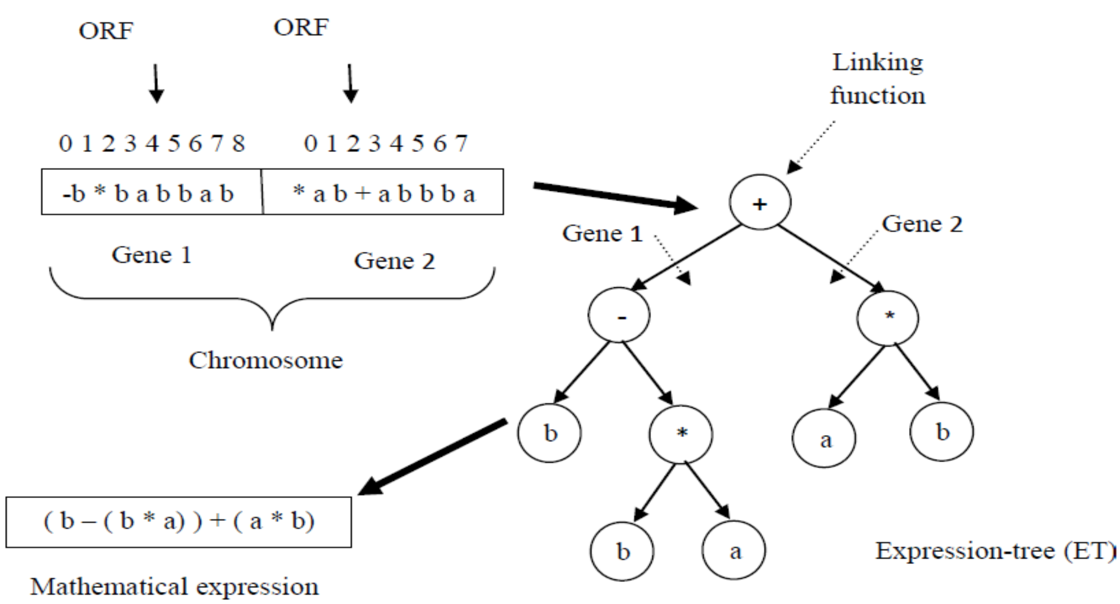


Fig. 2. Chromosome with two genes and its decoding in GEP (Kayadelen et al., 2009; Teodorescu and Sherwood, 2008)

The second major step is selection of terminals T and functions F to create chromosomes. In this case terminals set include of the independent variable, i.e., T = $\{C, W, FA, CA, MC, SP\}$ and about functions, four basic arithmetic operators$(+, -, *, .)$ and some basic mathematical functions $(Sqrt, x^3, \ldots)$ were used the models.

The third major step is to choose the chromosomal architecture, namely the length of the head and the number of genes. In this case we used 3 and 4 genes and length of heads 8 and 12 for models GEP-I and GEP-II respectively.

The fourth major step to use GEP is selection the linking function to link the sub-ETs which in this problem we used multiplication and addition.

And finally, all genetic operators (mutation, transposition and crossover…) combine and was applied as set of genetic operators. All details of used parameters were presented in Table 2.

Explicit formulations based on the approach models for f_c were obtained by Eq. (2).

$$f_c = f(C, W, FA, CA, MS, SP) \tag{2}$$

For the GEP-I and GEP-II approach models, Figures 2 and 3 show the expression trees of Eqs. (3) and (4), respectively. In these Equations, d0, d1, d2, d3, d4 and d5 refer to cement, coarse aggregate, fine aggregate, silicafume, superplastisizer and water respectively. The constant for formulation of each models specified by applied software.

$$f_c = \left[\sqrt[2]{\big((d2 + d0) + d2\big) * (d0/(d5 + d5))}\right]$$
$$\left[\sqrt[2]{d3} + ((d0 + d0)/d1)^3\right] + d4\Big]$$
$$\left[d3 * \sqrt[2]{(d4/(-8.09 + d5) + (9.05 * d4))}\right] \tag{3}$$

$$f_c = \sqrt[3]{\sqrt[2]{d2}}$$
$$* \sqrt[3]{\begin{matrix}(d2 - [(((d1 * d4) + d5^2)/\ 5.06^2)) \\ -((5.06 + \ 5.06) * (5.06 + d4))]\end{matrix}}$$
$$*$$
$$\left[\sqrt[2]{\sqrt[2]{\mathrm{d2}} + \sqrt[2]{\sqrt[2]{3.53} * (\mathrm{d3} - 3.53) * (\mathrm{d2} * \mathrm{d3})}^{3}}^{3}\right]$$
$$\left[(-4.35/{-4.35})^3\right] \tag{4}$$

Table 2. Parameters of GEP approach models

Parameter Definition	GEP-I	GEP-II
Function set	$+, -, *, /, \mathrm{Sqrt}, x^2, x^3, 3\mathrm{rt}$	$+, -, *, /, \mathrm{Sqrt}, x^2, x^3, 3\mathrm{rt}$
Chromosomes	30	40
Head size	8	12
Number of genes	3	4
Linking function	Addition	Multiplication
Mutation rate	0.044	0.044
Inversion rate	0.1	0.1
One-point recombination rate	0.3	0.3
Two-point recombination rate	0.3	0.3
Gene recombination rate	0.1	0.1
Gene transposition rate	0.1	0.1

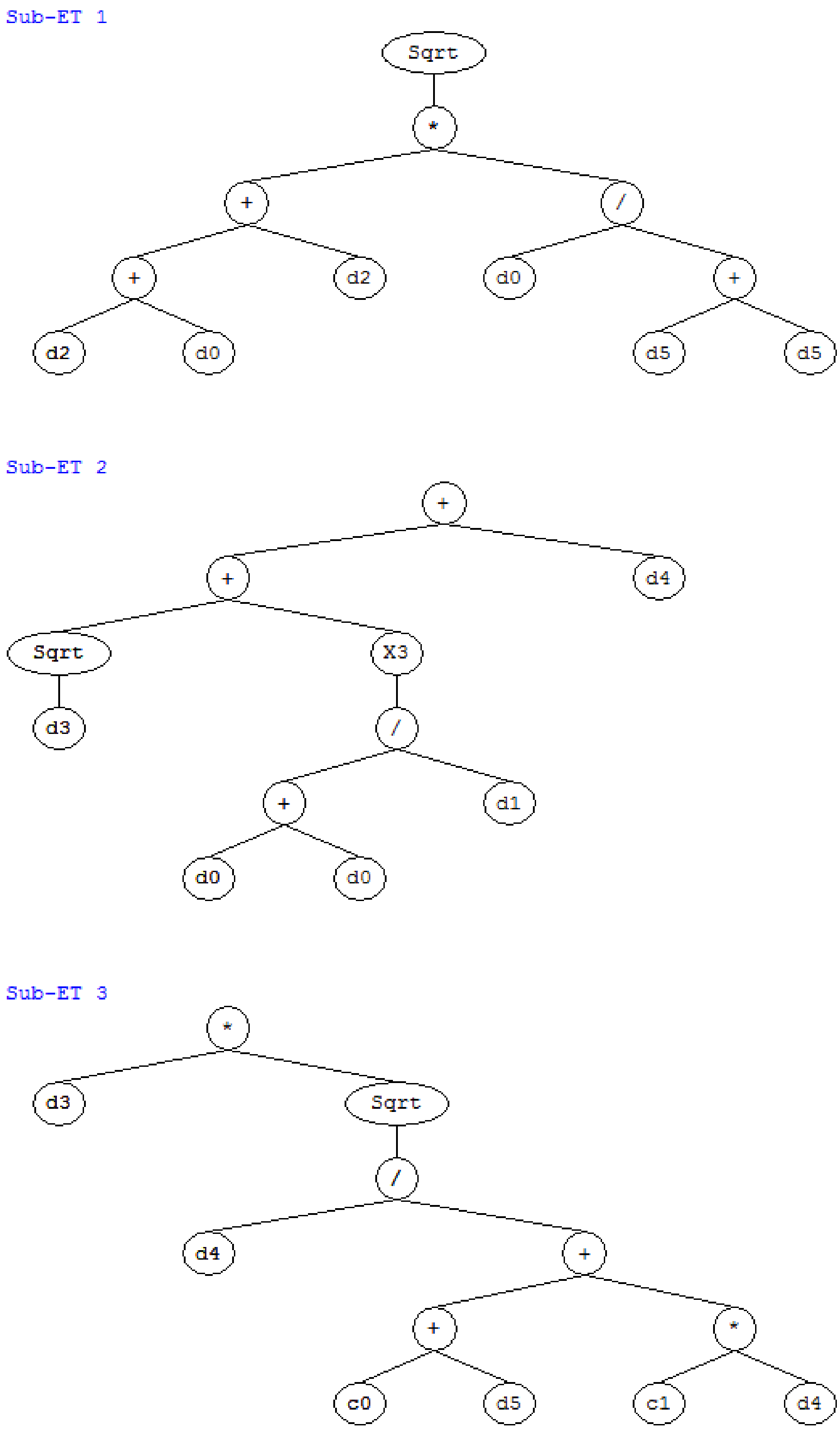


Fig. 3. Expression trees of GEP-I approach model

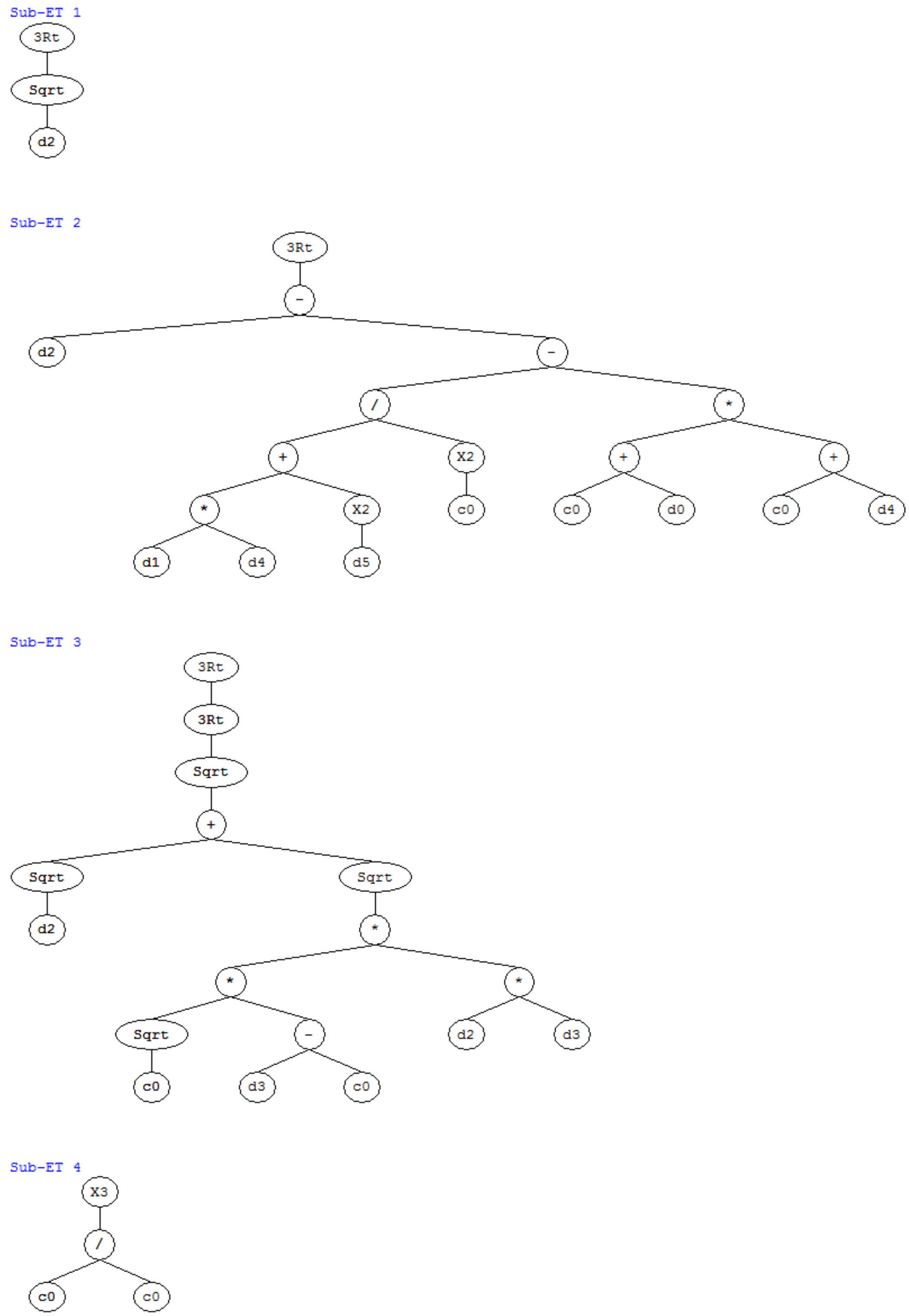


Fig. 4. Expression trees of GEP-II approach model

OUTCOMES AND DISCUSSION

Absolute fraction of variance (R^2) mean absolute error (*MAE*), root mean square error (*RMSE)* were presented in this paper as statistical evaluations for inevitable errors while training and testing the models according to the Eqs. (5-7), respectively.

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

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|t_i - o_i| \quad (6)$$

$$RMSE = \sqrt{\frac{1}{n}\sum\nolimits_{i=1}^{n}(t_i - o_i)^2} \quad (7)$$

Here *t*: is the target value, *o*: is the output value and *n*: is the number of all collected data. Statistical errors amounts for both training and testing the models were displayed in Table 4. If R^2 amounts are above 0.7 and closer to 1, predicted outcomes are closer to experimental outcomes. Also, if (*MAE, RMSE*) amounts increase, reduce models performance. There were few differences between experimental and predicted amounts statistically.

The obtained outcomes by experimental investigations and predicted amounts in training and testing phases of the GEP-I, GEP-II models, are displayed on Figures 3-4, respectively. The linear least square fit and fit line and the model R^2 amounts are displayed on Figures 5-6 for the training and testing phases. Also, inputs amounts and experimental outcomes with testing outcomes obtained from models were given and compared in Table 3.

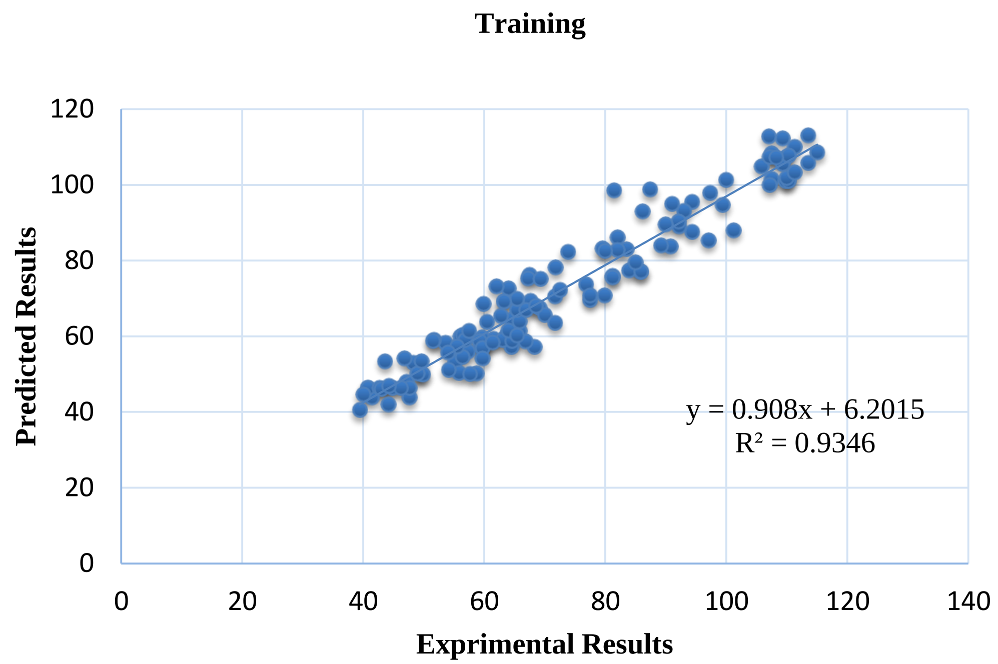
All outcomes show GEP is also a good approach for predicting of f_c amounts of HSC.

Table 3. GEP models outcomes compared with experimental outcomes are used as test sets

Data Used in Models Construction						Compressive Strength (MPa)		
Cement (kg/m³)	**Water** (kg/m³)	**Fine Aggregate** (kg/m³)	**Coarse Aggregate** (kg/m³)	**Silicafume** (kg/m³)	**Super-plastisizer** (kg/m³)	**Exp**	**GEP-I**	**GEP-II**
450	130	1187	667	50	19.5	105.8	111.7209747	111.936117
580	140	620	1025	70	13.3	103	100.5882589	97.04527961
550.6	138	612	770	0	6.3	66.5	68.72457022	63.24775205
413	190	767	1092	0	0	40	46.43367175	41.3984819
548	191.8	680	1020	0	4.7	56	58.14897191	58.48259891
320	148	750	1175	25	3.9	62.6	57.14877602	60.69051039
474.8	156.7	603.5	1127.5	47.5	9	85	76.36106043	79.46417193
402	188	643.1	1094.9	15.7	5.1	57.8	54.30094316	61.0633902
610	152.5	697	1045	0	16.3	81	81.1999886	82.35691674
350	195	749	1092	0	1.1	46	42.0876195	40.56048595
446	223.2	660	990	50	0.9	54	53.88215476	49.55817726
266	161	873	1100	40.3	3.7	67.5	56.60796513	59.39898544
391.5	178	700	1097	0	0.9	50	45.64997016	43.38333825
404	208	1086	726	0	0	52.4	51.3954756	44.23148742
560	155.3	698	1047	61	14.8	100	97.21441046	99.53713846
450	130	1231	623	45	19.5	108.2	111.7190627	112.3155351
450	135	770	1025	50	8.5	90.8	84.04857129	83.06491851
426	184	1148	768	0	8.5	67.2	65.99918773	69.11265227
362.7	178	711	1100	0	0.5	40	43.42817747	41.0158857
425	190	730	1000	0	4.3	50.7	50.82952203	52.98314309
413	190	767	1092	0	0	46.9	46.43367175	41.3984819
422	125	643	1172	42.2	20.5	93.3	92.05027679	94.80035971
385.7	137.5	587.5	1216.6	57.9	11.1	85.7	78.47328342	78.31607473
563.8	148	440.3	1216.6	28.2	14.8	76	79.91381758	86.53620814
519	120	725	1120	41	14.9	94.1	97.42666008	96.75327086
425	160	934	1172	0	12.8	63.9	68.36653989	72.42301954
342	171	670	1200	0	0	43.5	41.19738631	38.52545594
465	162	1204	450	0	2.7	66.5	75.73978907	61.87168524

Table 4. The f_c statistical amounts of GEP-I and GEP-II approach models

Statistical Parameters	GEP-I		GEP-II	
	Training Set	**Testing Set**	**Training Set**	**Testing Set**
R^2	0.9345	0.9511	0.9082	0.9494
MAE	4.4259	3.8290	5.072	4.5458
RMSE	5.5088	4.7520	6.4659	5.2604



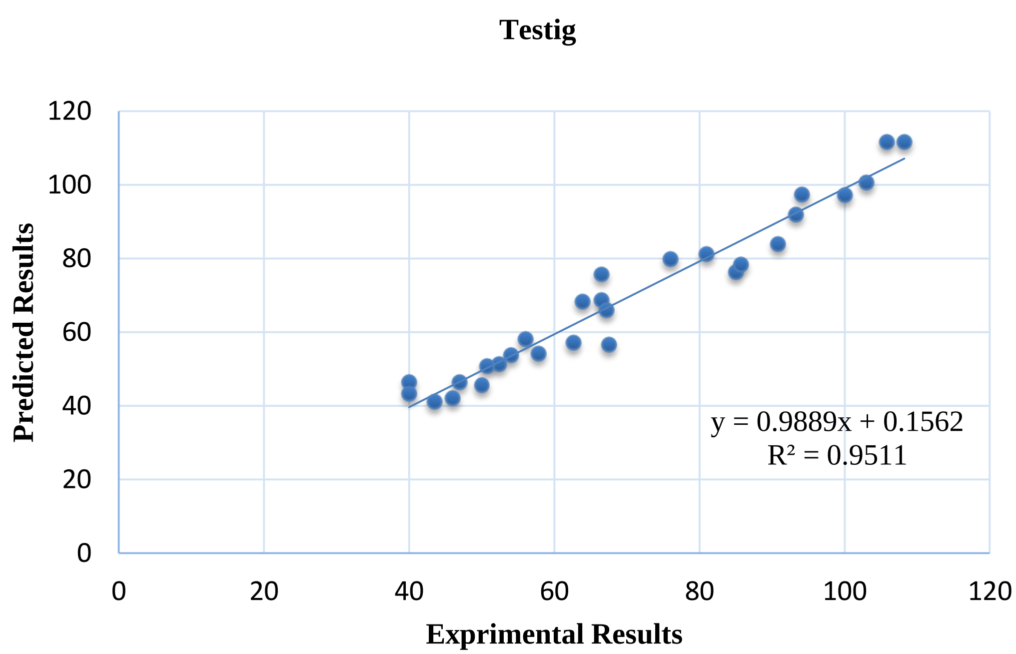
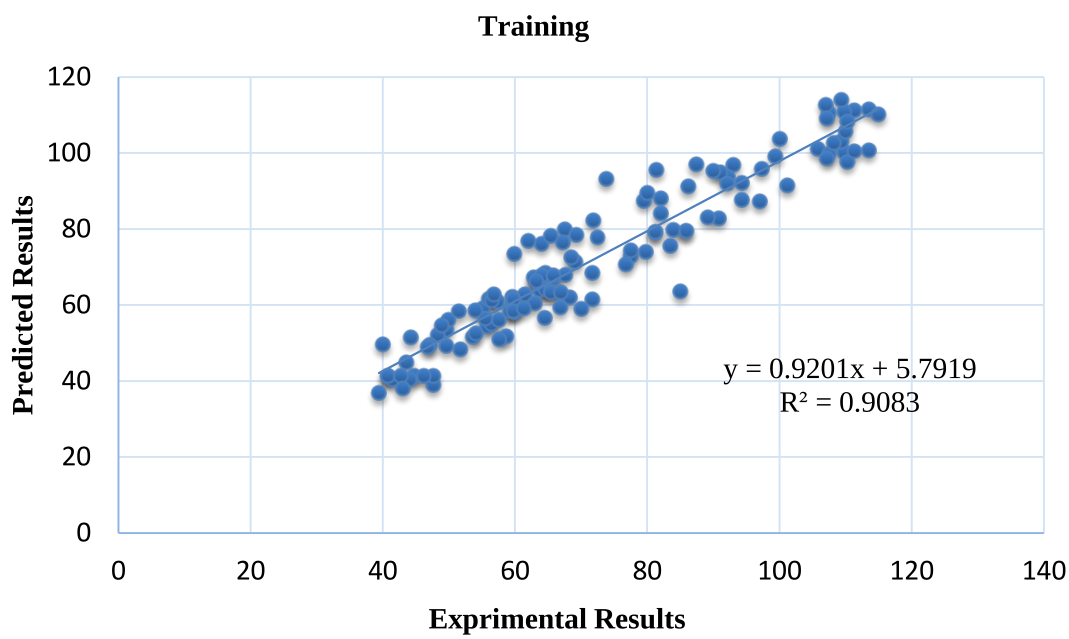


Fig. 5. Scattering diagram of predicted vs. experimental for training and testing models of GEP-I



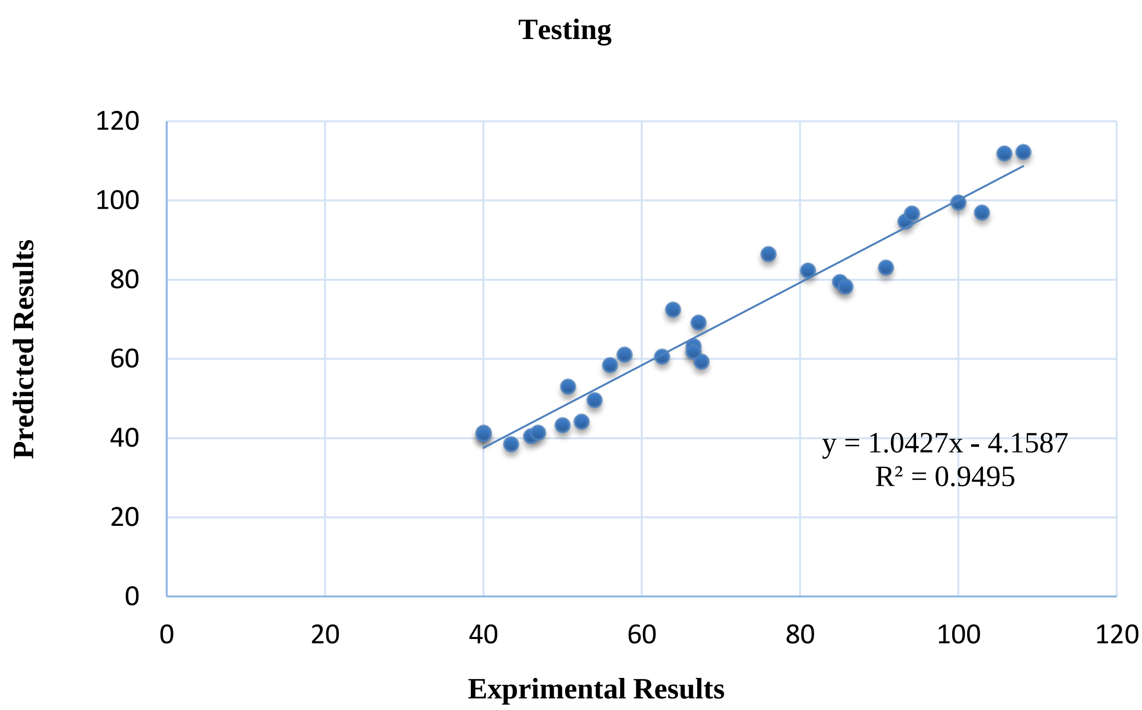


Fig. 6. Scattering diagram of predicted vs. experimental for training and testing models of GEP-II

This situation can be confirmed by the obtained R^2 amounts given on the figures. Furthermore, the statistic parameter outcomes of all the GEP models are given in Table 4. The outcomes of statistical analysis in the training and testing phases of the GEP models illustrate a good relevancy between the input parameters and the f_c amounts of HSC at various proportions. In Table 4, as can be noticed, R^2 of the models for the training and testing phases are higher than 0.90. The best value of R^2 is 0.9511 for testing phase in the GEP-I model, while the minimum value of R^2 is 0.9082 for training phase in the GEP-II model. All of the statistical parameter amounts noticed in Table 4 illustrate that the proposed models of GEP predict the f_c amounts of HSC with good accuracy. Accordingly, it can be concluded GEP provide a successful alternative way to the experimental studies or artificial neural networks and fuzzy logic used to model the f_c of all kind of concrete with different mixtures.

One of the other outcomes obtained from these models was Eqs. (3) and (4) that was brought in last section. These equations can be used for designing different mixers of HSC as it has been done before in literature about another concrete (Sarıdemir, 2014).

CONCLUSIONS

According to the obtained outcomes, GEP can be as an appropriate tool for modeling the compressive strength amounts of HSC at different proportions. The models, GEP-I and GEP-II, with different parameters are proposed to predict the compressive strength amounts of concrete. The experimental outcomes from a widely spread database of compressive strength amounts of HSC that have been already published, are used for developing the models. The models have been discovered be highly capable to predict the compressive strength amounts of concrete in connection with cement, water, fine aggregate, coarse aggregate, silica fume and super-plasticizer. The suggested GEP models can predict the compressive strength amounts of HSC for each of training and testing phases according to the statistical parameters of R^2, MAE and RMSE.

REFERENCES

Abdollahzadeh, G.R., Jahani, E. and Kashir, Z. (2016). "Predicting of compressive strength of recycled aggregate concrete by genetic programming", *Computers and Concrete*, 18(2), 155-163.

Bhargava, S., Dulikravich, G.S., Murty, G.S., Agarwal, A. and Colaco, M.J. (2011). "Stress corrosion cracking resistant aluminum alloys: Optimizing concentrations of alloying elements and tempering", *Materials and Manufacturing Processes*, 26(3), 363-374.

BS 8110-1 (1997). Code of practice for structural use of concrete Part 1: Design and construction, British Standard, BSI, London.

Castelli, M., Vanneschi, L. and Silva, S. (2013). "Prediction of high performance concrete strength using genetic programming with geometric semantic genetic operators", *Expert Systems with Applications*, 40(17), 6856-6862.

Chou, J.S. and Tsai C.F. (2012). "Concrete compressive strength analysis using a combined classification and regression technique". Automation in Construction 24, 52-60.

Ferreira, C. (2006). "Automatically defined functions in gene expression programming, In Genetic Systems Programming", *Springer Berlin Heidelberg*, 21-56.

Ferreira, C. (2001). "Gene expression programming: a new adaptive algorithm for solving problems", *Complex Systems*, 13(2), 87-129.

Gandomi, A.H., Mohammadzadeh, D., Pérez-Ordóñez, J.L. and Alavi, A.H. (2014). "Linear genetic programming for shear strength prediction of reinforced concrete beams without stirrups", *Applied Soft Computing*, 19, 112-120.

Ganguly, S., Datta, S. and Chakraborti, N. (2009). "Genetic algorithm-based search on the role of variables in the work hardening process of multiphase steels", *Computational Materials Science*, 45(1), 158-166.

Kayadelen, C., Günaydın, O., Fener, M., Demir, A. and Özvan, A. (2009). "Modeling of the angle of shearing resistance of soils using soft computing systems", *Expert Systems with Applications*, 36(9), 11814-11826.

Milani, A.A. and Nazari, A. (2012). "Modeling ductile-to-brittle transition temperature of functionally graded steels by gene expression programming", *International Journal of Damage Mechanics*, 21(4), 465-492.

Nazari, A., Khalaj, G. and Var, N.D. (2012). "Computational investigations of the impact resistance of aluminum-epoxy-laminated composites", *International Journal of Damage Mechanics*, 12, 623-646.

Nazari, A. and Riahi, S. (2011). "Prediction split tensile strength and water permeability of high strength concrete containing TiO_2 nanoparticles by artificial neural network and genetic programming", *Composites Part B: Engineering,* 42(3), 473-488.

Podgornik, B., Leskovšek, V., Kovačič, M. and Vižintin, J. (2011). "Residual stress field analysis and prediction in nitrided tool steel", *Materials and Manufacturing Processes*, 26(9), 1097-1103.

Sarıdemir, M. (2014). "Effect of specimen size and shape on compressive strength of concrete containing fly ash: Application of genetic programming for design", *Material and Design*, 56, 297-304.

Sarıdemir, M. (2011). "Empirical modeling of splitting tensile strength from cylinder compressive strength of concrete by genetic programming", *Expert Systems with Applications*, 38(11), 14257-14268.

Samayiinejaad, A. (2001). "The application of neural networks and genetic algorithms to determine the high-strength concrete mix design", M.Sc. thesis, Isfahan University of Technology, Isfahan, Iran.

Teodorescu, L. and Sherwood, D. (2008). "High energy physics event selection with gene expression programming", *Computer Physics Communications*, 178(6), 409-419.