



# New empirical formulations for indirect estimation of peak-confined compressive strength and strain of circular RC columns using LGP method

Mohamad Fazel Rostami<sup>1</sup> · Ehsan Sadrossadat<sup>2</sup> · Behnam Ghorbani<sup>2</sup> · Seyed Morteza Kazemi<sup>1</sup>

Received: 2 August 2017 / Accepted: 2 January 2018 / Published online: 9 January 2018  
© Springer-Verlag London Ltd., part of Springer Nature 2018

## Abstract

Reinforced concrete (RC) columns have been basically designed to withstand compressive loads by means of strain and ductility of the longitudinal and transverse reinforcing materials. The objective of this paper is to propose new predictive models of confined compressive strength and strain at confined peak stress of circular-reinforced concrete columns using a powerful evolutionary-based computational technique, namely, linear genetic programming (LGP). For this aim, a collection of data is utilized to develop new models. The models obtained in this study characterize peak-confined compressive strength and corresponding strain factors in terms of the compressive strength of unconfined concrete cylinder specimens, core diameter of circular column, yield strength of transverse reinforcement, ratio of volume of lateral reinforcement to volume of confined concrete core, spacing of lateral reinforcement or spiral pitch, and ratio of longitudinal steel to area of core of section in addition to the column height. These factors have also been considered as the most significant input variables in several models proposed by scholars in the existing literature for approximation of the peak-confined compressive strength and corresponding strain of RC columns. To evaluate the validity of the obtained models, several analyses are conducted and the results are compared with those provided by other researchers to validate and verify the capability of the proposed models. Consequently, the results explicitly approve that the proposed models are of a notably better performance than the traditional models in the literature.

**Keywords** Confined compressive strength · Strain at confined peak stress · Reinforced concrete columns · Indirect estimation · Linear genetic programming

## 1 Introduction

In general, columns are regarded as the most important structural members that resist under applied loads via deformation, strain, and ductility and transmit them to underlying supports. Concrete columns are mainly designed to withstand compressive loads due to the inherent properties of concrete. The failure of such members as subjected by lateral loads is mainly due to the lack of sufficient shear bearing capacity and enough ductility. The strength and ductility

characteristics of concrete columns can be enhanced by means of reinforcing with other materials such as steel, particularly, when they are subjected to lateral loads such as seismic and wind loads. Column confinement with transverse reinforcement, commonly in the form of closely spaced steel cross-ties, hoops, or spirals, substantially affect the strength and ductility of reinforced concrete (RC) structural members [1–3]. In other words, the most significant design consideration for ductility in plastic hinge regions of RC columns is the provision of sufficient transverse reinforcement [1, 2].

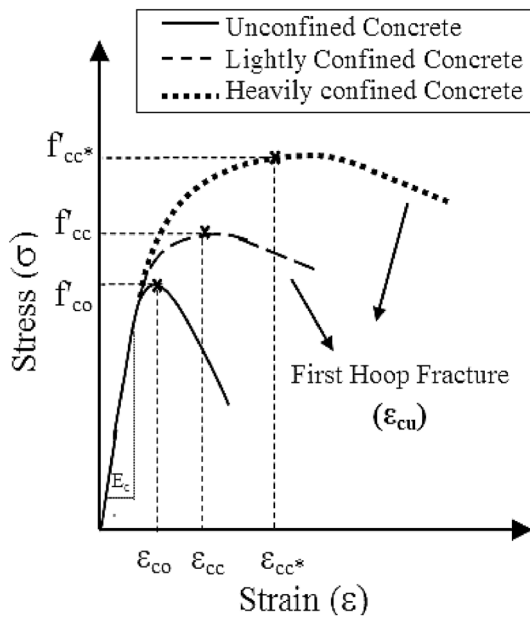
The relationship between the stress and strain for RC columns may be achieved through experimental studies by recording the amount of the column strain at specific intervals of applying stress. The result of such experiments can be demonstrated as a stress–strain curve (see Fig. 1).

As illustrated in Fig. 1,  $E_c$  is the concrete modulus of elasticity, AND  $\epsilon_{co}$  and  $\epsilon_{cc}$  are the corresponding strains to

✉ Ehsan Sadrossadat  
ehsan.sadrossadat@mshdiau.ac.ir

<sup>1</sup> Department of Civil Engineering, Kashmar Branch, Islamic Azad University, Kashmar, Iran

<sup>2</sup> Young Researchers and Elite Club, Mashhad Branch, Islamic Azad University, Mashhad, Iran



**Fig. 1** Stress–strain curve for axial loading of confined and unconfined concrete specimens

$f'_{co}$  and  $f'_{cc}$  that are referred to the peak concrete strengths which represent the maximum load carried by the RC column for unconfined and confined cases, respectively. In pre-peak phase, the form of the stress–strain curve for confined concrete is mainly controlled by three parameters, i.e.,  $f'_{cc}$ ,  $\epsilon_{cc}$ , and  $E_c$  [1, 2]. Based on the rate of confinement, it can be figured out that load-bearing capacity and corresponding ductility or strain in stress–strain curve grow with increase of the transverse confinement [4]. Besides, it can be seen that the confinement reinforcement increases the compressive strength of the concrete specimens and causes a more ductile response until reaching the ultimate concrete compressive strain of the concrete,  $\epsilon_{cu}$ , where the first hoop fractures which mainly occur in the post-peak phase [1, 2].

To provide an efficient design, engineers are required to precisely realize the stress–strain behavior of RC columns which are the most important members in RC buildings and structures. In this regard, researchers may evaluate the problem through experimental studies to examine theories and hypotheses through varying and manipulating the variables of the experiment to realize how they affect the system behavior. However, experimental studies in this major are limited due to the high cost, cumbersome procedure of tests and lack of high-capacity testing equipment. Therefore, over the last few decades, several researchers have proposed various analytical-based models for indirect estimation of  $f'_{cc}$  and corresponding  $\epsilon_{cc}$  of confined RC columns considering various variables affecting the system behavior which mainly include compressive strength of unconfined concrete, the amount, placement, and configuration of longitudinal

and transverse reinforcement, ratio of the core and gross cross-sectional areas, yield strength of transverse reinforcement, spacing of hoops or lateral ties, cross-sectional shape, e.g., circular or noncircular and dimensions of the structural member and the applied load [1, 2, 5–7].

Mander et al. [1] conducted experimental and theoretical studies on circular, rectangular, and square full-scale RC columns at seismic strain rates to explore the effect of transverse reinforcement on confinement behavior and overall performance of concrete compressive strength and strain of RC columns. Thereafter, Mander et al. [2] tried to find an optimal model using their test results. According to their study, their obtained model for confined RC columns with circular cross section is as follows:

$$f'_{cc} = f'_{co} \left( -1.254 + 2.254 \sqrt{1 + 7.94 f'_1 / f'_{co} - 2 f'_1 / f'_{co}} \right) \quad (1)$$

where  $f'_1 = k_e \rho_s f_{yh} / 2$  and  $k_e = (1 - s/2d)^n / (1 - p_{cc})$ .

in which  $n$  must be considered equal to 2 and 1, where circular hoops and spirals are used, respectively:

$$\epsilon_{cc} = \epsilon_{co} [1 + 5(f'_{cc} / f'_{co} - 1)] \quad (2)$$

where  $\epsilon_{co}$  should be considered equal to 0.002 if not available.

Saatcioglu and Razavi [4] observed that high-strength concrete columns needed more lateral ties to maintain the ductility than normal-strength concrete columns. It was observed that confined concrete strength and corresponding strain should be expressed in terms of equivalent uniform confinement pressure provided by the transverse reinforcement cage. The equivalent uniform pressure is obtained from average lateral pressure computed from sectional and material properties. Besides, the strain level was expressed in terms of confinement parameters [4]:

$$f'_{cc} = f'_{co} + k_1 (\rho_s f_{yh} / 2) \quad (3)$$

$$\epsilon_{cc} = \epsilon_{co} [1 + 5k_1 [\rho_s f_{yh} / (2f'_{co})]] \quad (4)$$

where  $k_1 = 6.7 (\rho_s f_{yh} / 2)^{-0.17}$  and  $\epsilon_{co}$  should be considered equal to 0.002 if not available.

Hoshikuma et al. [6] proposed a stress–strain model of concrete that takes confinement effects into account is developed, based on test results of a series of compression loading tests of reinforced concrete column specimens. The specimens have circular, square, and wall-type cross sections, with various arrangements of hoop reinforcement so as to cover practical bridge column sections designed in Japan. The final obtained model is represented as follows [7]:

$$f'_{cc} = f'_{co} + 3.83 \rho_s f_{yh} \quad (5)$$

$$\epsilon_{cc} = 0.00218 + 0.0332 \rho_s f_{yh} / f'_{co} \quad (6)$$

Besides, another model is proposed by Sakai [8] which has also been demonstrated by Oreta and Kawashima [7] in which parameters such as tie spacing and column dimension had been considered for indirect estimation of  $f'_{cc}$  and corresponding  $\epsilon_{cc}$  are of circular cross-sectional confined RC columns. The obtained equation is represented as follows:

$$f'_{cc} = f'_{co}(0.94 + 4.7C) \tag{7}$$

$$\epsilon_{cc} = 0.0024 + 0.064C \tag{8}$$

in which  $C = K_s [\rho_s f_{yh} / (2f'_{co})]$  and  $K_s = [1 - s / (d \tan 30)] \geq 0$ .

In addition to several other studies conducted in the same field, different guidelines and specifications recommend different methods and models for computing the appropriate amount of transverse reinforcement for confinement purposes, e.g., ACI 318-08 [9], AASHTO [10], Caltrans [11] and Priestley [12], and each one considers different common and uncommon parameters which characterize the concrete, steel reinforcement, and cross-sectional details. However, the models and formulas suggested by guidelines and codes are often obtained based on analytical methods which are represented before, e.g., widespread used Mander’s model [2]. Besides, although numerous researchers have made great efforts to develop accurate stress–strain models of confined RC columns using analytical methods and assumptions, they often made their conclusions by optimizing the analytical results with their obtained experimental data through empirical methods. On the other hand, such models are designated through simplifying assumptions, and more importantly, they are initiated from few observations and controlling few models through simple statistical regression analyses to find the best model. This is due to the complex parameters affecting the system behavior which cannot thoroughly be considered for producing constitutive models of  $f'_{cc}$  and corresponding  $\epsilon_{cc}$  of confined RC columns. Thus, the complexity of analysis of stress–strain behavior of RC columns accounting for the influences of different parameters implies that there is still the necessity for obtaining more comprehensive models.

With regard to the progress of computer software and hardware systems, many computational techniques have been motivated from the analogy of biological activities and natural evolution, e.g., genetic algorithms, neural networks, support vector machines, and such other algorithms which are branches of data mining, artificial intelligence, and soft-computing techniques. These approaches have been utilized for solving real-world problems, because they can learn from data to find an optimal model relating the output variable to inputs. It means that they do not require prior assumptions about the system of the problem under study. In addition, these approaches can definitely be employed for generating models for engineering problems using experimental results. The outcomes of such

modeling techniques will be definitely a great combination of experimental studies and numerical methods. In addition, they have been successfully applied to many civil engineering prediction problems [7, 13–23].

Despite extending promotions in such artificial intelligence (AI) methods, there are few efforts in the literature for using them for prediction of confined compressive strength and strain of circular concrete columns. Recently, Oreta and Kawashima [7] have utilized artificial neural networks (ANNs), Tsai [13] and Tsai and Pan [14] demonstrated the capability of weighted genetic programming (WGP) and genetic programming polynomials (GPP) for modeling  $f'_{cc}$  and corresponding  $\epsilon_{cc}$  of circular cross-sectional confined RC columns. It is worth mentioning that the models obtained by those AI approaches have represented better performance in comparison with the traditional analytical formulas; however, the ANN model is a black-box model which has not been stated as an explicit formula to be used for hand calculation purposes. In addition, the GP-based models developed by Tsai [13] and Tsai and Pan [14] were based on different combinations of input variables and their structures are also complex.

This paper explores the capability of a variant of genetic programming (GP), namely, linear genetic programming (LGP), for indirect estimation of the confined compressive strength or peak stress,  $f'_{cc}$ , and the strain at peak-confined compressive stress,  $\epsilon_{cc}$ , of confined RC columns with circular cross sections using a database comprised of several costly test results conducted on different specimens previously published in the literature by Oreta and Kawashim [7]. With regard to the literature and the available database, two new models are obtained for predicting each of  $f'_{cc}$  (MPa) and corresponding  $\epsilon_{cc}$  (%) considering significant input variables including (1) compressive strength of unconfined concrete cylinder specimen,  $f'_c$  (MPa). It is worth mentioning that the unconfined concrete cylinder strength  $f'_c$  is selected as the input variable instead of the unconfined concrete strength of the column,  $f'_{co}$  (MPa), considering that the associated information with  $f'_c$  would be obtained easier as obtaining  $f'_{co}$  requires using the same equipment for testing confined column specimen with the same column geometry which is a high-capacity equipment and testing would be highly costly; (2) core diameter of circular column,  $d$  (mm); (3) column height,  $H$  (mm); (4) yield strength of lateral or transverse reinforcement,  $f_{yh}$  (MPa); (5) ratio of volume of lateral reinforcement to volume of confined concrete core,  $\rho_s$  (%); (6) spacing of lateral reinforcement or spiral pitch,  $s$  (mm); and (7) ratio of longitudinal steel to area of core of section,  $\rho_{cc}$  (%). Finally,  $f'_{cc}$  and  $\epsilon_{cc}$  are considered to be a function of the following parameters:

$$f'_{cc} = f(f'_c, f_{yh}, \rho_s, \rho_{cc}, d, s, H) \tag{9}$$

$$\epsilon_{cc} = f(f'_c, f_{yh}, \rho_s, \rho_{cc}, d, s, H). \tag{10}$$

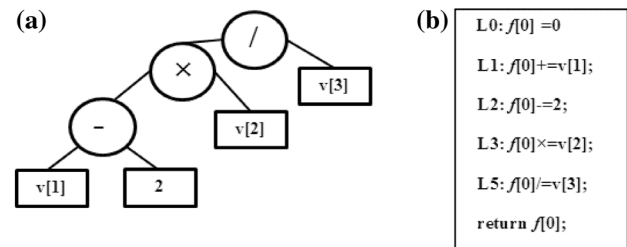
To show the capability of LGP method, out of data range, a new set of data which are not used in training of the model, considered as test data, used after training process. Consequently, to assess the accuracy of the obtained model validation, verification analyses are carried out and the results are compared with those provided by other researchers in the literature.

## 2 Linear genetic programming

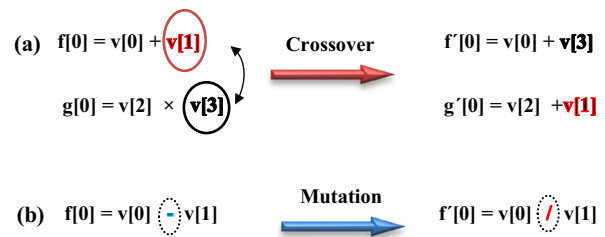
In 1958, Freidberg [24] attempted to solve simple problems by writing computer programs, i.e., the lingua franca for expressing the solutions to a problem. Thereafter, in 1985, Cramer [25] introduced the concept of tree-based genetic programming by applying an evolutionary algorithm (EA) to the programs that were represented as tree-shaped structures. In 1992, Koza [26] extended the work conducted by Cramer [25] and introduced genetic programming (GP) as a paradigm for searching the most fit computer programs that produce some preferred output for specific inputs.

Genetic programming is a subdivision of bio-inspired computational intelligence approaches which is basically created using the synthesis of Darwinian ideas of genetic inheritance to solve complicated problems. The essential goal in GP is to find predictive tree-shaped models relating the output to input variables, where experimental data samples are known based on the concept of survival of the fittest. Linear genetic programming (LGP) is an extended variant of GP with a linear structure similar to the DNA molecule in biological genomes and the name “linear” in LGP refers to the structure of the program representation. LGP evolves developing sequences of instructions from an imperative programming language (C or C++) rather than expressions of a functional programming language such as LISP [27]. Therefore, LGP systems can run faster than comparable tree-based GP interpreting systems and the programs are much simpler to read and to operate and these result in obtaining better models with less customization [28]. Figure 2 represents a comparison between three different structure of a program evolved by tree-based and linear GP.

In general, in GP and similarly in LGP, a population of programs is randomly initialized and individuals of the population are evaluated using a fitness function. Those members with higher fitness are given a higher chance to become a parent for the next generation, the offspring. Then, selected members are simply copied or transformed, by chance, into new members mainly via reproduction, mutation and recombination or crossover. Finally, comparing members of generation through a fitness function, the process repeat until the convergence conditions are satisfied and the optimal program is selected [15, 21, 29].



**Fig. 2** Comparison of a program structure evolved by tree-based (a) and linear-based (b) GP for the expression of  $y=f(v[i])=(v[1]-2) \times v[2]/v[3]$



**Fig. 3** Typical genetic operations in LGP: **a** crossover and **b** mutation

During the process of training, there may be some individuals or offspring which have an acceptable error while evaluated by the fitness function. In this regard, reproduction is performed by simply copying a selected member from the current generation to the next generation without modifications. Crossover occurs between two instruction blocks, whereas mutation occurs on a single instruction or chromosome. The crossover operation works by exchanging continuous sequences of instructions between parents. As it is seen in Fig. 3a, a segment of random position and arbitrary length is selected in each of the two parents [ $f(0)$  and  $g(0)$ ] and exchanged. If one of the two children would exceed the maximum length, crossover is aborted and restarted with exchanging equally sized segments. Two commonly used types of the standard LGP mutations are micro- and macro-mutation. The micro-mutation can change an operation type (e.g., change  $v[1] \times v[2]$  to  $v[1] - v[2]$ ), or can change the arguments of an operation (e.g., change  $v[1] + 1$  to  $v[1] + v[1]$ ). The macro-mutation can delete an operation (e.g., change  $v[1] + v[2]$  to  $v[2]$ ), or insert an operation (e.g., change  $v[1] + v[2]$  to  $v[1] + (v[1] \times v[1])$ ) [30, 31]. Figure 3 illustrates typical crossover and mutation operation in LGP.

The fitness function is a particular type of objective function that is used to assess how close an obtained solution is to achieving the set aims. In other words, how close the obtained results are to real experimental output data. In the present paper, the fitness function used to develop LGP

programs which may simply be computed through a difference or error function is represented as follows:

$$f = \sum_{i=1}^N (|E_i - P_i|) \quad (11)$$

where  $E_i$  is the output experimental or real value,  $P_i$  is the corresponding predicted value through the LGP solution case  $i$ , and  $N$  is the number of output data.

### 3 Experimental database

A reliable database containing 38 RC column load test results is acquired from Oreta and Kawashima [7] which is a data collection of previous studies conducted by Mander et al. [2], Sakai et al. [32], and Sakai [8]. It is true that more data may exist in the literature, but they may be not readily available or published or in some cases, and the published data are not complete considering the same variables. Thus, the same database is used for the development of the LGP models here which is efficient. The employed database is also brought here for more considerations (Table 1).

There are 15 column specimens from Mander et al. [2] and are 500 mm diameter and 1500 mm height. The core diameters of the columns are 438 mm. All columns consist of both lateral and longitudinal bars with varying sizes and spacing. Besides, seven columns are available from Sakai et al. [32] which are of 200 mm diameter and 600 mm height, and provided with both lateral ties and longitudinal bars 10 bars with 6.35 mm diameter. The N series are those provided with only one layer of lateral reinforcement, while the D series consists of two layers of lateral reinforcements. The 16 columns of Sakai [8] are of 300 mm diameter 280 mm core diameter and 900 mm in height. Sixteen longitudinal bars with diameter of 9.53 mm were provided. C1, C2, and C3 correspond to one, two, and three layers of lateral reinforcement, respectively. The descriptive statistics of the variables in the database is given in Table 2 for the aim of considering the statistical consideration of variables.

## 4 Development of LGP-based models

### 4.1 Data preprocessing

Data preprocessing such as data normalization is often employed in modeling process through machine learning techniques due to different scales of data, distributions of variables data sets and considering the limited range of data [33]. In this regard, a remarkable superiority of LGP and GP subdivisions over several computational and machine learning methods is the fact that input and output variables

are directly expressed in terms of computer programs produced by genetic programming consist of functions that are natural for the problem domain and suitable for employed data sets [15, 26]. Therefore, normalization can be eliminated. However, there is another problem in model generalization which is called overfitting. It is a case in which the error on the training set is driven to a very small value, but when new data are presented to the model, the error become large. A commonly used approach to avoid overfitting is to test the model on another group of data which are not used in the training process. To avoid overfitting, it is proposed that the available database should be classified into three sets: (1) training; (2) validation; and (3) test subsets [15, 33, 34]. The training set is utilized to fit the models, the validation set is used to estimate prediction error for model selection, and the test set which is a group of unseen data sets is used for the evaluation of the generalization error of the selected model. Regarding to several suggested values in the literature, the training, validation, and testing data are taken as 50–70, 15–25, and 15–25% of all data, respectively, for the development of models in artificial intelligence (AI) techniques [29, 33]. Herein, among 38 data, approximately 70% data sets are used for learning phase, i.e., training and validation, and about the 30% remaining data sets were used for the testing of the obtained models as the same sets of data were utilized by Oreta and Kawashima [7] for development of ANN-based models for prediction of peak-confined compressive strength and strain of circular RC columns. It is worth mentioning that if the fitness values on training validation and test data are close, it can be concluded that overfitting is avoided [15, 26, 27].

### 4.2 Initialization of LGP training parameters

Herein, the Discipulus software was used for the development of each LGP-based model. Similar to other evolutionary algorithms (EAs), GP and its variants such as LGP are of many degrees of freedom which should be taken into account to obtain an accurate model. Proper initialization of parameters highly affects finding efficient and accurate models. In this regard, quite a few runs may be required to be done and LGP parameters would be changed for each run which are basically obtained via trial and error. However, these parameters may be applied with the aid of previously suggested values [15, 34]. Herein, the model development process was controlled by considering evolutionary parameters, as represented in Table 3.

A run in LGP, similar to GP, proceeds through randomly generating initial populations of individuals or programs. Technically, the upper bound for the initial program length is a user-defined parameter and is illustrated in Table 3 as the maximum program size. The lower bound might be equivalent to the minimum length of a program, e.g., one instruction such

**Table 1** Employed database for development of LGP models [7]

Ref.	Specimen	$f'_c$ (MPa)	$f'_{co}$ (MPa)	$d$ (mm)	$H$ (mm)	$f_{yh}$ (MPa)	$\rho_s$ (%)	$s$ (mm)	$\rho_{cc}$ (%)	$f'_{cc}$ (MPa)	$\epsilon_{cc}$ (%)
Mander et al. [2]	M-a	28	24	438	1500	310	2	52	1.6	38	0.8
	M-b	31	30	438	1500	340	2	52	1.6	48	0.42
	M-c	33	32	438	1500	340	2	52	1.6	47	0.58
	M-1 <sup>a</sup>	28	29	438	1500	340	2.5	41	1.6	51	0.73
	M-2	28	29	438	1500	340	1.5	69	1.6	46	0.5
	M-3 <sup>a</sup>	28	29	438	1500	340	1	103	1.6	40	0.4
	M-4	28	29	438	1500	320	0.6	119	1.59	36	0.33
	M-5	28	29	438	1500	320	2	36	1.59	47	0.65
	M-6	28	29	438	1500	307	2	93	1.63	46	0.58
	M-7	31	32	438	1500	340	2	52	3.27	52	0.57
	M-8	27	30	438	1500	340	2	52	3.3	49	0.58
	M-9 <sup>a</sup>	31	32	438	1500	340	2	52	2.34	52	0.54
M-10	27	30	438	1500	340	2	52	3.2	50	0.64	
M-11	27	30	438	1500	340	2	52	4.8	54	0.45	
M-12	31	32	438	1500	340	2	52	3.2	52	0.56	
Sakai et al. [32]	N-1 <sup>a</sup>	29.8	24.6	185	600	376	0.57	120	1.18	29.6	0.37
	N-2 <sup>a</sup>	29.8	24.6	185	600	376	1.14	60	1.18	29.7	0.58
	N-2b	29.8	24.6	185	600	376	1.14	60	1.18	34.4	0.49
	N-3	29.8	24.6	185	600	376	1.71	40	1.18	35.9	0.59
	D-1 <sup>a</sup>	29.8	24.6	185	600	376	0.57	240	1.18	31.1	0.32
	D-2	29.8	24.6	185	600	376	1.14	120	1.18	36	0.4
D-3	29.8	24.6	185	600	376	1.71	80	1.18	36.1	0.57	
Sakai [8]	C1-20	19.45	21	280	900	363	2.26	20	1.85	35.4	1.3
	C1-30	19.45	21	280	900	363	1.51	30	1.85	29.7	0.8
	C1-40	19.45	21	280	900	363	1.31	40	1.85	27	0.62
	C1-60 <sup>a</sup>	19.45	21	280	900	363	0.75	60	1.85	24	0.47
	C1-80	19.45	21	280	900	363	0.57	80	1.85	22.8	0.42
	C1-120	19.45	21	280	900	363	0.38	120	1.85	19.8	0.31
	C1-160	19.45	21	280	900	363	0.28	160	1.85	19.3	0.24
	C2-40	19.45	21	280	900	363	2.26	40	1.85	33.8	1.38
	C2-60	19.45	21	280	900	363	1.51	60	1.85	27.8	0.71
	C2-80 <sup>a</sup>	19.45	21	280	900	363	1.31	80	1.85	25.4	0.58
	C2-120	19.45	21	280	900	363	0.75	120	1.85	22.3	0.37
	C2-160	19.45	21	280	900	363	0.57	160	1.85	20.1	0.34
	C3-60	19.45	21	280	900	363	2.26	60	1.85	34.1	1.06
	C3-80 <sup>a</sup>	19.45	21	280	900	363	1.7	80	1.85	26.7	0.79
C3-120	19.45	21	280	900	363	1.13	120	1.85	22.4	0.48	
C3-160 <sup>a</sup>	19.45	21	280	900	363	0.85	160	1.85	20.3	0.33	

<sup>a</sup>Data sample used as test data in the modeling process

as one variable, or can be an expression, i.e., more than one instruction. Although the success of finding a high accuracy model relies on the initial and maximum program size parameter, the optimal model is not necessarily the largest. In this regard, a major concern in development of models via GP is the way in which computer programs or individuals grow in size as the search progresses, what is known as bloat. Bloat is excessive code growth in size of the individuals of the evolving population without a notable improvement in fitness. Many

methods have been proposed in the literature to control the bloat in modeling process [35]. To cope with this issue, LGP uses limited variable-size chromosomes to a maximum number of instructions, the upper bound, which should be defined in advance by the user [36]. Besides, LGP take advantage of using demes that are semi-isolated subpopulations due to the fact that the evolution process develops faster in them in comparison with a single population of equal size [18, 27]. These remarkable features of LGP in comparison with GP and

**Table 2** Descriptive statistics of variables in the database

Statistical Indexes	$f'_c$ (MPa)	$f'_{co}$ (MPa)	$d$ (mm)	$H$ (mm)	$f_{yh}$ (MPa)	$\rho_s$ (%)	$s$ (mm)	$\rho_{cc}$ (%)	$f'_{cc}$ (MPa)	$\epsilon_{cc}$ (%)
Mean	25.10	25.11	324.87	1081.58	353.61	1.45	79.97	1.90	35.57	0.58
Median	27.50	24.60	280.00	900.00	363.00	1.51	60.00	1.85	34.90	0.57
Mode	19.45	21.00	280.00	900.00	363.00	2.00	52.00	1.85	52.00	0.58
Standard deviation	5.03	4.19	98.79	359.32	18.95	0.63	46.08	0.74	11.05	0.25
Sample variance	25.32	17.56	9760.39	129110.95	359.22	0.40	2123.05	0.55	122.04	0.06
Range	13.55	11.00	253.00	900.00	69.00	2.22	220.00	3.62	34.70	1.14
Minimum	19.45	21.00	185.00	600.00	307.00	0.28	20.00	1.18	19.30	0.24
Maximum	33.00	32.00	438.00	1500.00	376.00	2.50	240.00	4.80	54.00	1.38

**Table 3** Different combinations of settings and training parameters for LGP model development

Training parameters	Values
Population size	500, 700, 1000
Number of generations	500
Arithmetic operators	+, −, ×, /
Mathematical functions	sqrt(√)
Random constants	− 10 to 10
Maximum program size	64 and 128 bytes
Number of demes	20
Probability of crossover	0.5, 0.95
Probability of mutation	0.95, 0.5

its other variants make LGP faster which results in obtaining higher accuracy models.

### 4.3 The proposed LGP-based models

The optimal LGP programs would be achieved after performing the training process or after the convergence criteria have been reached in C++ format. The obtained code can be linked to the optimizer or may be called from the optimization routines [37]. For further considerations, the derived LGP-based codes for both  $f'_{cc}$  and  $\epsilon_{cc}$  are represented in “Appendix” at the end of the paper. In the present paper, to facilitate the use of the obtained codes for hand calculation aims and design purposes, they are simplified and converted into a functional representation by successive replacements of variables. The best LGP-based formulas of  $f'_{cc}$  and corresponding  $\epsilon_{cc}$  of circular confined RC columns can be calculated using the following equations:

$$f'_{cc} = f'_c + \rho_s \sqrt{f'_c \rho_{cc} \rho_s \frac{d}{H} \sqrt{\rho_s \frac{d}{s} + \frac{\rho_s f_{yh} c}{H}}} \quad (12)$$

$$\epsilon_{cc} = \frac{d}{H} + \frac{(3.5 \rho_s f_{yh} \rho_{cc} - 5.25 s \rho_s \rho_{cc}^2 + \rho_s H + H) \rho_s - 3s + 2d}{f'_c H} \quad (13)$$

It is noteworthy that the statistical regression models, e.g., semi-empirical or empirical models, are mostly obtained after controlling only few equations established in advance. Thus, these models cannot efficiently take into account the interactions between output and independent input variables. In fact, the best model generated by the LGP technique is obtained after controlling numerous linear and non-linear preliminary models, e.g., the proposed design equation for  $f'_{cc}$  is selected among a total of 628, 531,896 programs evolved and assessed by the LGP method during several runs. Besides, the obtained models have not been established by chance which represent the capability of LGP in finding optimal solutions.

## 5 Analyses and results

### 5.1 Prediction performance analysis of the model

To select an appropriate model, supplementary studies must be done and the performance and accuracy of the model should be precisely evaluated. This may be achieved through statistical analyses. According to Smith [38], if a model gives a correlation coefficient ( $R$ ) > 0.8 ( $R^2 > 0.64$ ), there is a strong correlation between the predicted and observed values which means that the prediction capability of the model is acceptable. However, the negative point of considering  $R$  only as a model performance evaluation criterion is the fact that it is insensitive to additive and proportional differences between model predictions and observed values. Therefore, other criteria are required to be considered. The root mean square error (RMSE) and mean absolute error (MAE) may be used as other criteria. These indices are more sensitive to differences between predicted and observed values, particularly occasional large errors. RMSE serves to aggregate the magnitudes of the errors in predictions for various times into a single measure of predictive power and gives disproportionate weight to very large errors. Besides, MAE represents how big of an error can be expected from the forecast on average. The lower the RMSE and MAE values, the better

the model performance would be. These parameters are calculated using the following equations:

$$R = \frac{\sum_{i=1}^n (o_i - \bar{o}_i)(p_i - \bar{p}_i)}{\sqrt{\sum_{i=1}^n (o_i - \bar{o}_i)^2 \sum_{i=1}^n (p_i - \bar{p}_i)^2}} \tag{14}$$

$$RMSE = \left[ \frac{\sum_{i=1}^n (p_i - o_i)^2}{n} \right]^{1/2} \tag{15}$$

$$MAE = \frac{\sum_{i=1}^n |p_i - o_i|}{n} \tag{16}$$

where  $o_i$  and  $p_i$  are the actual and predicted output values for the  $i$ th output, respectively,  $\bar{o}_i$  and  $\bar{p}_i$  are the average of the actual and predicted outputs, and  $n$  is the number of samples.

To find out whether the obtained models overfit, it is suggested that the performance of the model must be checked on different divisions of data. In this regard, the aforementioned performance measure tools,  $R$ , RMSE, and MAE, may be used. Close  $R$ , MAE, and RMSE values between predicted and actual values on the training and testing divisions of data sets confirm that overfitting in model development is avoided [15]. Values of statistical criteria obtained by LGP-based models based on data classification are summarized in Table 4. According to the close calculated values for

different divisions of data in Table 4, it can be concluded that overfitting is avoided. Besides, the performance of the models on the training and testing data demonstrates that they have both good predictive abilities and generalization performance.

Furthermore, it is suggested that the arithmetic values of mean ( $\mu$ ) and standard deviation ( $\sigma$ ) of ratios predicted versus actual values ( $p/a$ ) can be regarded significant indicators of the accuracy and precision of the estimation models. Prediction via a more accurate model results in  $\mu(p/a)$  closer to 1 and  $\sigma(p/a)$  closer to 0, respectively, in comparison with other models [39].

A comparison between the existing models in the literature and obtained LGP models could indicate the level of accuracy or capability performance of the presented models.

There are several statistical indices proposed in the literature to evaluate the model performance. Herein, with regard to the aforementioned issues, the overall statistical performance results of models of  $f'_{cc}$  (MPa) and corresponding  $\epsilon_{cc}$  (%) are summarized in Tables 5 and 6, respectively.

It can be observed in Tables 5 and 6 that the values of  $R$ , the correlation coefficient index, for different models are close. Although the ANN model is ranked first in those tables, as was expressed, it is a black-box model which cannot be used as a formula to for hand calculation purposes. This demonstrates the superiority of LGP model over the ANN approach. In case the ANN model can be converted to a tractable equation, the structure of the model will be too complex, considering the number of biases and weights. In this regard, LGP can also generate higher precision models using more instructions in its programs length; albeit, the

**Table 4** Calculated values of criteria for proposed LGP models based on data division

Statistical index	$f'_{cc-LGP}$ (this work)			$\epsilon_{cc-LGP}$ (this work)		
	All data	Train data	Test data	All data	Train data	Test DATA
$R$	0.97	0.97	0.96	0.96	0.97	0.94
RMSE	2.87	2.74	3.19	0.066	0.069	0.057
MAE	2.16	2.02	2.55	0.044	0.044	0.045

**Table 5** Evaluation of performance of different  $f'_{cc}$  prediction models considered in this study

References	Analysis of correlation		Analysis of error			Arithmetic calculations of $Q_p/Q_a$		Rank
	$R$	$R^2$	RMSE	MAE	Max IREI	$\mu$	$\sigma$	
$f'_{cc-LGP}$ (this study)	0.97	0.933	2.870	2.160	8.386	1.005	0.079	2
Oreta and Kawashima [7] ANN's model	0.99	0.980	1.687	1.234	4.180	1.023	0.046	1
Hoshikuma et al. [6]	0.90	0.807	10.161	8.940	18.621	1.289	0.190	5
Mander et al. [2]	0.92	0.853	12.395	11.622	18.685	1.357	0.184	6
Saatcioglu and Razvi [4]	0.94	0.893	5.800	4.662	11.369	1.151	0.168	3
Sakai [8]	0.94	0.886	6.240	4.733	13.106	0.895	0.105	4



**Table 6** Evaluation of performance of different  $\epsilon_{cc}$  prediction models considered in this study

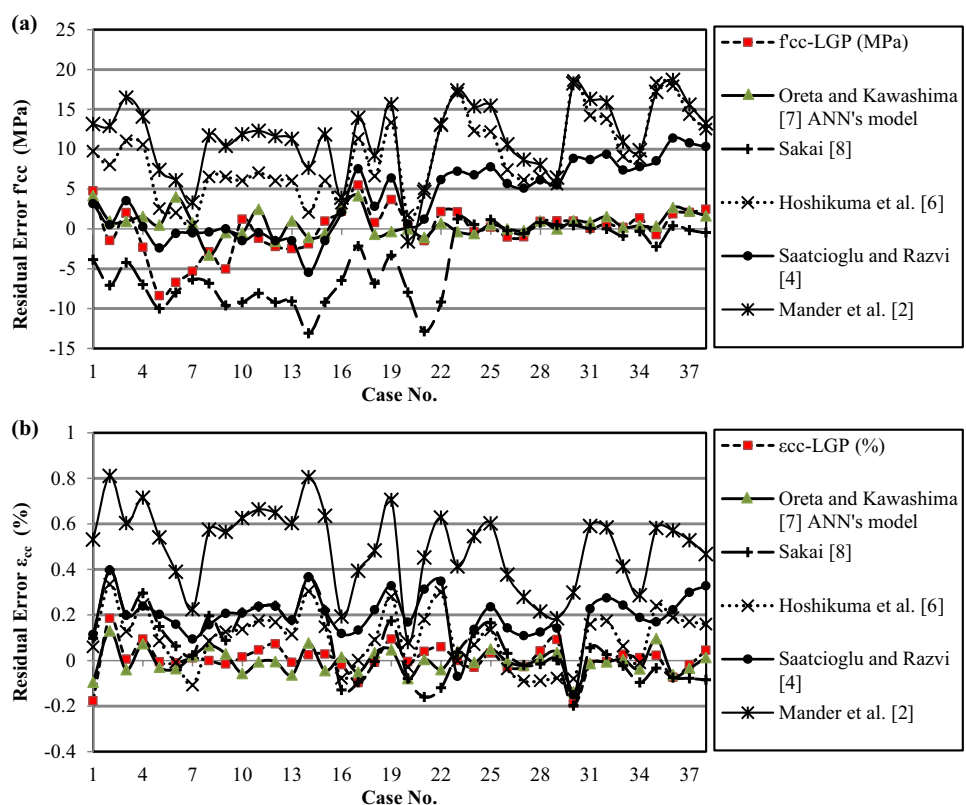
References	Analysis of correlation		Analysis of error			Arithmetic calculations of $Q_p/Q_a$		Rank
	$R$	$R^2$	RMSE	MAE	Max  RE	$\mu$	$\sigma$	
$\epsilon_{cc}$ -LGP (this study)	0.96	0.930	0.066	0.044	0.189	1.030	0.124	2
Oreta and Kawashima [7] ANN's model	0.98	0.952	0.054	0.043	0.138	0.997	0.100	1
Hoshikuma et al. [6]	0.91	0.821	0.155	0.130	0.335	1.163	0.261	3
Mander et al. [2]	0.86	0.738	0.525	0.495	0.810	1.923	0.351	6
Saatcioglu and Razvi [4]	0.91	0.821	0.224	0.208	0.397	1.411	0.234	5
Sakai [8]	0.85	0.730	0.159	0.126	0.396	1.103	0.296	4

objective of this study was to propose new simple design equations. It also should be considered that  $R$  gives a general viewpoint for the evaluation of the accuracy of the models and only this factor cannot be used to assess the degree of accuracy of models. The differences among models would be more as other indices are considered such as RMSE and MAE which represent an average error between predicted and experimental values. To explain this, another way of observing the systematic mismatches between the predictions and observations can be considered and that is the detection of the residual error (RE) between predicted and actual values for each datum. Figure 4 demonstrates a plot of the RE values obtained by proposed LGP models and those provided by traditional models. The maximum RE

values made by each model is also indicated in Tables 5 and 6. The indicated statistical analysis used for validation and verification of the LGP-based models can give the user an idea about the amount of reliability, accuracy or the factor of safety when using the model or selecting the best model among different models.

The RE values in Fig. 4 are calculated as predicted minus actual values. An important issue which may be realized from Fig. 4 is the fact that predicted  $f'_{cc}$  values made by Hoshikuma et al. [6], Mander et al. [1], and Mander et al. [2] models are more than actual  $f'_{cc}$  values which means those models overestimate the output, whereas that of Sakai [8] underestimate the values in many cases. Similarly, the results of RE plot represent that  $\epsilon_{cc}$  values are overestimated while

**Fig. 4** Comparison of RE values made by different models for all data sets **a**  $f'_{cc}$ , **b**  $\epsilon_{cc}$



using Mander's model. These consequences could not be realized through statistical indices which demonstrate the necessity of evaluating errors for models before using them.

Nevertheless, it can be also observed from Tables 5 and 6 that LGP-Based models are ranked after the ANN's model. In this regard, it is noteworthy that although prediction tools such as ANN are capable in predicting target values, they require the structure of the network to be identified in advance. This is usually done through a time-consuming trial and error procedure. Besides, these approaches are black-box tools. That is, such AI tools are not successful in generating practical prediction equations or if they are, the obtained model is highly complex and relies on several factors which often do not have the worth to be used. The main advantage of LGP over AI-based approaches is that it generates a transparent and structured representation of the system being studied via producing a practical model that can be easily utilized for hand calculating domains. Considering the values represented in Tables 5 and 6, and the results differences between predicted and actual values illustrated in Fig. 4, it can be concluded that LGP outperforms other models and methods in terms of accuracy and precision.

## 5.2 Parametric analysis

It is true that  $f'_{cc}$  must increase with the increase of  $f'_c$ ; in addition,  $\epsilon_{cc}$  must increase with increase of  $\rho_s$  as expressed by several experimental studies by various researchers. In this regard, the pertinent question is whether the models responses conform to the experimental study results.

In the present study, the results in the aforementioned studies on performance capability of the models represent that the LGP models are able to satisfy the required validation conditions as the models outperformed well-known models in the literature. However, the model response and behavior to different input variables must be examined to assess the obtained model from engineering viewpoints. Herein, to approve the results, they should be compared to the results that experimentally or theoretically provided in the literature. For this purpose, a comparative parametric analysis is performed in this study as recommended by several researchers, e.g., Sakai and Sheikh [5], Saatcioglu and Razavi [4], Mander et al. [1], Mander et al. [2], Hoshikuma et al. [6], and Oreta and Kawashima [7]. The parametric analysis represents the response of the LGP-based  $f'_{cc}$  and corresponding  $\epsilon_{cc}$  models to variations of input variables. The methodology is based on changing only one predictor variable at a time, while the other variables are kept constant at their specified mean values. For this purpose, a set of synthetic data is generated for each variable regarding to its range in the database. These values are presented to the prediction models and  $f'_{cc}$  and  $\epsilon_{cc}$  are calculated. This procedure

is repeated using another variable until the model response is obtained for all of the predictor variables [18, 19].

Figure 5 represents the responses of the  $f'_{cc}$  and corresponding  $\epsilon_{cc}$  prediction models to the variations of the affecting variables, i.e.,  $f'_c$  (MPa),  $f'_{co}$  (MPa),  $d$  (mm),  $H$  (mm),  $f_{yh}$  (MPa),  $\rho_s$  (%),  $s$  (mm), and  $\rho_{cc}$  (%). To confirm the accuracy of tendency of LGP models, the procedure of parametric study is comparatively conducted for different models in the literature. As can be seen,  $f'_{cc}$  values continuously increase with increasing  $f'_c$  which can be regarded as similar as  $f'_{co}$ ,  $d$ ,  $f_{yh}$ ,  $\rho_s$ , and  $\rho_{cc}$  and traditional models tend to respond similarly, whereas they decreases with increase of  $H$  and  $s$ . Besides,  $\epsilon_{cc}$  values grow with the increase of  $d$ ,  $f_{yh}$  and  $\rho_s$ , while they decline when  $f'_c$ ,  $f'_{co}$ ,  $\rho_{cc}$ ,  $H$ , and  $s$  increase.

These trends highlight that the proposed models are efficient in accurately predicting the  $f'_{cc}$  and corresponding  $\epsilon_{cc}$  of RC columns in terms of physical behavior. Particularly, comparing the results with other methods which are basically obtained via experimental studies further verifies the accurate LGP models behavior to each variable from engineering viewpoints.

## 5.3 Sensitivity analysis

Another important concern involving the analysis of models is to find out how a model output is affected by changes in input values. For this aim, a sensitivity analysis (SA) may be performed. The results of SA can represent how different values of predictors will affect a dependent variable. In other words, it may be realized that parameters with greater value of the sensitivity coefficient have stronger influence than other variables on variation of the model output.

For each input process parameter, the SA percentage is calculated using the following formula [17]. The percent of each obtained output difference for each variable is computed which is often referred to as the sensitivity index (SI). Herein, the SI (%) value can be calculated as follows:

$$N_i = f_{\max}(x_i) - f_{\min}(x_i) \quad (17)$$

$$SI_i(\%) = \frac{N_i}{\sum_{j=1}^n N_j} \times 100 \quad (18)$$

where  $f_{\max}(x_i)$  and  $f_{\min}(x_i)$  are maximum and minimum of the predicted output by LGP model for the  $i$ th input variable, where other input values remain at their mean values and  $n$  is the number of involved variables in the model.

The obtained SI value demonstrates the relative importance of each variable in model variation. This procedure is performed for all variables involved in LGP models and the results are represented in Figs. 6 and 7.

These figures indicate that the LGP model of  $f'_{cc}$  is more sensitive to  $f'_c$ ,  $\rho_s$ , and  $\rho_{cc}$  as compared to other variables,

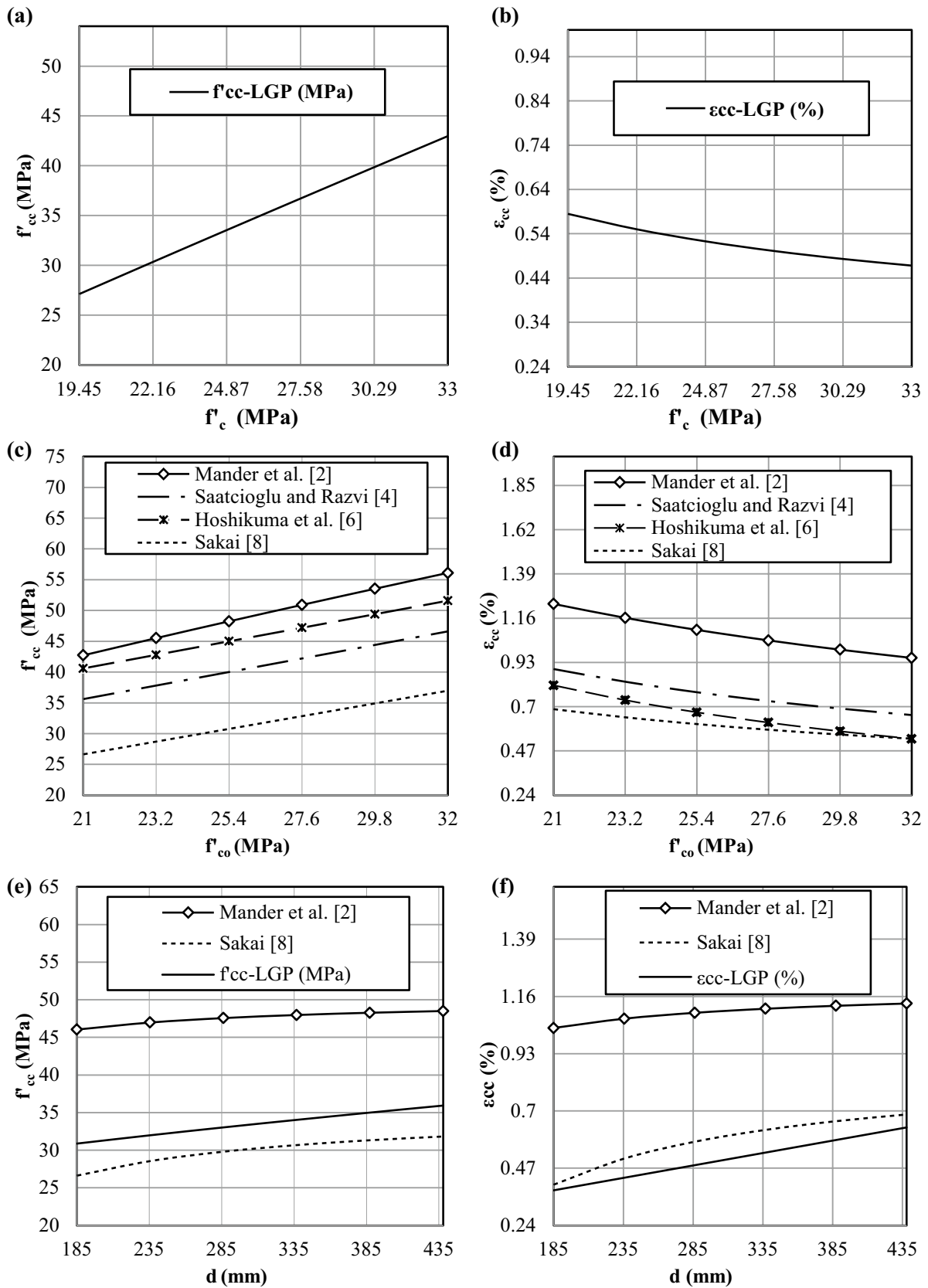


Fig. 5 Comparative parametric studies of various  $f'_{cc}$  and corresponding  $\epsilon_{cc}$  models for circular confined RC columns

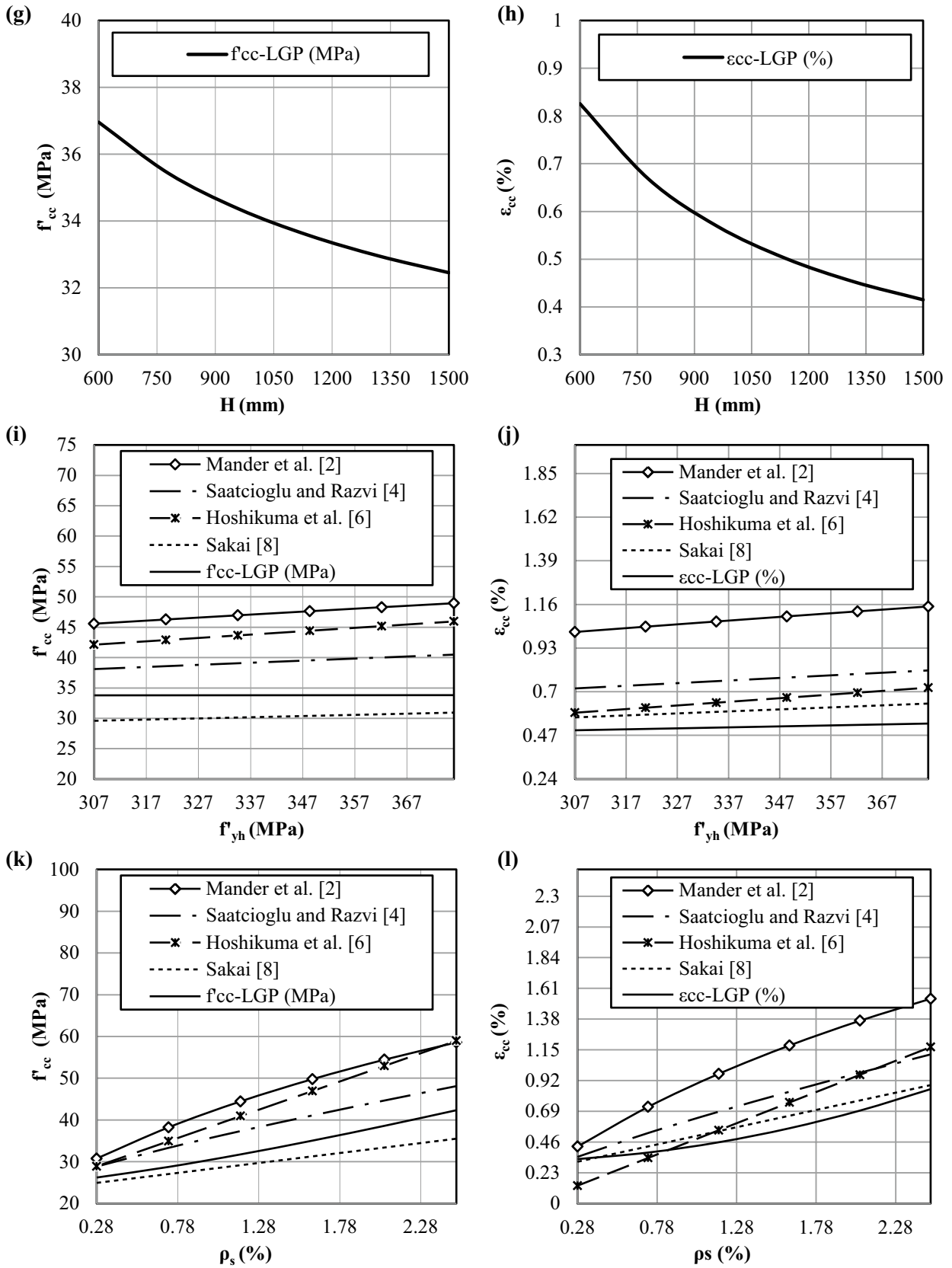


Fig. 5 (continued)

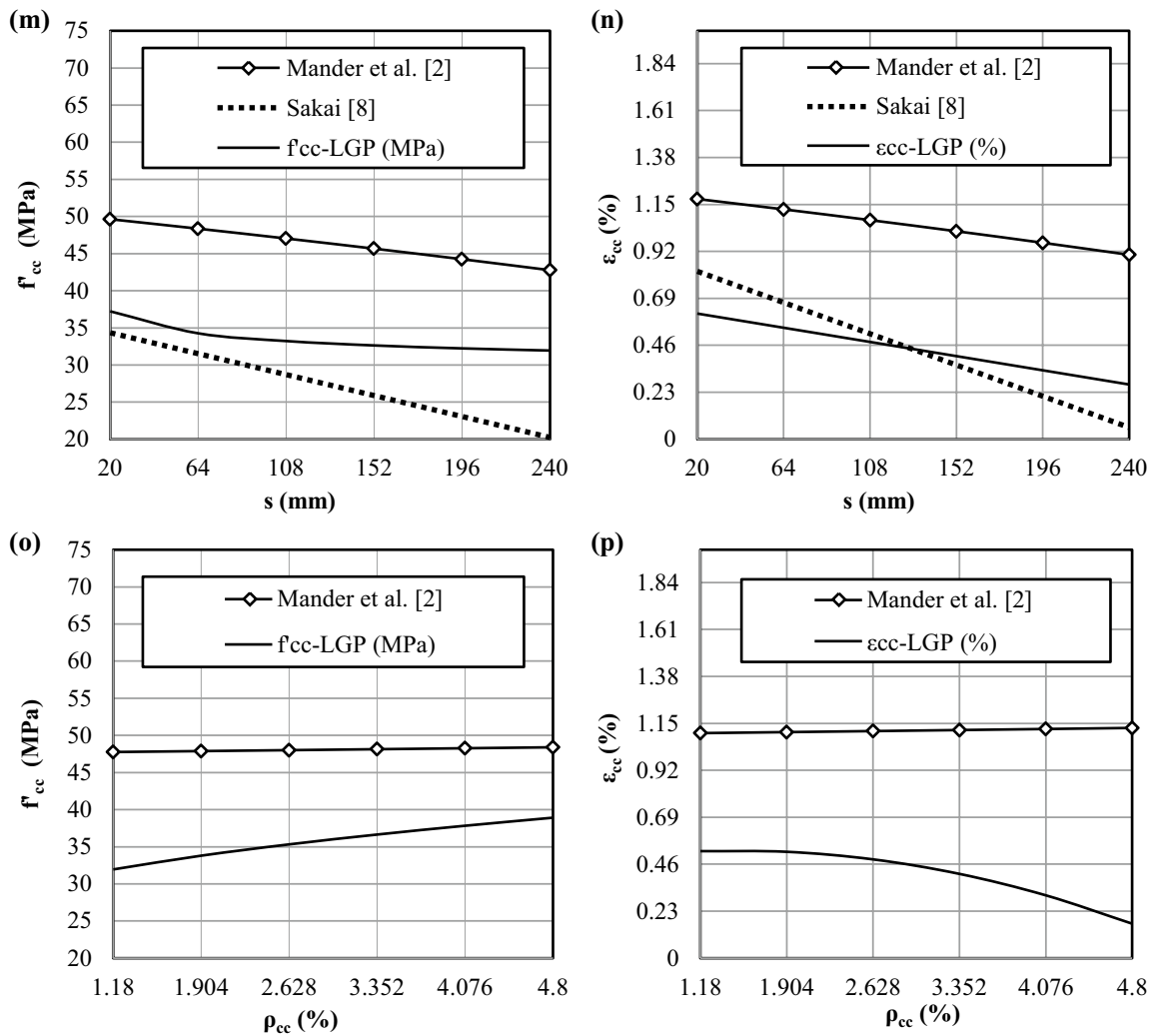


Fig. 5 (continued)

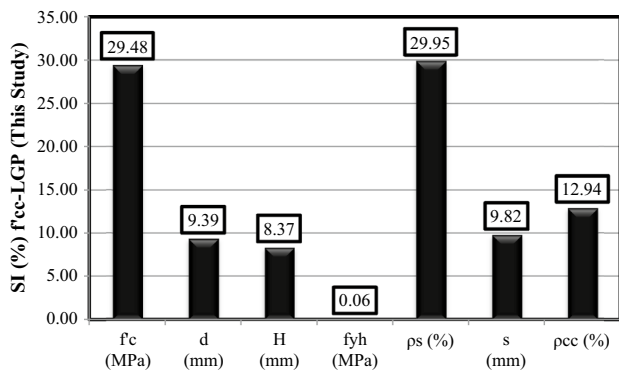


Fig. 6 Relative importance of the each variable in obtained LGP-based model for prediction of  $f'_{cc}$  circular confined RC columns

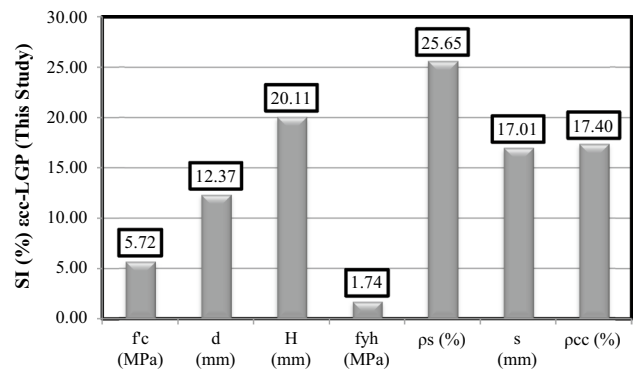


Fig. 7 Contributions of the each variable in obtained LGP-based models for prediction of  $\epsilon_{cc}$  circular confined RC columns

while that of  $\varepsilon_{cc}$  is basically relies on the variation of  $\rho_s$ ,  $H$ ,  $\rho_{cc}$  and  $s$ .

Moreover, according to Fig. 5 which represents the results of the comparative parametric study, the relative importance of each input variable can also be figured out for each model. In other words, the greater the variation of each model for each variable in its range, the more significant the involved variable affects the model. Besides, the significance of each variable obtained by each model can also be compared to those provided by other models. As a notable example, consider Fig. 5 (1), as  $\rho_s$  increases from 0.28 to 2.5, its range in the employed database,  $\varepsilon_{cc}$  value calculated by the LGP model varies from 0.33 to around 0.92 whereas that of Mander's model increases from almost 0.44 to around 1.52 which means that the variation of  $\varepsilon_{cc}$  is more in Mander's model in comparison with LGP model. This also means that the variation of  $\rho_s$  further affects the Mander's  $\varepsilon_{cc}$  model in comparison with other models. Similarly, the results can be extended for other variables and other models.

Moreover, it should be noted that results of a sensitivity analysis should be taken to account while using a formula. This is a serious issue which should be considered while varying a variable and calculating the amount of a variable via formulas for design purposes. In this regard, each model or formula gives distinct result; however, in the present study, all represented models give correct results and the main distinction is the amount of accuracy in their prediction performance. However, these results may be different for other proposed models. Finally, it is noteworthy that considering the range of the data used for model calibration and also the formulation in addition to the results of parametric study and SA definitely give the user a great comprehension about the design and final conclusion.

Finally, comparing the obtained results of comparative parametric and sensitivity analyses approve that the LGP model can accurately and correctly predict  $f'_{cc}$  and  $\varepsilon_{cc}$  in terms of physical behavior and engineering viewpoints as the obtained results by LGP models conform to those provided by well-known experimentally obtained models in the literature.

## 6 Conclusion

The objective of the present paper was to examine potential of LGP as an alternative approach for modeling a considerable structural and civil engineering problem which is estimation of  $f'_{cc}$  and  $\varepsilon_{cc}$  of RC columns. Reviewing the literature, many mathematical models have been proposed considering different common and uncommon parameters as independent input variables. In this regard, this study essentially resulted in deriving new robust models for prediction of those factors incorporating  $f'_c$  (MPa),  $f'_{co}$  (MPa),

$d$  (mm),  $H$  (mm),  $f_{yh}$  (MPa),  $\rho_s$  (%),  $s$  (mm), and  $\rho_{cc}$  (%) as affecting input variables for prediction of  $f'_{cc}$  and  $\varepsilon_{cc}$  of RC columns with circular cross section. Various analyses were conducted for evaluating the validation and generalization of the proposed models. The results mainly demonstrate that the proposed LGP models notably outperform the well-known models in the literature regarding the prediction capability. To assess the accuracy of the LGP models in terms of physical behavior and from engineering viewpoints, a comparative parametric analysis was carried out and the responses of obtained LGP models to input variables were compared to those obtained by other models. Besides, to find out how the proposed models outputs are affected by varying the inputs, a sensitivity analysis procedure was also performed. The results of comparative parametric and sensitivity analyses approved that LGP models can accurately predict  $f'_{cc}$  and  $\varepsilon_{cc}$  of circular RC columns and have appropriately captured the input/output non-linear relationships from engineering viewpoints as were conformed to those experimentally or theoretically provided in the literature. It is noteworthy that the main superiority of LGP over other computational and artificial intelligence modeling approaches is the fact that it is able to generate practical models which can be used via hand calculations during design procedures. Besides, the results of this study also represented that LGP can be utilized as a good alternative approach for indirect estimation of  $f'_{cc}$  and  $\varepsilon_{cc}$  of RC columns with circular cross section as the proposed models are able to predict those parameters with a high degree of accuracy. Furthermore, it should be considered that numerical models highly depend on the data used in their process of model development. The capability of these models is typically limited to the range of the data used for their calibration and generalization, and also the formulation depends on the available variables, in the database. To tackle this, the model can be retrained and improved to make more precise predictive models for a wider range of data and more variables. However, the authors believe that the proposed LGP models can be used for pre-design purposes.

## Appendix

### The optimum LGP program for the prediction of $f'_{cc}$

The following LGP program can be run in the Discipulus interactive evaluator mode or can be compiled in C++ environment. Note:  $v[0]$ ,  $v[1]$ , ...,  $v[6]$ , respectively, are  $f'_c$  (MPa),  $d$  (mm),  $H$  (mm),  $f_{yh}$  (MPa),  $\rho_s$  (%),  $s$  (mm), and  $\rho_{cc}$  (%) and  $f[0]$  holds the output which is  $f'_{cc}$  (MPa).

```

float DiscipulusCFunction(float v[])
{
  long double ff[8];
  long double tmp = 0;
  int cflag = 0;
  ff[0]=ff[1]=ff[2]=ff[3]=ff[4]=ff[5]=ff[6]=ff[7]=0;

  L0:  ff[0]+=v[3];
  L1:  ff[1]+=v[1];
  L2:  ff[0]/=v[2];
  L3:  ff[1]/=v[5];
  L4:  ff[0]+=ff[1];
  L5:  ff[0]*=v[4];
  L6:  ff[0]=sqrt(ff[0]);
  L7:  ff[0]*=v[1];
  L8:  ff[0]*=v[0];
  L9:  ff[0]*=v[4];
  L10: ff[0]*=v[6];
  L11: ff[0]=sqrt(ff[0]);
  L12: ff[0]*=v[4];
  L13: ff[0]+=v[0];
  L14:
  if (!_finite(ff[0])) ff[0]=0;
  return ff[0];
}

```

### The optimum LGP program for the prediction of $\epsilon_{cc}$

The following LGP program can be run in the Discipulus interactive evaluator mode or can be compiled in C++ environment. Note:  $v[0]$ ,  $v[1]$ , ...,  $v[6]$ , respectively, are  $f'_c$  (MPa),  $d$  (mm),  $H$  (mm),  $f_{yh}$  (MPa),  $\rho_s$  (%),  $s$  (mm), and  $\rho_{cc}$  (%) and  $ff[0]$  holds the output which is  $\epsilon_{cc}$  (%).

```

float DiscipulusCFunction(float v[])
{
  long double f[8];
  long double tmp = 0;
  int cflag = 0;
  ff[0]=ff[1]=ff[2]=ff[3]=ff[4]=ff[5]=ff[6]=ff[7]=0;

  L0:  ff[0]-=v[4];
  L1:  ff[0]*=1.5f;
  L2:  ff[0]*=v[5];
  L3:  ff[0]+=v[2];
  L4:  ff[0]*=v[5];
  L5:  ff[0]*=3.5f;
  L6:  ff[0]+=v[6];
  L7:  ff[0]*=v[3];
  L8:  ff[0]+=v[6];
  L9:  ff[0]*=v[3];
  L10: ff[0]-=v[4];
  L11: ff[0]-=v[4];
  L12: ff[0]-=v[4];
  L13: ff[0]+=v[1];
  L14: ff[0]+=v[1];
  L15: ff[0]/=v[0];
  L16: ff[0]+=v[1];
  L17: ff[0]/=v[6];
  L18:
  if (!_finite(ff[0])) ff[0]=0;
  return ff[0];
}

```

### References

1. Mander J, Priestley M, Park R (1988) Observed stress-strain behavior of confined concrete. J Struct Eng 114(8):1827–1849

2. Mander JB, Priestley MJ, Park R (1988) Theoretical stress-strain model for confined concrete. *J Struct Eng* 114(8):1804–1826
3. Rocca S, Galati N, Nanni A (2008) Review of design guidelines for FRP confinement of reinforced concrete columns of noncircular cross sections. *J Compos Constr* 12(1):80–92
4. Saatcioglu M, Razvi SR (1992) Strength and ductility of confined concrete. *J Struct Eng* 118(6):1590–1607
5. Sakai K, Sheikh SA (1989) What do we know about confinement in reinforced concrete columns? *ACI Struct J* 86(2):192–207
6. Hoshikuma J, Kawashima K, Nagaya K, Taylor A (1997) Stress-strain model for confined reinforced concrete in bridge piers. *J Struct Eng* 123(5):624–633
7. Oreta AW, Kawashima K (2003) Neural network modeling of confined compressive strength and strain of circular concrete columns. *J Struct Eng* 129(4):554–561
8. Sakai J (2001) Effect of lateral confinement of concrete and varying axial load on seismic response of bridges. Doctor of Engineering Dissertation, Dept of Civil Engineering, Tokyo Institute of Technology, Tokyo
9. Committee A, Institute AC (2008) Standardization IOf Building code requirements for structural concrete (ACI 318-08) and commentary. In: American Concrete Institute
10. Aashto L (2012) Bridge design specifications, 6th edn. American Association of State Highway and Transportation Officials, Washington, DC
11. Caltrans S (2010) Caltrans seismic design criteria version 1.6. California Department of Transportation, Sacramento
12. Priestley MN, Seible F, Calvi GM (1996) Seismic design and retrofit of bridges. Wiley
13. Tsai H-C (2013) Polynomial modeling of confined compressive strength and strain of circular concrete columns. *Comput Concrete* 11(6):603–620
14. Tsai H-C, Pan C-P (2013) Improving analytical models of circular concrete columns with genetic programming polynomials. *Genet Program Evolvable Mach* 14(2):221–243
15. Tajeri S, Sadrossadat E, Bazaz JB (2015) Indirect estimation of the ultimate bearing capacity of shallow foundations resting on rock masses. *Int J Rock Mech Min Sci* 80:107–117
16. Ziaee SA, Sadrossadat E, Alavi AH, Shadmehri DM (2015) Explicit formulation of bearing capacity of shallow foundations on rock masses using artificial neural networks: application and supplementary studies. *Environ Earth Sci* 73(7):3417–3431
17. Kiani B, Gandomi AH, Sajedi S, Liang RY (2016) New formulation of compressive strength of preformed-foam cellular concrete: an evolutionary approach. *J Mater Civ Eng* 28(10):04016092
18. Sadrossadat E, Heidaripناه A, Ghorbani B (2016) Towards application of linear genetic programming for indirect estimation of the resilient modulus of pavements subgrade soils. *Road Mater Pavement Des* 1–15
19. Sadrossadat E, Heidaripناه A, Osouli S (2016) Prediction of the resilient modulus of flexible pavement subgrade soils using adaptive neuro-fuzzy inference systems. *Constr Build Mater* 123:235–247
20. Baykasoğlu A, Güllü H, Çanakçı H, Özbakır L (2008) Prediction of compressive and tensile strength of limestone via genetic programming. *Expert Syst Appl* 35(1):111–123
21. Alavi AH, Gandomi AH, Sahab MG, Gandomi M (2010) Multi expression programming: a new approach to formulation of soil classification. *Eng Comput* 26(2):111–118
22. Mousavi SM, Alavi AH, Mollahasani A, Gandomi AH, Esmaeili MA (2013) Formulation of soil angle of shearing resistance using a hybrid GP and OLS method. *Eng Comput* 29(1):37–53
23. Armaghani DJ, Mohamad ET, Hajihassani M, Abad SANK., Marto A, Moghaddam M (2016) Evaluation and prediction of flyrock resulting from blasting operations using empirical and computational methods. *Eng Comput* 32(1):109–121
24. Friedberg RM (1958) A learning machine: Part I. *IBM J Res Dev* 2(1):2–13
25. Cramer NL (1985) A representation for the adaptive generation of simple sequential programs. In: Proceedings of the first international conference on genetic algorithms, pp 183–187
26. Koza JR (1992) Genetic programming: on the programming of computers by means of natural selection, vol 1. MIT press
27. Brameier M, Banzhaf W (2001) A comparison of linear genetic programming and neural networks in medical data mining. *IEEE Trans Evol Comput* 5(1):17–26
28. Francone FD, Deschaine LM (2004) Extending the boundaries of design optimization by integrating fast optimization techniques with machine-code-based, linear genetic programming. *Inf Sci* 161(3):99–120
29. Alavi AH, Sadrossadat E (2016) New design equations for estimation of ultimate bearing capacity of shallow foundations resting on rock masses. *Geosci Front* 7(1):91–99
30. Banzhaf W, Nordin P, Keller RE, Francone FD (1998) Genetic programming: an introduction, vol 1. Morgan Kaufmann San Francisco
31. Gandomi AH, Alavi AH, Sahab MG (2010) New formulation for compressive strength of CFRP confined concrete cylinders using linear genetic programming. *Mater Struct* 43(7):963–983
32. Sakai J, Kawashima K, Une H, Yoneda K (2000) Effect of tie spacing on stress-strain relation of confined concrete. *J Struct Eng* A 46:757–766
33. Shahin MA, Maier HR, Jaksa MB (2004) Data division for developing neural networks applied to geotechnical engineering. *J Comput Civ Eng* 18(2):105–114
34. Sadrossadat E, Soltani F, Mousavi SM, Marandi SM, Alavi AH (2013) A new design equation for prediction of ultimate bearing capacity of shallow foundation on granular soils. *J Civ Eng Manag* 19(sup1):S78–S90
35. Trujillo L, Naredo E, Martínez Y (2013) Preliminary study of bloat in genetic programming with behavior-based search. *EVOLVE-A bridge between probability, set oriented numerics, and evolutionary computation IV*. Springer, pp 293–305
36. Oltean M, Grosan C (2003) A comparison of several linear genetic programming techniques. *Complex Syst* 14(4):285–314
37. Deschaine LM, Patel JJ, Guthrie RD, Grimski JT, Ades M (2001) Using linear genetic programming to develop a C/C++ simulation model of a waste incinerator. *Advanced Technology Simulation Conference*, Seattle, pp 22–26
38. Smith GN (1986) Probability and statistics in civil engineering. Collins Professional and Technical Books 244
39. Abu-Farsakh MY, Titi HH (2004) Assessment of direct cone penetration test methods for predicting the ultimate capacity of friction driven piles. *J Geotech Geoenviron Eng* 130(9):935–944