



# Predictive modeling of the lateral drift capacity of circular reinforced concrete columns using an evolutionary algorithm

Mostafa Rezvani Sharif<sup>1</sup> · Seyed Mohammad Reza Sadri Tabaei Zavareh<sup>1</sup>

Received: 25 April 2019 / Accepted: 9 December 2019 / Published online: 21 December 2019  
© Springer-Verlag London Ltd., part of Springer Nature 2019

## Abstract

The drift capacity of reinforced concrete (RC) columns is a crucial factor in displacement and seismic based design procedure of RC structures, since they might be able to withstand the loads or dissipate the energy applied through deformation and ductility. Considering the high costs of testing methods for observing the drift capacity and ductility of RC structural members in addition to the impact of numerous parameters, numerical analyses and predictive modeling techniques have very much been appreciated by researchers and engineers in this field. This study is concerned with providing an alternative approach, termed as linear genetic programming (LGP), for predictive modeling of the lateral drift capacity ( $\Delta_{\max}$ ) of circular RC columns. A new model is developed by LGP incorporating various key variables existing in the experimental database employed and those well-known models presented by various researchers. The LGP model is examined from various perspectives. The comparison analysis of the results with those obtained by previously proposed models confirm the precision of the LGP model in estimation of the  $\Delta_{\max}$  factor. The results reveal the fact that the LGP model impressively outperforms the existing models in terms of predictability and performance and can be definitely used for further engineering purposes. These approve the applicability of LGP technique for numerical analysis and modeling of complicated engineering problems.

**Keywords** Reinforced concrete columns · Drift capacity · Cyclic lateral loading · Predictive modeling · Linear genetic programming

## List of symbols

$a/L_s$	Aspect ratio	$f_c$	Standard compressive strength of unconfined concrete samples
$A_c$	The area of RC column core within perimeter transverse reinforcement (center-to-center)	$f_{yl}$	The yield strength of longitudinal steel reinforcement,
$A_g$	Gross area of the column section	$f_{yt}$	The yield strength of lateral or transverse steel reinforcement
$A_{sl}$	The total area of the longitudinal reinforcement	$h_c$	The column concrete core depth
$A_{st}$	The total area of transverse reinforcement	$L$	Column length
$b$	Column section width	$L/D$	The column slenderness ratio
$b_c$	The column concrete core width	$P$	The axial load applied to the column
$d$	Effective depth	$s$	The spacing between transverse reinforcement or spiral pitch
$D_{eq}$ or $D$	The diameter of the equivalent circular column cross section	$V_{\max}$	The maximum shear strength
		$\alpha$	The axial load ratio calculated by $P/A_g f_c$
		$\beta$	The transverse reinforcement index equal to $\rho_l f_{yt} / f_c$
		$\gamma$	The longitudinal reinforcement index equal to $\rho_t f_{yt} / f_c$
		$\Delta_{\max}$	The maximum lateral displacement or the drift capacity of circular RC columns

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00366-019-00904-z>) contains supplementary material, which is available to authorized users.

✉ Mostafa Rezvani Sharif  
rezvanisharif@kntu.ac.ir

<sup>1</sup> Department of Civil Engineering, K.N. Toosi University of Technology, Tehran, Iran

$\rho_l$	The longitudinal steel ratio
$\rho_t$	The ratio of the volume of lateral reinforcement to volume of confined concrete core

## 1 Introduction

The key consideration in the design and performance assessment of civil engineering structures and infrastructures is the fact that they must work efficiently during their service life. This is achieved through ensuring about the structural safety, i.e., strength and stability, and their serviceability. Particularly in seismic hazard zones, the performance and design of buildings and their structural members should also be examined according to seismic provisions recommended by codes and guidelines [1, 2]. In this regard, the shear strength and ductility of structural members under cyclic lateral loads, such as seismic loads, are two important factors should be scrutinized. Although the strength of structural elements is of paramount significance, the ductility and the deformability, in both elastic and plastic states, are also important, since the member might be still able to withstand larger loads through its deformation. In fact, the ductility also causes to dissipate the energy applied to the structure and its corresponding elements during the occurrence of an earthquake.

Amongst structural members, columns are usually considered as the major structural member, generally, as axial load-bearing components. In reinforced concrete (RC) members, the slight tensile strength and the brittle behavior of the concrete is commonly enhanced through augmenting it with materials such as steel bars owing to their higher deformability and tensile strength. Besides to the existing longitudinal reinforcement in RC columns, experimentations demonstrated that transverse reinforcement through steel stirrups or spirals arranged in different layers as confinement along the column length also increase their shear strength and ductility characteristics, particularly as they are subjected to earthquake and seismic loadings [3–6]. Therefore, it is obvious that the ductility of RC columns is provided by the longitudinal or transverse reinforcement commonly including steel bars. In many cases, lack of existing appropriate ductility in columns would lead to their brittle failure and this might also result in structural collapse, severe damage and fatalities. To rise above this, the lateral drift capacity or the maximum lateral displacement of RC columns is typically examined by the researchers and practicing engineers and this is a crucial factor in displacement-based design and seismic evaluation of structures, since the structure is expected to withstand a specific drift without structural damage [7–9].

Considering the expensive process of direct testing approaches to acquire and observe the drift capacity of RC columns under cyclic lateral and seismic loads, indirect estimation and numerical modeling techniques have been very

much appreciated by engineers and researchers, mostly for pre-design purposes. The displacement-based design procedure for RC columns and pertinent approaches require estimation models of drift capacity [9]. In this regard, several models have been proposed by different researchers using analytical and empirical numerical analysis methods. Considering the plastic hinge concept, the lateral deformation and corresponding drift of RC columns can be represented as follows.

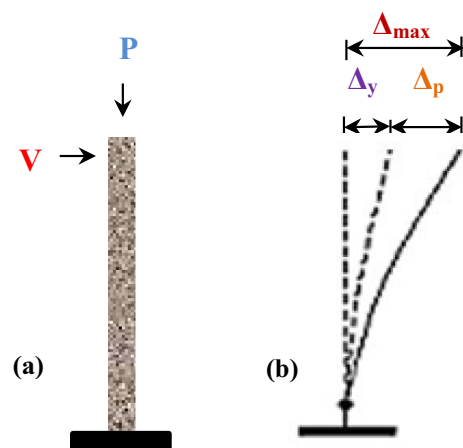
It can be realized from Fig. 1, the ultimate lateral drift capacity ( $\Delta_{\max}$ ) of the column can be considered as the summation of yield elastic displacement and plastic deformation,  $\Delta_y$  and  $\Delta_p$ , respectively, where  $L$  is the column length;  $P$  and  $V$  are axial and shear loads applied. The lateral drift capacity or ultimate deformation, i.e.,  $\Delta_{\max}$  is commonly defined as the lateral displacement of the column observed at 80% of peak shear strength ( $V_{\max}$ ) in pertinent load–displacement curves acquired by experimental studies and testing methods, according to [10].

Pujol and his colleagues [11] have explored the displacement of RC columns in their nonlinear range of response to reduce earthquake impacts via energy dissipation. In addition, it is observed that the column drift capacity is a function of displacement history and, as well as the shear strength, it decreases as the number and amplitude of cycles increases [12]. The maximum drift capacity model for RC columns proposed is as follows [10, 11]:

$$100 \frac{\Delta_{\max}}{L} = \frac{\rho_l f_{yt}}{v} \frac{a}{L_s} \leq \text{Min} \left( \frac{a}{L_s}, 4 \right) \quad (\text{psi units}), \quad (1)$$

where

$$v = \frac{V_{\max}}{bd}. \quad (2)$$



**Fig. 1** a Typical illustration of RC columns under shear ( $V$ ) and axial ( $P$ ) load in plastic hinge concept; b schematic deformation illustration of the RC column

Saatcioglu and Razavi [13] proposed a model based on column displacement analyses using different amounts of transverse reinforcement and levels of applying axial compressive loads, where the lateral displacement at 80% of lateral shear strength ( $V_{max}$ ) was considered as the  $\Delta_{max}$ . The model is given as follows:

$$\frac{\Delta_{max}}{L} = \frac{\rho_t f_{yt}}{14 f'_c} \frac{\sqrt{k_2}}{\left(\frac{A_g}{A_c} - 1\right)} \frac{P_O}{P} \quad (\text{MPa}). \quad (3)$$

In equation above

$$\rho_t = \frac{A_{st}}{bs}, \quad (4)$$

$$k_2 = 0.15 \sqrt{\frac{b_c b_c}{s s_t}} \leq 1.0, \quad (5)$$

$$P_O = 0.85 f'_c (A_g - A_{sl}), \quad (6)$$

and there exist some conditions as follows:

$$\frac{P}{P_O} \geq 0.2, \quad (7)$$

$$\frac{A_g}{A_c} - 1 \geq 0.3. \quad (8)$$

Furthermore, Elwood and Moehle [10] formulated the drift of columns at 80% in peak shear strength based on experimental data gained by tests carried out on RC columns specimens [10]. The empirical drift capacity formula has been acquired by regression analysis and is as follows:

$$\frac{\Delta_{max}}{L} = \frac{3}{100} + 4\rho'' - \frac{1}{40} \frac{v}{\sqrt{f'_c}} - \frac{1}{40} \frac{P}{A_g f'_c} \geq \frac{1}{100}, \quad (9)$$

where

$$v = \frac{V_{max}}{bd}. \quad (10)$$

From a different perspective, drift capacity models are also valuable to recognize the critical columns in seismic assessment and rehabilitation of structures. It is clear that the lateral drift capacity of RC members is a complicated problem, since there are numerous parameters and their changes, in turn, influencing the system performance under loadings. This can also be realized as the expression of the aforementioned empirical and analytical based models have incorporated distinct parameters, e.g.,  $P/A_g f'_c$ ,  $\rho_t$  and  $a/L_s$ , as independent input variables to estimate the  $\Delta_{max}$  of RC columns.

Nevertheless, there exist some defects associated with the models proposed. The most significant weakness of analytical methods is the fact that they cannot incorporate many parameters. This is due to that they are developed based on assumptions simplifying the problem. The empirical models based on regression analysis are usually calibrated and validated using few experimental data. In regression-based analysis few models are commonly checked to find the fit-test model. The drift capacity equations provided by Pujol et al. [11] and Elwood and Mohele [10] incorporate the shear strength ( $V_{max}$ ) of the RC column as an input variable. The  $V_{max}$  factor itself should be directly observed through experimental investigations and this is inefficient when indirect estimation is the purpose or it can be indirectly estimated through the numerical analysis which might produce negative effects on the accuracy of the drift capacity model in estimation. Furthermore, the design equation proposed by Saatcioglu and Razavi [13] is conditional, e.g., in Eq. 7, where  $P/P_O \geq 2$ , and covers a small range. This would lead to its deficiency in predictability and approximation capability.

When it comes to approximation and modeling of complicated real-world problems, data-mining methods and soft computing (SC) techniques, can be selected as useful alternative approaches. In general, SC techniques are naturally motivated algorithms which are developed based on the analogy of biological activities, the behavior of natural systems or biological systems of living creatures. SC methods are data-mining computational tools that attempt to simulate and find the structure of the models existing between the independent input and corresponding output variables. These data-driven and self-adaptive approaches such as artificial neural networks (ANNs) and support vector machines (SVM) and genetic programming (GP) and their subdivisions have been very much employed for predictive modeling of engineering problems [8, 14–18]. However, these approaches have been trivially employed for indirect estimation and simulating of the drift capacity of RC columns. Inel [8] applied ANNs to model the ultimate deformation and estimates of RC columns using experimental data in which flexural failure was the criterion. The model proposed incorporates the aspect ratio,  $a/H$ ,  $f'_c$  (MPa),  $f_{y1}$  (MPa),  $f_{yt}$  (MPa),  $s$  (mm),  $\rho_1$  (%),  $\rho_t$  (%), confinement effectiveness factor provided by [19] and the axial load ratio,  $P/f'_c A_g$  [8]. In addition, Gordon [20] also developed an ANN-based model for the prediction of the  $\Delta_{max}$  of RC columns using  $b$  or  $D$  for rectangular and circular RC column, respectively (m),  $d$  (m),  $P/\sqrt{f'_c} A_g$ ,  $\rho_1$  (%),  $\rho_t$  (%) and  $a/L$ . Although the results of the ANN models produced were expressed to be promising in estimation, ANNs have been criticized to be black-box tools and this is a substantial deficit. This means that the ANN-based developed models can only be computed through computer. In other words, the mathematical

structure of the obtained model cannot be presented as an explicit formula or an expression [6, 21, 22].

This paper is concerned with providing an alternative perspective on the numerical analysis and predictive modeling of the drift capacity of RC columns of circular cross section at their lateral strength with normal strength concrete. A robust SC technique, linear genetic programming (LGP) is employed to generate new predictive models. The superiority of LGP approach over other SC techniques is to create models that can be converted to explicit formulae and they can be employed for further design and calculation purposes. LGP has been successfully used to solve various engineering problems [16, 23–28]. A database consisting of 63 experimental results conducted on different types of circular RC columns is compiled for development of drift capacity models. Several runs have been done and many models have been developed using LGP approach for indirect estimation of the  $\Delta_{\max}$  of RC columns. The optimal model has been selected based on the assessment of the generated models from various accuracy perspectives.

## 2 LGP approach

Genetic programming (GP) is a symbolic optimization technique that generates computer programs or models based on the principle of Darwinian natural selection [23, 24, 29]. GP programs are programmed using a functional language such as LISP [24, 25]. LGP is an improved subdivision of GP having a linear program assembly comparable to the structure of DNA existing in natural genomes. This algorithm develops variable-length sequences of a programming language, C or C++ [16, 29, 30]. The functions commonly used in LGP include arithmetic operations, e.g., +, −, / and ×, unary functions such as Log, abs and Sqrt, conditionals such as if–then statements, and some of Trigonometric functions, Sin and Cosine. LGP algorithm can develop programs faster than other types of GP variants due to the simple formations of generated programs and the fact that these programs are not required to be interpreted [6, 22]. Figure 2 represents a linear shaped solutions generated by LGP.

In general, LGP starts with producing an initial number of individual programs or solutions. These are models, and therefore, their accuracy is assessed by means of a fitness

**Fig. 2** LGP program expressing  $y = f(v[i]) = \sqrt{3 - v[0]}$

```
L0 : ff[0]=0;
L1 : ff[0]=v[0];
L2 : ff[0] +=3;
L3 : ff[0]=Sqrt(ff[0]);
Return ff[0];
```

function. After running a tournament, typically four programs are randomly chosen from the initial individuals [23, 26]. Those members with better fitness acquire a greater opportunity to grow into a parent in the next production of programs. Subsequently, those fittest individuals are reproduced or transformed randomly to new programs by means of genetic operators. To tackle bloating problem, i.e., the undue development of the program size without a noteworthy fitness improvement, LGP restricts the size of programs which can be predefined prior to the model development [16, 22, 31]. After the evaluation of the programs regarding their fitness, the procedure repeats till the stop conditions are reached and the best program is obtained [27, 32, 33].

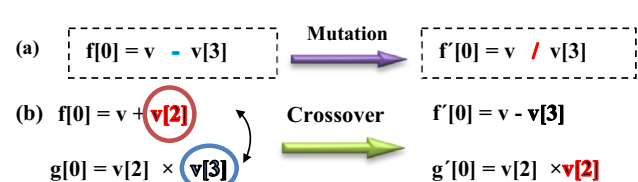
### 2.1 Genetic operators

The main operation in evolutionary algorithms (EAs) is performed by genetic operators mainly including reproduction, mutation, crossover or recombination [23, 28, 31, 33]. GP optimizes the population of computer programs according to a fitness function. Hence, the fitness function is the objective function that GP optimizes [26]. In the procedure of training, some parts of programs are modified by genetic operators to enhance the program accuracy. Reproduction is basically carried out by duplicating a fitting program within the existing population to the next generation without any changes applied. In addition, crossover and mutation change, the programs and structures of individuals are illustrated in Fig. 3.

### 2.2 Fitness assessment

The fitness function is used to assess how close the estimated and observed or experimental values are. Therefore, it can also be used as a selection means of producing individuals with higher accuracy in the training process [15, 27]. Here, the a linear sum of errors function is employed as the fitness function to develop LGP programs and this can be calculated using a linear mean of errors function which is given as follows:

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



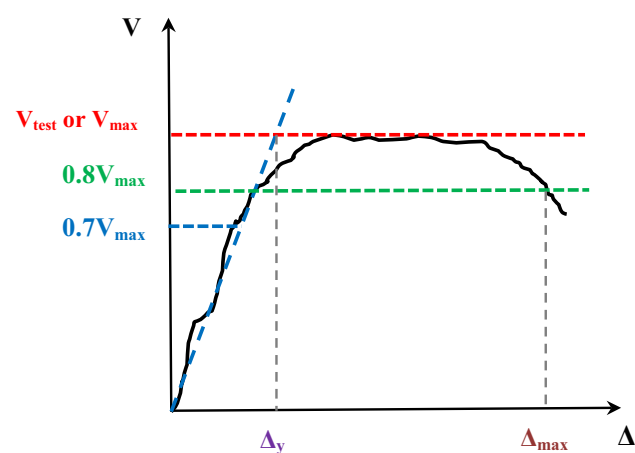
**Fig. 3** a Crossover and b mutation operators used in LGP

where  $E_i$  is the value of the outputs observed in experiments, and  $P_i$  is the equivalent predicted value using the LGP model for case  $i$  and  $N$  is the number of output data [16].

### 3 Experimental database and model variables

It is well-known that SC algorithms are data-driven. In present study, a valuable and reliable database already presented by the Pacific Earthquake Engineering Research Structural Performance Database (PEER-SPD) [34] is employed for the development of LGP models for indirect estimation of  $\Delta_{max}$  of RC columns. This comprehensive database is a collection of experimental studies previously conducted by several researchers in the literature to examine the precisions of seismic performance predictions for RC columns of different cross sections. The database contains the key RC columns material characteristics, mainly circular and rectangular cross section, geometry and reinforcement factors, the applied loads and observed damage, test results and so forth [20, 34]. The database used in this paper is presented in Appendix A.2. Test configurations considered are cantilever, double-curvature, double-ended, hammerhead, and cantilever-flexible-base testing methods [20, 34]. The samples have been tested under cyclic horizontal loading until failure. It is noteworthy that the shear failure was assumed and was defined by the maximum displacement,  $\Delta_{max}$ , where shear resistance fell below 80% of the maximum shear recorded [10, 20, 35].  $\Delta_y$  and  $\Delta_{max}$  acquired by load–displacement test results has been defined by Elwood and Mohele [10], as represented in Fig. 4.

$\Delta_s$  or  $\Delta_{max}$  is the displacement at 80% of  $V_{max}$ , and  $\Delta_y$  is the displacement at the point of intersection of a horizontal line at the peak shear, and a line formed by the origin and the point on the force–displacement envelope, where



**Fig. 4** Representation of columns drift capacity in shear load–displacement curves reported in the database

the shear is at 70% of its peak value [20]. Nevertheless, the purpose of this study is to find a predictive model for the  $\Delta_{max}$ . The experimental data in which the shear failure of samples under cyclic lateral loading occurred were selected in presents study for the development of models. Two modes of shear failure reported in the database, i.e., flexural-shear and shear failure modes, are typically considered as the main reason to narrow down the PEER-SPD. Therefore, the LGP model is merely useful for these two type of failure modes. The descriptive statistical details of the database is summarized and given in Table 1. For further considerations, the whole experimental data used here is attached to this article as a supplementary file for online publication.

Considering the models proposed by different researchers and the incorporated variables previously mentioned, it is possible to find the significant variables to estimate the  $\Delta_{max}$  factor. The LGP-based  $\Delta_{max}$  model in this study is also assumed to be a function of the key independent input variables. Considering the available information in the experimental database, some significant variables such as  $L/D$ , which represents the slenderness of the column, can also be used to develop LGP models. It is worth mentioning that analytical models are commonly developed based on assumptions for simplifying the problem. These models might not be capable of capturing the impact of several important variables such as  $L/D$ . It is also noteworthy that the LGP model is free from the impact of  $V_{max}$ , since new experiments are required to calculate this parameter and calculating this parameter from analysis may impact the accuracy of the model. This is the superiority of LGP over other approaches and also the fact that LGP models are based on experimental observations without any assumptions. The variables employed here are considered to be dimensionless and meaningful from engineering perspectives. The LGP  $\Delta_{max}$  (mm) function can be demonstrated as follows:

$$\Delta_{max} = f\left(\frac{L}{D}, \rho_t, \rho_l, \alpha, \beta, \gamma, \frac{s}{d}, \frac{a}{L}\right), \tag{12}$$

where  $L/D$  is the slenderness ratio which is defined as the column length to diameter or the circular column,  $\rho_t$  and  $\rho_l$  are the transverse and longitudinal reinforcement ratio (%),  $\alpha$  is assumed to be the axial load ratio,  $P/A_g f_c'$ , and  $\beta$  is considered to be the transverse reinforcement index equal to  $\rho_f y_l / f_c'$ , and  $\gamma$  is the longitudinal reinforcement index equal to  $\rho_f y_l / f_c'$ ,  $s/d$  is the transverse reinforcement spacing to depth ratio and  $a/L_s$  is the aspect ratio.

### 4 Data classification

To assess the potential of prediction models, i.e., to see whether they underfit or overfit and their generalization performance, many researchers recommended that the data sets

**Table 1** Statistical factors of input variables considered for the drift capacity model development

Parameter	$D_{eq}$ (mm)	$L$ (mm)	$f_c$ (MPa)	$\rho_1$ (%)	$f_{yt}$ (MPa)	$\rho_t$ (%)	$f_{yt}$ (MPa)	$s$ (mm)	$d$ (mm)	$P$ (KN)	$V_{max}$ (KN)	$a/L_s$	$\Delta_s$ or $\Delta_m$ (mm)
Mean	407.87	762.23	33.00	2.89	405.14	0.73	355.45	82.02	370.10	399.07	351.21	1.99	21.78
Standard deviation	207.79	497.22	6.52	1.08	57.21	0.70	79.45	51.12	204.32	640.38	385.75	0.58	16.38
Range	1547	3408	38.05	5	268	4	492	279.8	1545	4300	3048	2	83
Minimum	282	250	18.9	1	240	0	200	25	199	0	85	1	3
Maximum	1828.8	3658	56.95	5.21	508	4.18	691.5	304.8	1743.8	4300	3132.98	3.44	85.82
Sum	25,695.6	48,020	2079	182	25,524	46.1	22,393	5167	23,316	25,142	22,126.1	125	1372.4

be mainly classified into three groups, learning, validation and test data [6, 27, 36]. The learning group is applied to fit the models, i.e., LGP-generated programs, and the validation is utilized for the fitness assessment of learned programs. It is noteworthy that the learning and validation process is known as the training process. Another data set which is not employed in the training process and only to confirm the predictability of the obtained model and this set is so-called test data. Considering the recommended values by researchers, the training and testing data are usually taken 80% and 20% of all data, respectively, for model development through SC techniques [6, 21, 22, 33, 37]. Similarly, these values are considered in this study to explore the accuracy degree of the generated models. It is obvious that in case the performances of the model on different sets of data, e.g., learning, validation and test data, are good and close, overfitting and underfitting are avoided [6, 16].

### 5 LGP training parameters

To find the estimation models of  $\Delta_{max}$  for circular RC columns, Discipulus software provided by [38], was utilized in present study. In SC algorithms, There exist several parameters in the algorithms should be fixed or predefined to perform a run. Herein, the initial number of program population was considered to be first 500 and second 700. The number of generation of programs without improvement was fixed at 200 which mean that the generation of programs would be stopped when the fitness of 200 programs was not acceptable in that iteration. Various operators, +, -, ×, /, mathematical operator, √, and random constants between - 10 and 10 have been selected to develop programs. These sorts of operators are considered here to find a simple LGP model which can be used via manual calculations. A significant problem in GP is the fact that the size of produced programs gradually increase without any improvement in their accuracy after a while and this is known as bloat. To rise above this, LGP restricts the size of programs generated to a number of orders [31–33]. In this study, the maximum program size was fixed at 120 bytes. In addition, demes, i.e., semi-isolated subgroups, classify program populations of equal size and this leads to a better evolution of generation in each run in LGP [6, 21]. The number of demes was considered to be equal to 20. There exist some other parameters such as the probability of crossover and mutation which were equal to 0.5 and 0.95, respectively. These parameters should be appropriately initialized to produce an accurate model. To adjust these parameters' values, numerous runs might be required to be carried out through a process of trial and error. Albeit, those previously suggested values of the parameters by other researchers can also be used [6, 22, 31, 32].



## 8 Results and discussion

### 8.1 Evaluation of the performance of models

As was stated, during and after the process of modeling and performing training and validation procedures, several models are generated by LGP. With the intention to examine the fitness of models, some statistical indices,  $R$ , RMSE and MAE, are initially employed after the assessment of models used through a fitness function. The performance of the optimal model should be examined for different division of data, i.e., training data and test data. The last is known as the generalization performance or the predictability of the model [6]. Furthermore, PI can be used to evaluate the model incorporating the simultaneous impact of  $R$ , RMSE and/or MAE. These indices are calculated for different classes of data employed here In Table 2.

It is proposed that if PI value is close to zero (e.g.,  $PI \leq 0.2$ ), the model predicts the actual values well and is the recommended acceptance threshold [25]. According to the PI values in Table 2, the performance of the LGP model proposed here is acceptable. Furthermore, a traditional way to assess the accuracy and fitness of a model in comparison to those existing in the literature for estimation of the drift capacity of RC columns is to indicate the model output results versus the experimental values [6, 17, 40]. Figure 5 shows a scatter plot of the obtained values of the  $\Delta_{max}$  of cyclically loaded circular RC columns obtained by the LGP and other models proposed by well-known researchers versus those experimentally observed.

The calculated values of  $R$ , RMSE, MAE and PI for all, training and test data presented in Table 2 imply that the LGP model does not underfit which means that it is capable of estimating  $\Delta_{max}$  factor with an acceptable degree of precision. In addition, the closeness of  $R$ , RMSE and MAE values calculated for training and test data indicates this fact that overfitting is eliminated. It can be figured out from Fig. 6 that the model provided in this study by LGP method considerably outperforms those well-known models suggested by researchers for indirect measurement of  $\Delta_{max}$  of RC columns.

Although the models provided by Elwood and Mohele [10] and Pujol [11] can to estimate the  $\Delta_{max}$  factor with an

acceptable degree of accuracy, they present poor results for several cases. It can be seen in Fig. 5 that the PI calculated values are more than 0.2, while that of LGP is equal to 0.15 which demonstrates better performance of LGP model compared to these models. This is mainly due to the fact that these models incorporate few input variables for the estimation of  $\Delta_{max}$  factor, e.g., Pujol [11] merely considered  $L$ ,  $b$ ,  $d$ ,  $V_{max}$ ,  $\rho_t$  and  $f_{yt}$ . Therefore, in case an ignored parameter changes cause variation of the model output, e.g.,  $P$ ,  $\rho_l$  and  $D$  which have not been considered in Pujol's model, this impacts the model performance and cause poor estimations and inefficiency of that model for such data, similar to this study. In addition, the experimental data employed in other studies were mainly conducted on columns samples of rectangular cross section; albeit, the generated models are used for estimation of  $\Delta_{max}$  of RC columns of circular cross section. It can be also seen that the Elwood and Mohele's model almost overestimated the  $\Delta_{max}$  factor in several cases. It is worth noting that the empirical model proposed by Saatcioglu and Razvi [13] was not able to estimate all  $\Delta_{max}$  values considered in the database employed here. The main reason is the fact that the model is conditional and developed based on limitations which cause deficiency and poor results and inapplicability for several data. Thus, the aforementioned model could not be presented here.

### 8.2 Additional indices for the evaluation of the model predictability

Some additional factors have been recommended by researchers in the existing literature which can be used to evaluate the generalization performance of mathematical models on unseen data [6, 17, 21, 33]. In present study, these criteria are considered for the assessment of the predictability of the LGP-based  $\Delta_{max}$  model and the obtained results are summarized in Table 3. Numbers in this table indicate the fact that the proposed model fulfills all of the conditions required and this approves the robustness of LGP model in prediction for unseen or test data.

According to [25],  $R$  does not change significantly by shifting the output values of a model equally, and error functions (e.g., RMSE and MAE) only shows the error. Therefore, a better criteria can be used which can simultaneously incorporate the impact of  $R$ , RMSE and/or MAE such as performance index (PI) [25]. PI is also calculated to evaluate the predictability of the proposed LGP model and is given in Table 2.

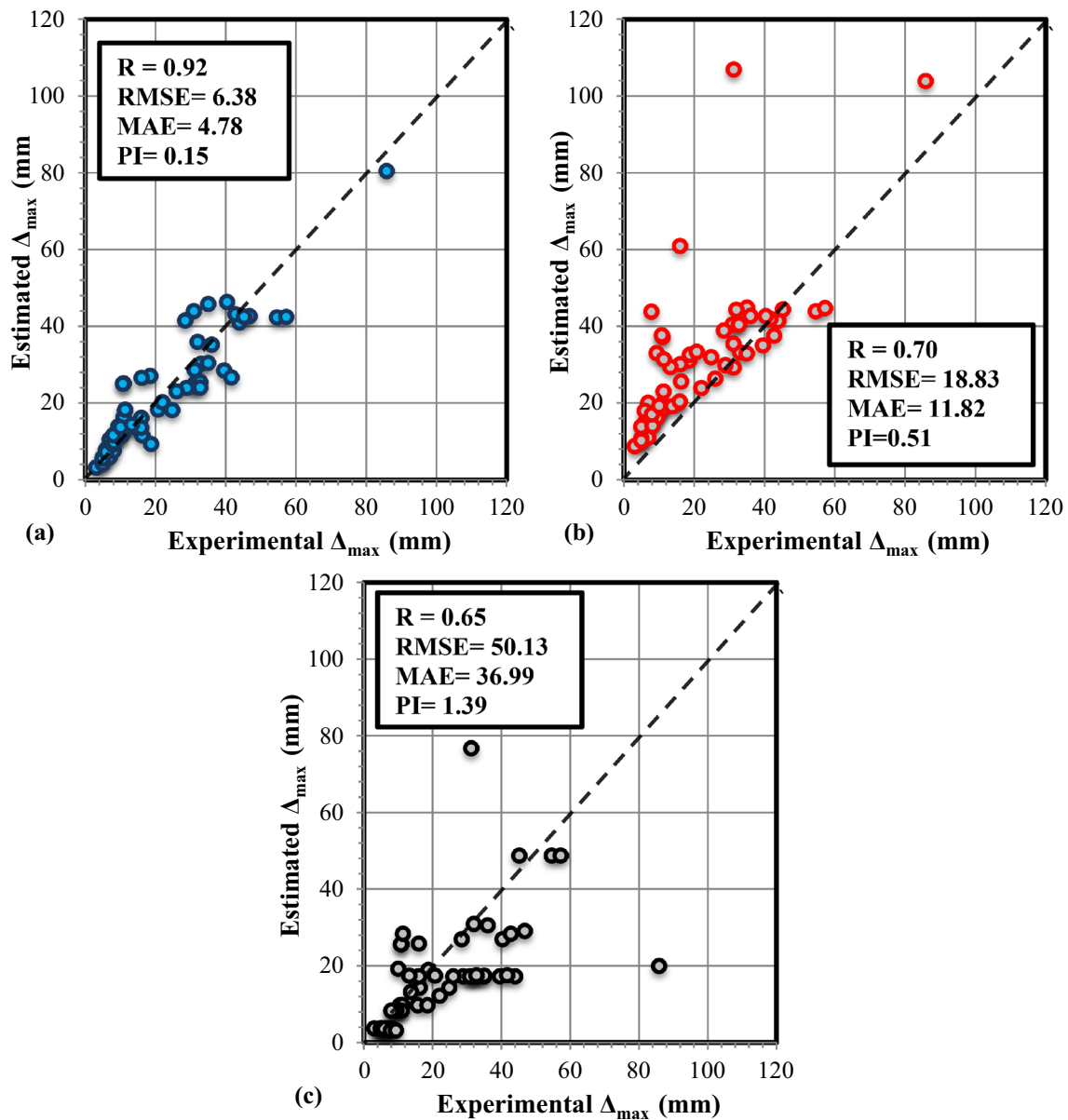
### 8.3 Parametric analysis of the LGP model

Considering the coefficients of different variables in  $\Delta_{max}$  models proposed by Elwood and Mohele [10], Pujol [11] and Saatcioglu and Razavi [13], the trend of  $\Delta_{max}$  can be

**Table 2** Performance of the LGP model on different divisions of data

Item	All data	Training data	Test data
$R$	0.92	0.922	0.916
RMSE	6.38	6.19	6.96
MAE	4.78	4.72	5.01
PI	0.15	0.15	0.17



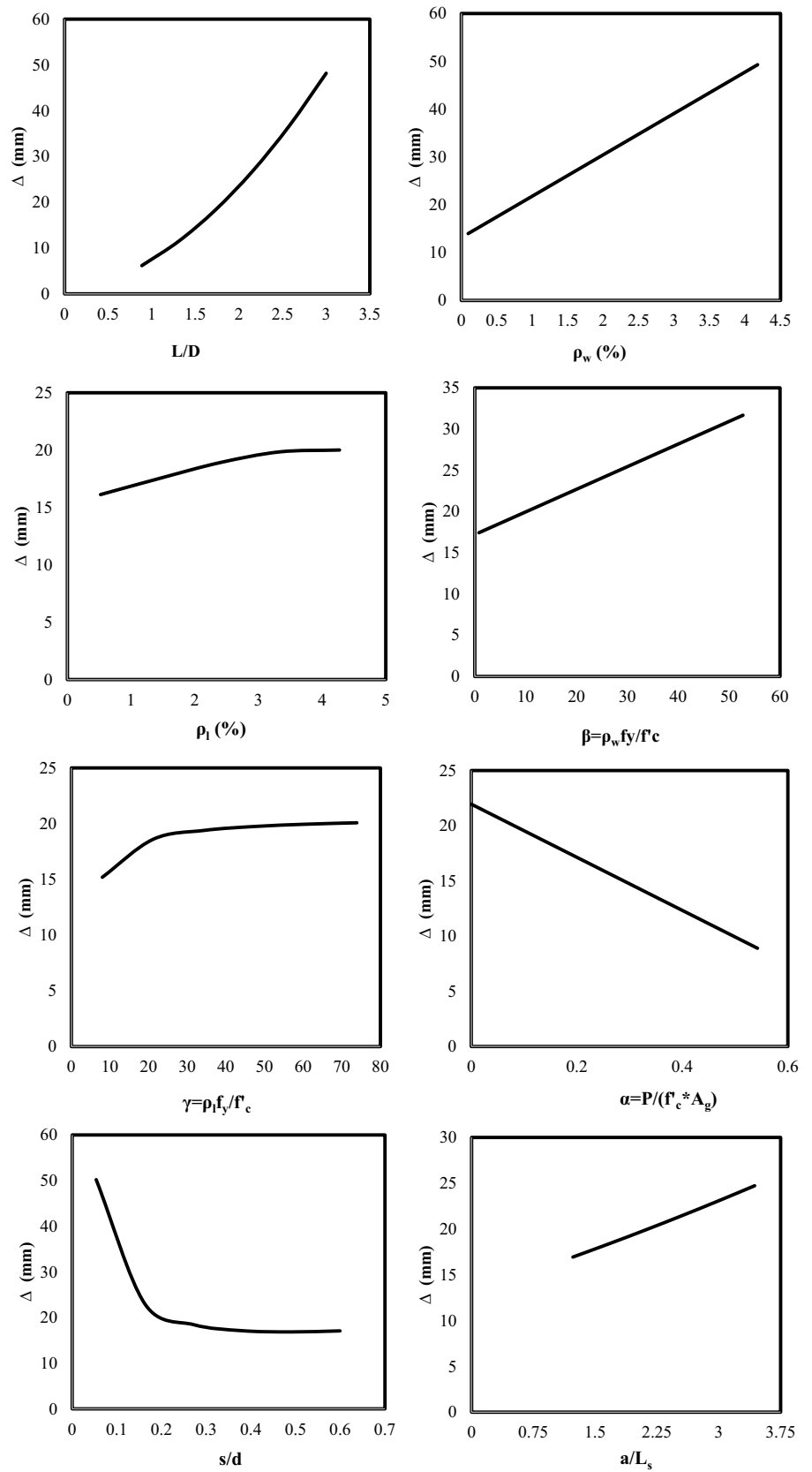


**Fig. 5** Estimated versus experimental values of  $\Delta_{max}$  calculated by **a** LGP model (this study); **b** Elwood and Moehle’s Model [10], and **c** model proposed by Pujol et al. [11]

realized when each variable varies. For instance, when  $P/A_g f'_c$  increases,  $\Delta_{max}$  decreases which demonstrates that it has a negative coefficient in the model proposed by Elwood and Moehle [10]. Accordingly, it can be expected that  $\Delta_{max}$  must increase with increasing  $\rho_r$ ,  $\rho_l$ , and  $a/L_s$ . The trend of the model to the changes of its independent variables must be proper and meaningful. Definitely, if the model behavior conforms to the outcomes in the literature, the model can be accepted from engineering viewpoints. In present study this behavior is explored for LGP models using a parametric analysis [6, 25, 33].

The methodology utilized here is dependent on merely varying one independent variable, while the other variables are considered to be equal to their mean values [6, 25, 33]. For this aim, a group of simulated data is generated for each input variable accounting for the variable’s range in the database. Then, the output value, which is the  $\Delta_{max}$  of circular RC columns here, is calculated by the predictive model using the virtual values of inputs. The aforementioned procedure is repeated similarly for another variable till the model response to all input variables has been obtained. Figure 6 shows the results of the parametric analysis carried out for the LGP model to assess its response to the changes

**Fig. 6** Parametric analysis of the proposed LGP-based model for indirect estimation of the  $\Delta_{max}$  of circular RC columns



**Table 3** Additional parameters for evaluation of the LGP model predictability proposed for prediction of  $\Delta_{max}$  of RC columns

Item	Formula	Condition	The LGP model
1	$R$	$0.8 < R$	0.916
2	$k = \frac{\sum_{i=1}^n (h_i \times t_i)}{\bar{h}_i^2}$	$0.85 < k < 1.15$	0.994
3	$k' = \frac{\sum_{i=1}^n (h_i \times t_i)}{t_i^2}$	$0.85 < k' < 1.15$	0.936
4	$R_m = R^2 \left( 1 - \sqrt{ R^2 - Ro^2 } \right)$	$0.5 < R_m$	0.503
where	$Ro^2 = 1 - \frac{\sum_{i=1}^n (t_i - h_i^o)^2}{\sum_{i=1}^n (t_i - t_i)^2}$	Should be close to 1	0.999
	$h_i^o = k \times t_i$		
	$Ro'^2 = 1 - \frac{\sum_{i=1}^n (h_i - t_i^o)^2}{\sum_{i=1}^n (h_i - h_i)^2}$ and $t_i^o = k' \times h_i$	Should be close to 1	0.987

$h_i$ : the actual outputs for the  $i$ th output.  $t_i$ : the calculated outputs for the  $i$ th output.  $n$ : the number of samples.  $\bar{h}_i$ : the average of the actual outputs.  $\bar{t}_i$ : the average of the calculated outputs

of incorporated predictor variables,  $L/D$ ,  $\rho_v$ ,  $\rho_l$   $\alpha = P/A_g f_c'$ ,  $\beta = \rho_v f_y / f_c'$ ,  $\gamma = \rho_l f_y / f_c'$ ,  $s/d$  and  $a/L_s$ . The results of the parametric analysis demonstrates that LGP model can perfectly capture the physical behavior of the system. This also represents the robustness of LGP as a soft computing algorithm for solving complicated engineering problems.

### 8.4 Sensitivity analysis

The crucial question when a model is used for estimation aims is the fact that how the output value varies as the values of each input variable changes. To explore this, a sensitivity analysis (SA) procedure is useful. This sort of analysis of the model can indicate how predictors which are independent input variables incorporated in a model impact the amount of changes of the dependent variable [6, 16, 25, 31, 32]. For this purpose, the sensitivity index (SI) can be utilized. In this study, the SI (%) can be obtained via the following equations:

$$N_i = f_{\max}(x_i) - f_{\min}(x_i) \tag{19}$$

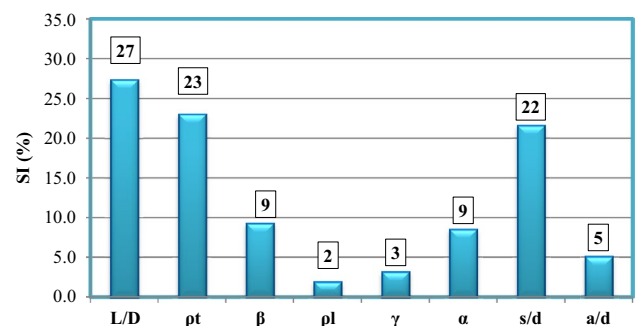
$$SI_i(\%) = \frac{N_i}{\sum_{j=1}^n N_j} \times 100, \tag{20}$$

where  $f_{\max}(x_i)$  is maximum value of the LGP model prediction for the  $i$ th input variable  $x$ , while  $f_{\min}(x_i)$  is calculated minimum value. These values are calculated by the model, where other inputs incorporated, except  $x_i$ , are considered constant at their mean values and  $n$  is the number of independent input variables in the model.

It can be suggested that that input variables with higher value of the SI have greater impact compared to other input variables on the model performance and the output value and vice versa. This shows the significance of variables. In

case the SI value become small or close to zero, the relevant variable poorly or does not influence on the model variations which means that it can be omitted. The SA process is carried out for the LGP model proposed in this study and the results are illustrated in Fig. 7.

Sezen [41] has stated that the shear deformation of RC columns greatly relevant to  $\rho_v$ ,  $a/L_s$ ,  $P$  and some other parameters [41]. In addition, the amount of transverse reinforcement radically impacts both the shear strength and ultimate drift capacity of RC columns [4, 6, 42]. A significant and new factor considered in this study was the slenderness ratio of the column, denoted by  $L/D$ . Although this factor can be regarded as a key parameter, it has not been incorporated in previous models proposed by researchers in the literature. As can be seen in Fig. 6, the SI (%) values obtained by the LGP model proposed in this study suggest that the  $L/D$  is the most influential input variable, followed by  $\rho_l$  and  $s/d$ , whereas the  $\rho_v$  and  $\gamma$  slightly impact the performance of the LGP-based  $\Delta_{max}$  model. The results precisely conform to those provided by researchers, e.g., [3, 10–13, 41]. It should be noted that the SI is a unique factor for every single of



**Fig. 7** SI (%) values input variables and their importance on the  $\Delta_{max}$  in LGP model

estimation models which should be given very much attention whenever an equation is used.

## 9 Summary and conclusion

This paper explored the applicability of a robust SC approach, called LGP, for predictive modeling of the  $\Delta_{\max}$  of circular RC columns with normal strength concrete. A valuable experimental database, PEER-SPD, has been utilized to develop LGP-based estimation models. The new LGP design equation proposed in this study incorporates  $L/D$ ,  $\rho_t$ ,  $\rho_1$ ,  $\alpha = P/A_g f'_c$ ,  $\beta = \rho_f f_y / f'_c$ ,  $\gamma = \rho_f f_y / f'_c$ ,  $s/d$  and  $a/L_s$  as independent input variables which have also been presented as key parameters by other researchers for indirect estimation of the  $\Delta_{\max}$  factor. The predictability and performance of the proposed LGP model was indicated through conducting some statistical analyses and was checked via comparing the results with some well-known models in the literature. It was stated that the model proposed by Saatcioglu and Razavi [13] was not able to estimate all data employed in this study and this is due to the fact that it is conditional and was developed based on simple regression-based analysis. The superiorities of the LGP model in comparison to other models are the consideration of more number of parameters to describe the  $\Delta_{\max}$  of circular RC columns, accounting for the column slenderness ratio as an input, higher accuracy in estimation and the predictability and generalization performance, and being a computational intelligence based model. The last means that the optimal LGP model is selected amongst numerous models after performing a comprehensive process of trial and error and improvement and optimization of the models which does not occur in regression-based numerical modeling methods. To represent the influence of each variable on the model, a sensitivity analysis was carried out. In this study, the slenderness ratio of the RC columns is also incorporated for the prediction of the  $\Delta_{\max}$ . The parametric analysis of the LGP model demonstrated that it can successfully capture the physical behavior off the system. The SA results presented that all variables contribute to the changes of  $\Delta_{\max}$  factor, where  $L/D$ ,  $\rho_t$  and  $s/d$  were the most significant influential variables. The accuracy of the LGP model proposed here taking into account several parameters compared to other conventional models confirm the capability of LGP approach for resolving complicated engineering problems. The authors are confident that the proposed LGP model is useful and valuable for the design of RC columns, particularly for pre-design aims.

## Appendix A

### A.1. The optimum LGP program for the prediction of $f'_c$ .

The following LGP program can be run in the Discipulus interactive evaluator mode or can be compiled in C++ environment. (Note:  $v[0]$ ,  $v$ ,  $v[3]$ , ...,  $v[7]$ , respectively, are  $L/D$ ,  $\rho_t$ ,  $\beta = \rho_f f_y / f'_c$ ,  $\rho_1$ ,  $\gamma = \rho_f f_y / f'_c$ ,  $\alpha = P/A_g f'_c$ ,  $s/d$  and  $a/L_s$ )

```
float DiscipulusCFunction(float v[])
{
long double f;
long double tmp = 0;
int cflag = 0;
f[0]=f=f[2]=f[3]=f[4]=f[5]=f=f[7]=0;
L0:  f[0]+=v[2];
L1:  f[0]*=v[6];
L2:  f[0]*=v[7];
L3:  f[0]-=v[3];
L4:  f[0]*=v[3];
L5:  f[0]+=v[4];
L6:  f[0]*=v[6];
L7:  f[0]-=v[7];
L8:  f[0]*=v[7];
L9:  f[0]*=v[1];
L10: f[0]/=v[4];
L11: f[0]-=v[5];
L12: f[0]*=v[3];
L13: f[0]+=v[0];
L14: f[0]/=v[6];
L15: f[0]-=v[5];
L16: f[0]*=v[0];
L17: f[0]+=v[0];
L19:
if (!_finite(f[0])) f[0]=0;
return f[0];
}
```

A.2. The database used in the present paper is attached as a supplementary file.

## References

1. ACI Committee 318-14 Building code requirements for structural concrete (ACI 318 M-14) and commentary (ACI 318RM-14). In: 2015, American Concrete Institute
2. FEMA 356 (2000) Prestandard and commentary for the seismic rehabilitation of buildings. Federal Emergency Management Agency, Washington DC
3. Mander J, Priestley M, Park R (1988) Observed stress–strain behavior of confined concrete. J Struct Eng 114(8):1827–1849
4. Pujol S, Sözen M, Ramirez J (2000) Transverse reinforcement for columns of RC frames to resist earthquakes. J Struct Eng 126(4):461–466
5. Caglar N, Garip ZS (2013) Neural network based model for seismic assessment of existing RC buildings. Comput Concr 12(2):229–241
6. Sadrossadat E, Ghorbani B, Hamooni M, Moradpoor Sheikhanloo MH (2018) Numerical formulation of confined compressive

- strength and strain of circular reinforced concrete columns using gene expression programming approach. *Struct Concr* 19(3):783–794
7. Priestley M, Park R (1987) Strength and ductility of concrete bridge columns under seismic loading. *Struct J* 84(1):61–76
  8. Inel M (2007) Modeling ultimate deformation capacity of RC columns using artificial neural networks. *Eng Struct* 29(3):329–335
  9. Zhu L (2005) Probabilistic drift capacity models for reinforced concrete columns. University of British Columbia, Columbia
  10. Elwood KJ, Moehle JP (2005) Drift capacity of reinforced concrete columns with light transverse reinforcement. *Earthq Spectr* 21(1):71–89
  11. Pujol S, Ramfrez J, Sozen MA (1999) Drift capacity of reinforced concrete columns subjected to cyclic shear reversals. *Spec Publ* 187:255–274
  12. Pujol S (2002) Drift capacity of reinforced concrete columns subjected to displacement reversals. Purdue University, West Lafayette
  13. Saatcioglu M, Razvi SR (2002) Displacement-based design of reinforced concrete columns for confinement. *Struct J* 99(1):3–11
  14. Sadrossadat E, Heidariapanah 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
  15. Sadrossadat E, Heidariapanah 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* 19(1):139–153
  16. Rostami MF, Sadrossadat E, Ghorbani B, Kazemi SM (2018) New empirical formulations for indirect estimation of peak-confined compressive strength and strain of circular RC columns using LGP method. *Eng Comput* 34(4):865–880
  17. 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
  18. Khandelwal M, Faradonbeh RS, Monjezi M, Armaghani DJ, Abd Majid MZB, Yagiz S (2017) Function development for appraising brittleness of intact rocks using genetic programming and non-linear multiple regression models. *Eng Comput* 33(1):13–21
  19. Sheikh SA, Uzumeri S (1982) Analytical model for concrete confinement in tied columns. *J Struct Div* 108(12):2703–2722
  20. Gordon N (2015) Prediction of shear strength and ductility of cyclically loaded reinforced concrete columns using artificial intelligence. Dissertation, University of Nevada, Las Vegas
  21. 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
  22. 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
  23. Gandomi AH, Yang XS, Talatahari S, Alavi AH (2013) Metaheuristic applications in structures and infrastructures. Elsevier, London
  24. Gandomi AH, Yun GJ, Alavi AH (2013) An evolutionary approach for modeling of shear strength of RC deep beams. *Mater Struct* 46(12):2109–2119
  25. Gandomi AH, Roke DA (2015) Assessment of artificial neural network and genetic programming as predictive tools. *Adv Eng Softw* 88:63–72
  26. Gandomi AH, Alavi AH, Ryan C (2015) Handbook of genetic programming applications. Springer, Switzerland
  27. Sadrossadat E, Basarir H (2019) An evolutionary-based prediction model of the 28-day compressive strength of high-performance concrete containing cementitious materials. *Adv Civil Eng Mater* 8(3):484–497
  28. Sadrossadat E, Basarir H, Karrech A, Durham R, Fourie A, Bin H (2019) The optimization of cemented hydraulic backfill mixture design parameters for different strength conditions using artificial intelligence algorithms. In: Proceedings of the 28th international symposium on mine planning and equipment selection - MPES 2019, 2020, pp 219–227
  29. Koza JR (1992) Genetic programming II, automatic discovery of reusable subprograms. MIT Press, Cambridge
  30. Brameier MF, Banzhaf W (2007) Linear genetic programming. Springer, New York
  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. Alavi AH, Ameri M, Gandomi AH, Mirzahosseini MR (2011) Formulation of flow number of asphalt mixes using a hybrid computational method. *Constr Build Mater* 25(3):1338–1355
  33. 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
  34. Berry M, Parrish M, Eberhard M (2004) PEER structural performance database, user's manual (version 1.0). University of California, Berkeley
  35. Baradaran Shoraka M, Elwood K (2013) Mechanical model for non ductile reinforced concrete columns. *J Earthq Eng* 17(7):937–957
  36. 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
  37. Armaghani DJ, Faradonbeh RS, Momeni E, Fahimifar A, Tahir MM (2018) Performance prediction of tunnel boring machine through developing a gene expression programming equation. *Eng Comput* 34(1):129–141
  38. Conrads M, Dolezal O, Francone F, Nordin P (2004) Discipulus—fast genetic programming based on AIM learning technology. Register Machine Learning Technologies Inc, Littleton CO
  39. Smith GN (1986) Probability and statistics in civil engineering. Nichols Publishing Company, New York
  40. Ghorbani B, Sadrossadat E, Bazaz JB, Oskooei PR (2018) Numerical ANFIS-based formulation for prediction of the ultimate axial load bearing capacity of piles through CPT data. *Geotech Geol Eng* 36(4):2057–2076
  41. Sezen H (2008) Shear deformation model for reinforced concrete columns. *Struct Eng Mech* 28(1):39–52
  42. Vu NS, Yu B, Li B (2016) Prediction of strength and drift capacity of corroded reinforced concrete columns. *Constr Build Mater* 115:304–318

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.