Neural Computing and Applications (2024) 36:9481–9501
https://doi.org/10.1007/s00521-024-09578-1

ORIGINAL ARTICLE

A prediction model for plastic hinge length of rectangular RC columns using gene expression programming

Sema Alacali[1] · Guray Arslan[1]

Received: 25 August 2022 / Accepted: 5 February 2024 / Published online: 4 March 2024
© The Author(s) 2024

Abstract
In the reinforced concrete (RC) columns which are exposed extreme loads such as earthquake effects, the plastic hinge length can be defined as the length of the region where flexural moments exceed the yielding capacity, and the plastic deformations are concentrated. More accurate estimation of plastic hinge length increases the reliability of the seismic design. However, a sensitivity prediction of plastic hinge length is difficult due to a large number of model parameters. Therefore, this study aims to predict the plastic hinge length using the gene expression programming (GEP). An experimental database of 133 RC columns gathered from the literature was utilized for prediction with GEP. The results of GEP model are statistically compared with those of 13 models existing in the literature proposed by various researchers. The comparison results reveal that the proposed GEP-based formulation has the best efficiency among all models. Furthermore, a sensitivity analysis and parametric study are conducted to identify the most influential parameters affecting the GEP formulation.

Keywords Gene expression programming · Plastic hinge length · Reinforced concrete column, sensitivity analysis

1 Introduction

The performance of reinforced concrete (RC) columns plays a crucial role in the overall performance of structural systems with their load transfer, stability and durability properties [1, 2]. Columns transfer the loads from beams and slabs to foundation in structures. Seismic failure modes of the RC columns are classified into three major types: shear, bond splitting and flexural modes. In flexural failure mode, the plastic deformation is localized in a small region namely the plastic hinge region after the yielding of the member. The performance of the plastic hinge region is critical for flexural members as it governs the load carrying and deformation capacities of the column. The length of the plastic hinge region is an important design parameter where intense confinement is used to increase the ductility of the member. Therefore, accurate estimation of the length of this region will contribute significantly to a more realistic prediction of the behavior of the structure. When a column is subjected to large earthquake loading, significant inelastic curvatures in the end regions of the column may occur. It is assumed that the inelastic curvature, which is distributed nonlinearly, remains constant over a specific length known as the plastic hinge length (L_p). The L_p is a critical parameter for seismic analysis and the design of RC structures [3–5]. Inaccurate prediction of the L_p may result in inadequate seismic design, reducing the ability of the structure to withstand dynamic forces. The prediction of plastic hinge length L_p of the RC columns in real-world scenarios has a significant effect on the design, evaluation, and strengthening of reinforced concrete columns. It enables cost-effective solutions and improves engineering techniques by optimizing the use of materials. The prediction models may assist in the development of building codes, encourage further structural engineering research and affect how structures are built in the future.

In general, there are many factors that can affect L_p. The factors affecting the L_p in RC members subjected to monotonic loading could be different from those of columns under cyclic loading [6, 7]. Due to material uncertainties and complex interactions among constituent

✉ Guray Arslan
aguray@yildiz.edu.tr

[1] Department of Civil Engineering, Yildiz Technical University, 34220 Istanbul, Turkey

materials, extensive research has been conducted on the plastic hinge behavior of RC columns, employing both experimental and analytical approaches [8]. First empirical models in the literature for the L_p were proposed by Baker [9], Baker and Amarakone [10], Corley [11] and Mattock [12], respectively. These models reveal that two crucial factors such as the shear span and the section depth have a primary impact on the L_p. Since these formulations were calibrated mainly based on experimental tests on beam elements, the fixed-end rotation contribution was not clearly evaluated. In formulations calibrated related to the RC column, this contribution is explicitly and prominently represented [13]. Paulay and Priestley [14] proposed an equation by considering the effect of the displacement due to the longitudinal reinforcement slipping out of the joint or foundation.

Besides, the use of different confinement models may significantly influence on the L_p. For example, Tanaka [15] is concerned with the effects of lateral confinement reinforcement on the ductile behavior of RC columns. Ho and Pam [16] examined the post-elastic behavior of low axially loaded high-strength RC columns with transverse reinforcement. Légeron and Paultre et al. [17] demonstrated that the increase in L_p with the axial loads could also be observed for high-strength concrete columns. Ho [18] introduced the effect of the transverse and longitudinal reinforcement ratios, axial load level, compressive strength of concrete, yield strength of longitudinal bar, and cross-sectional depth on the L_p. In the parametric study by Ou et al. [3], it is examined the effects on the L_p of circular RC columns of the longitudinal reinforcement ratio, the shear span-to-depth ratio, the axial force ratio, and the concrete compressive strength. This study, where the yield strengths of the longitudinal reinforcement are taken as 414 MPa and 685 MPa, shows that L_p increases when the axial load, longitudinal reinforcement, and shear span-to-depth ratios increase. In addition, Ou et al. [3] have shown that as the concrete strength increases, the L_p decreases for the 414 MPa case but increases for the 685 MPa case. Mortezaei [19] performed hundreds of elastic-time history analyses to predict plastic hinge length of RC columns under the combined effect of near-fault vertical and horizontal ground motions, by considering influence of various parameters such as axial load level, height-depth ratio, and strength of concrete. Geng and Zhou [20] presented a predictive model to evaluate the effective plastic hinge length for existing RC columns under lateral load reversals, using data from 24 test columns to evaluate the effects of slenderness ratio, axial load and transverse reinforcement. In the study of Barbagallo et al. [21], the most appropriate value for the L_p was calibrated by comparing laboratory and numerical test results. Based on optimal L_p values obtained from laboratory tests specifically to identify elements that accurately represent columns in existing RC buildings, a simple relationship depending on the volumetric ratio of the transverse reinforcement is proposed to determine L_p. In this study, it is emphasized that the optimum L_p decreases with the increase in the volumetric ratio of the transverse reinforcement.

Studies have been carried out to determine the L_p with machine learning algorithms, which have been widely used in recent years. Wakjira et al. [22] used various ensemble machine learning algorithms to propose a model. Feng et al. [23] used the adaptive boosting method based on an ensemble machine learning algorithm to predict the L_p of RC columns. The performance of the model has been assessed using the tenfold cross-validation technique. Bayrak et al. [24] has extensively studied the use of the Levenberg–Marquardt Algorithm with artificial neural network (ANN) models to predict the L_p of columns and shear walls. The findings revealed that the proposed ANN model represents a practical and valuable approach to predict the L_p of both columns and shear walls.

Genetic Expression Programming (GEP), one of the machine learning approaches in evolutionary computing, combines the components of functional programming with genetic algorithms to produce extraordinary, efficient and comprehensive results. Unlike traditional statistical methods and some machine learning models, GEP develops mathematical expressions for a given data set, thanks to its flexible gene structure that adapts to variable lengths during the optimization process. The evolutionary nature of GEP enhances the model's ability to more effectively explore various nonlinear functions. This feature allows direct understanding of models by providing a clear and interpretable form in scientific research or engineering design. This is in contrast to many black box machine learning models, where the relationship between input features and output can be difficult to understand.

The aim of this study is to derive a new equation to predict plastic hinge length L_p by GEP. As can be seen from the above expressions, major factors affecting on the L_p are sectional depth, axial force level, aspect ratio, longitudinal reinforcement ratio, concrete compressive strength, and transverse reinforcement ratio. However, none of the proposed equations include all the above parameters. Therefore, it is aimed to predict a new equation by considering all the important parameters affecting L_p. In the prediction of L_p, 133 RC column specimens collected by Ning and Li [25] were taken into account. The

Table 1 Experimental database and ranges of parameters [25]

Number	References	B(mm) min max	H(mm) min max	L(mm) min max	f_c'(MPa) min max	d_b(mm) min max	ρ_s min max	f_y(MPa) min max	s(mm) min max	ρ_v min max	f_{yv}(MPa) min max	P/P_o min max	L/H min max	$L_{p,exp.}$(mm) min max
4	Bae and Bayrak [26]	440	440	3050	36.5	16	0.0125	400	86	0.0072	427	0.17	5.00	286.70
		610	610		43.4	22		503	152	0.0204	496	0.43	6.93	402.60
8	Pam and Ho [27]	325	325	1895	50.0	12	0.0610	339	70	0.0173	339	0.12	5.83	299.00
					96.4	32	0.0090	531		0.0320	572	0.63		666.25
6	Sheikh and Khoury [28]	305	305	1841	31.3	19	0.0244	508	95	0.0130	464	0.47	6.04	259.25
					33.2				114	0.0306	508	0.77		353.80
3	Sheikh et al. [29]	305	305	1841	53.6	19	0.0244	508	76	0.0168	464	0.62	6.04	271.45
					54.7				108	0.0430	508	0.64		329.40
10	Yang et al. [30]	300	300	1200	72.0	16	0.0298	483	50	0.0120	952	0.45	4.00	261
					88.0				80	0.0248		0.57		402
8	Elmenshawi et al. [6]	250	350	630	26.6	15	0.0137	494	140	0.0080	467	0	1.80	269.50
				900	175	20	0.0206			0.0160			2.57	420
8	Paultre et al. [31]	305	305	2000	78.7	18	0.0215	446	55	0.0196	418	0.35	6.56	311.10
					109.5			481	130	0.0426	825	0.53		600.85
40	Barrera et al. [32]	150	140	1500	29.5	8	0.0140	531	50	0.00800.0310	530	0.00	7.50	70.00
			200		100.4	12	0.0320	538	100		531	0.59	10.71	340.20
8	Tanaka [15]	400	400	1650	25.6	20	0.0125	474	80	0.0170	325	0.10	3.00	172.00
		550	550	1800	32.1	20	0.0157	511	110	0.0255	333	0.30	4.50	473.00
4	Bayrak and Sheikh [33]	250	350	2010	74.1	20	0.0274	521	75	0.0183	850	0.34	5.74	297.50
					74.2				140	0.0354		0.53		350.00
6	Legeron and Paultre [17]	305	305	2150	92.4	18	0.0215	456	60	0.01960.0426	391	0.14	7.05	311.10
					104.3			469	130		418	0.39		741.15
4	Ohno and Nishioka [34]	400	400	1600	24.8	19	0.0082	362	100	0.0032	325	0.03	4.00	400.00
												0.08		
4	Ho and Pam [16]	325	325	1895	83.2	12	0.0086	556	85	0.0038	339	0.11	5.83	256.75
					89.6				220	0.0173	378			279.50
20	Bayrak [35]	250	250	2010	56.2	20	0.0258	454	70	0.0183	463	0.31	5.74	247.50
		350			112.1		0.0274	521	150	0.0672	542	0.51	8.04	444.50

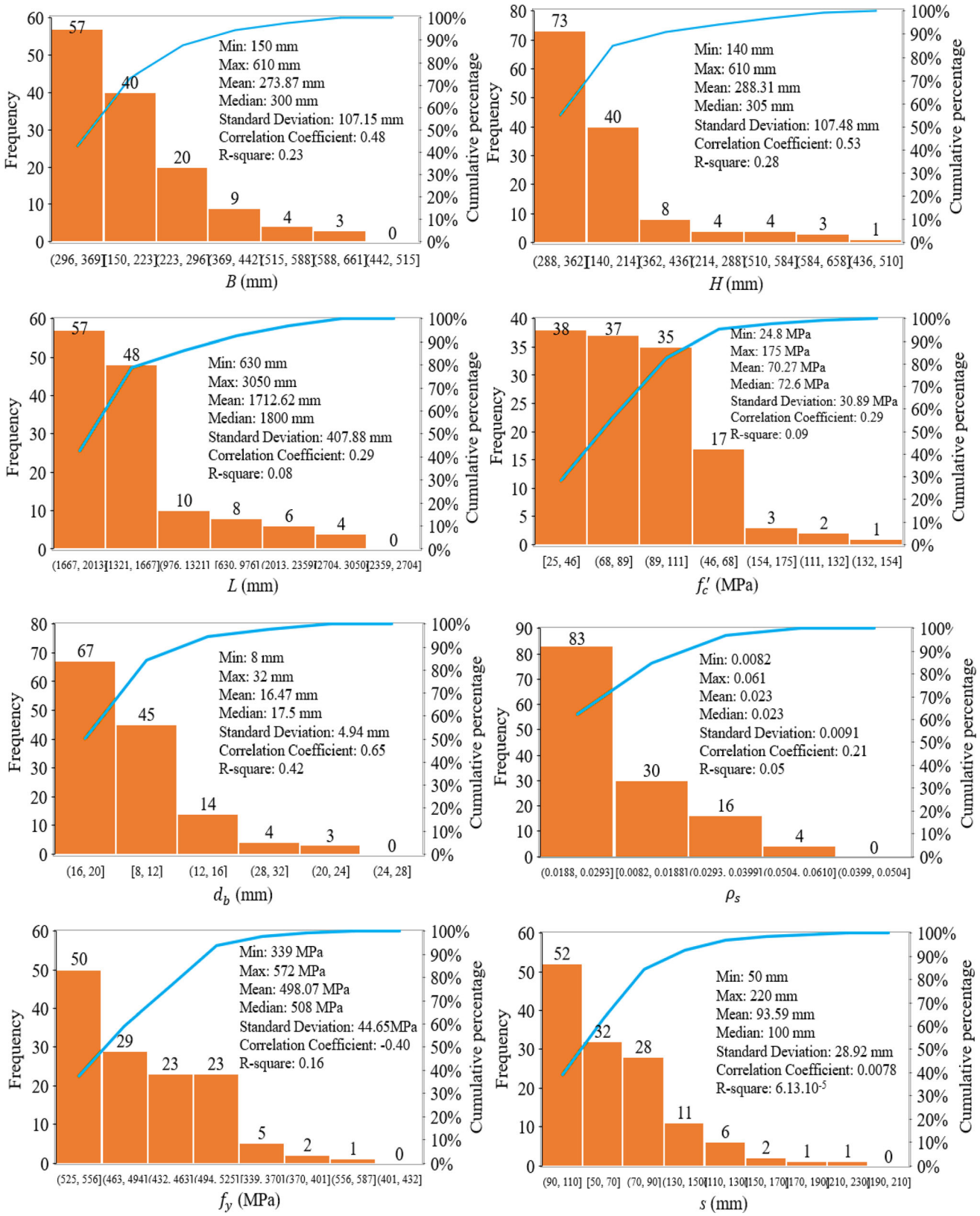


Fig. 1 Frequency histograms of each input and output parameter in the GEP model

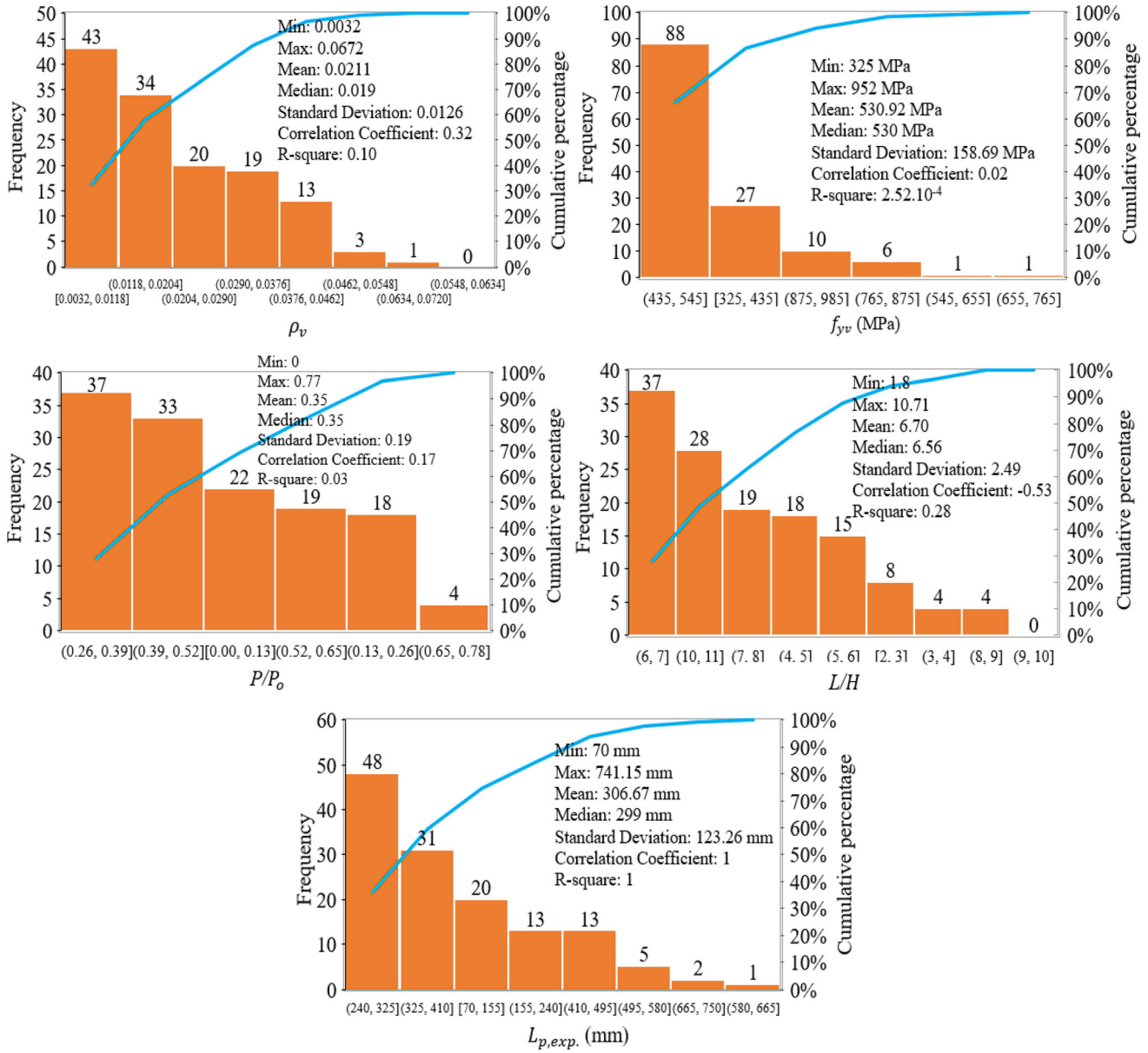


Fig. 1 continued

performance of the GEP model has been demonstrated by comparing the experimental and the results of the proposed equations in the literature. Besides, sensitivity and parametric analysis were performed to examine the cumulative effect of each input parameter on the overall performance and predictive capacity of the proposed GEP models. This detailed analysis has contributed to a comprehensive understanding of the complex relationships between the input variables and the resulting model outputs.

2 Summary of experimental data

In this study, the experimental database of 133 RC rectangular column specimens was taken from the study of the Ning and Li [25] to predict formulation with the GEP model. Table 1 shows the experimental database and ranges of parameters considered in this study. As seen in Table 1, the upper and lower limits of the design variables comprise a wide enough range of values. These parameters are cross-sectional width of the column B (mm), cross-sectional depth of the column H (mm), shear span L (mm), compressive strength of concrete $f_c^{'}$ (MPa), the diameter of

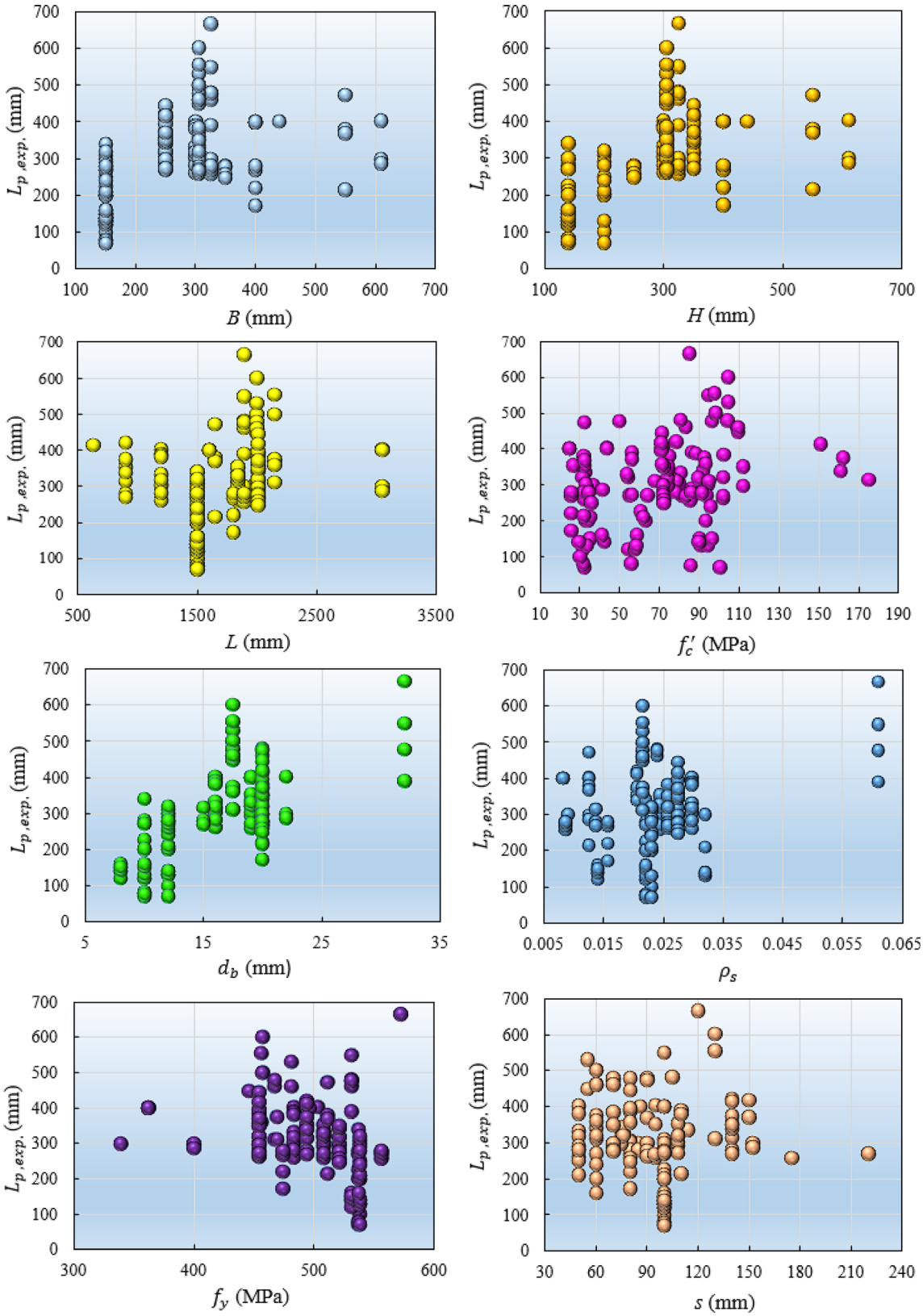


Fig. 2 Relationship between each input parameter and the $L_{p,exp.}$ of the RC columns included in the experimental database

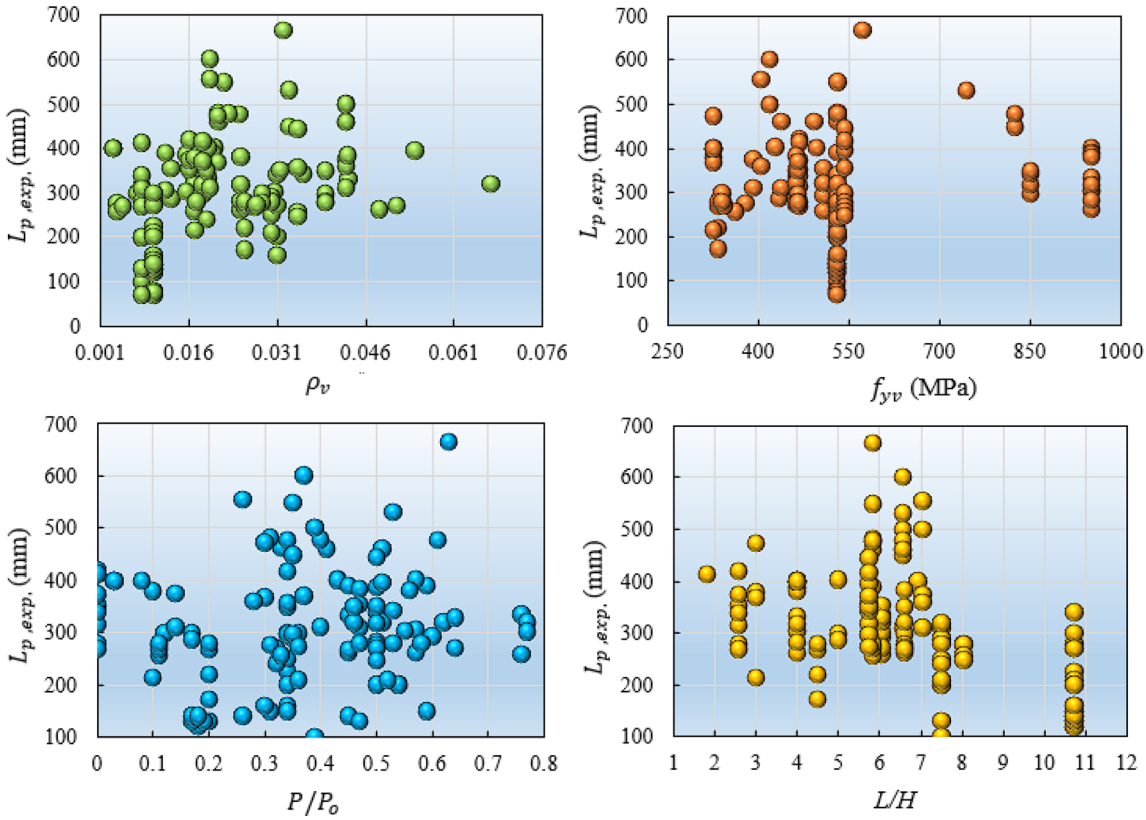
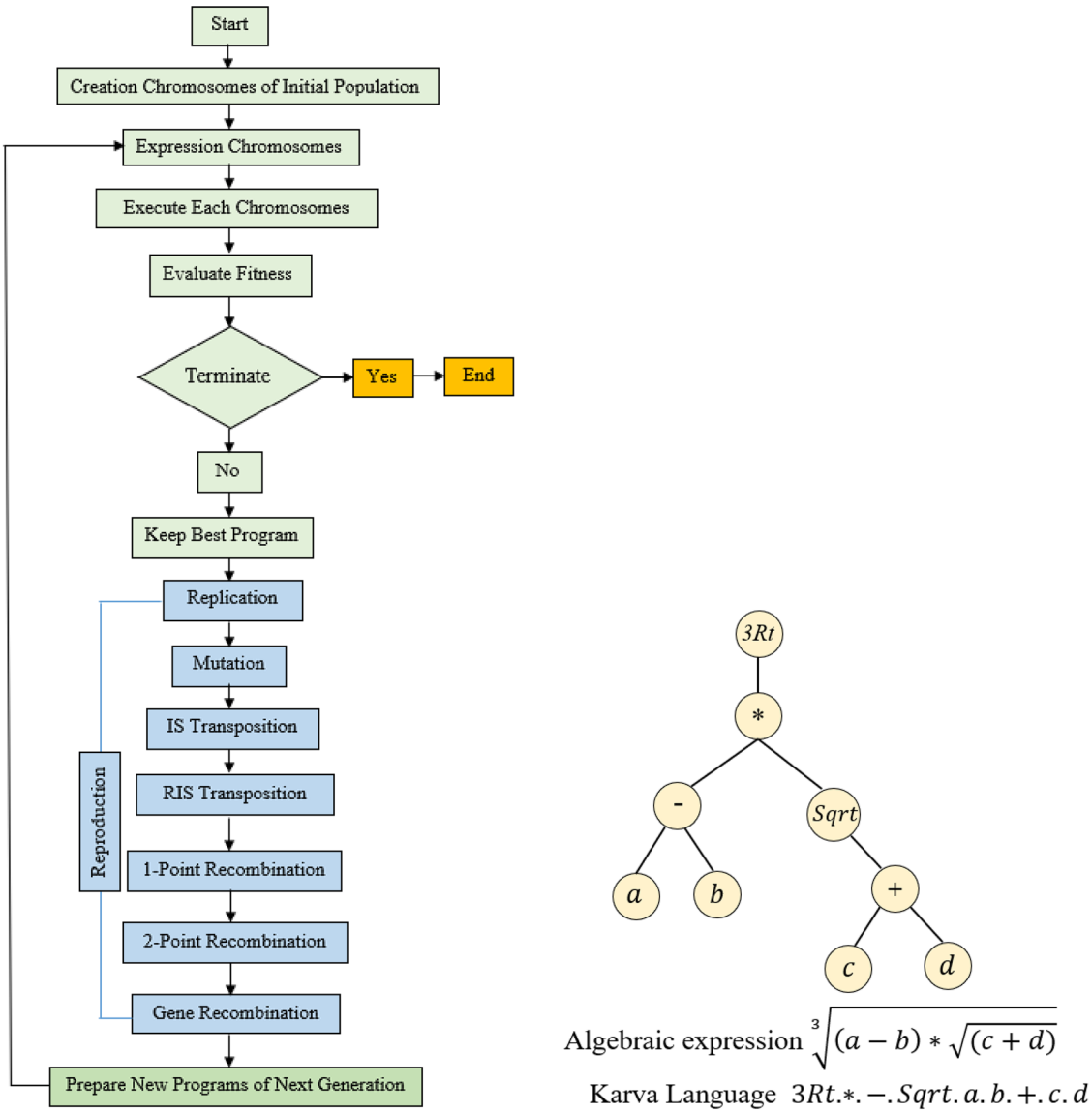


Fig. 2 continued

Table 2 Proposed models for L_p of RC columns

Model	Plastic hinge length, (L_p)
Priestley and Park [36]	$L_p = 0.08L + 6d_b$
Paulay and Priestly [14]	$L_p = 0.08L + 0.022f_y d_b$
Sheikh and Khoury [28]	$L_p = 1.0H$
Priestley et al. [37]	$L_p = 0.08L + 0.022f_y d_b \leq 0.044f_y d_b$
Panagiotakos and Fardis [38]	$L_p = 0.12L + 0.014\alpha_{sl} f_y d_b$
Ho [18]	$L_p = \left[20\sqrt{P/P_o}\left(f_c^{'}/f_y\right)^{1.5}(\rho_s/\rho_v)^{0.5} + 0.6\right]H$
Lu et al. [1]	$L_p = 0.077L + 8.16d_b$
EN 1998–3 EC8 [39]	$L_p = L/30 + 0.2H + 0.11d_b f_y/\sqrt{f_c^{'}}$
Berry and Eberhard [40]	$L_p = 0.05L + 0.1f_y d_b/\sqrt{f_c^{'}} \leq 0.25L$
Berry et al. [41]	$L_p = 0.0375L + 0.12f_y d_b/\sqrt{f_c^{'}}$
Bae and Bayrak [26]	$L_p = (0.3P/P_0 + 3\rho_s - 0.1)L + 0.25H \geq 0.25H$
Biskinis and Fardis [42]	$L_p = 0.2H[1 + (1/3)\text{minimum}(9, L/H)]$
Ning and Li [25]	$L_p = (0.042 + 0.072P/P_0)L + 0.298H + 6.407d_b$



a) The flowchart of a gene expression algorithm **b)** Expression tree of a chromosome

Fig. 3 Flowchart of a gene expression algorithm and expression tree of a chromosome [43]

longitudinal bars d_b (mm), longitudinal reinforcement ratio ρ_s, yield strength of the longitudinal reinforcement f_y (MPa), spacing of transverse reinforcement s (mm), transverse reinforcement ratio ρ_v, yield strength of transverse reinforcement bars f_{yv}(MPa), axial load ratio P/P_o, and aspect ratio L/H, respectively.

Also, Fig. 1 shows frequency histograms of each input parameter taken into account in the GEP model.

In addition, Fig. 2 shows the relationship between each input parameter and the experimental plastic hinge lengths ($L_{p,exp.}$) of the columns included in the experimental database.

3 Existing models for plastic hinge length of the RC columns

In the literature, many researchers have proposed different formulations in prediction of the L_p. However, substantial disparities exist among these various formulations, and there has been no systematic assessment of their accuracy in predicting of the L_p [26]. A summary of the most frequently used formulas in the literature regarding the L_p of the columns is shown in Table 2. As seen in, the number of the input parameters which are included in the proposed equations shows differences.

Table 3 Parameters used in GEP model

Parameter definition	Values
Input parameters: B(mm), H(mm), L(mm), f_c'(MPa), d_b(mm), ρ_s, f_y(MPa), s(mm), ρ_v, f_{yv}(MPa), $P/P_o, L/H$	Output parameter: L_p(mm)
Genes	4
Function set	$+, -, *, /, \sqrt{}, \tan$
Chromosomes	30
Head Size	10
Linking function between ETs	Addition
Mutation	0.00138
Inversion rate	0.00546
One point recombination rate	0.00277
Two-point recombination rate	0.00277
Gene recombination	0.00277
Training records	113
Validation records	20

The form of the commonly used equation for the L_p of RC columns is given below [7, 13].

$$L_p = \alpha L + \beta(f_y d_b) \quad (1)$$

In the two-term relationship of the L_p, the first term includes the effects of bending moment distribution along the L, whereas the second term considers the yield penetration of the longitudinal reinforcement into the column base or the joint [7]. α and β in Eq. 1 are model coefficients proposed differently by various researchers. The models proposed by researchers, shown in Table 2, are discussed in the following sections.

Priestley and Park [36] proposed an equation based on the L and the d_b. The values of $\alpha = 0.08$ and $\beta = 6/f_y$ in the form of the Eq. 1 represent the model coefficients, respectively. The relationship proposed by Paulay and Priestley was obtained by taking the model coefficients $\alpha = 0.08\, and\, \beta = 0.022$ in the form of Eq. 1. Due to their experiments on columns subjected to high axial loads, Sheikh and Khoury [28] introduced an equation predicting that L_p is approximately equivalent to the H of the column, without accounting for the influence of any other parameters. Priestley et al. [37] proposed the same equation derived by Paulay and Priestley [14] for the L_p and limited the result of the equation to $0.044 f_y d_b$. Panagiotakos and Fardis [38] proposed a equation to predict the L_p for RC columns subjected to cyclic loads. This equation includes model coefficients of $\alpha = 0.12$ and $\beta = 0.014\alpha_{sl}$, with the constant α_{sl} representing the fixed-end rotation due to the longitudinal bar slippage. If the slippage of longitudinal reinforcement bars from its anchorage zone is 1, otherwise α_{sl} is 0. Ho [18] derived an equation for theL_p, taking into account the parameters of the $H, f_c', f_y, \rho_s, \rho_v$, and P/P_o. Lu et al. [1] took into account the form of the equation proposed by Priestley and Park [36], by taking the model coefficients of $\alpha = 0.077$ and $\beta = 8.16/f_y$, respectively. In EN 1998–3 EC8 [39], L_p was given for members with detailing for earthquake resistance and no lapping of longitudinal bars near the section where yielding is expected, taking into account the parameters of the L, H, f_c', d_b, and f_y. In the form of Eq. (1), the model coefficients $\alpha = 1/30$ and $\beta = 0.11/\sqrt{f_c'}$ are taken, respectively. Berry and Eberhard [40] proposed an equation for the L_p and limited the result of the equation to $0.25L$. Thus, the model coefficients calibrated with experimental results correspond to $\alpha = 0.05$ and $\beta = 0.1/\sqrt{f_c'}$, respectively. Berry et al. [41] revised the same equation by taking the model coefficients of $\alpha = 0.0375$ and $\beta = 0.12/\sqrt{f_c'}$, respectively. Bae and Bayrak [26] demonstrated that the P/P_o, L, H and ρ_s are the main parameters in the proposed equation using the least squares method for L_p. It has been observed that the L_p is equal to approximately $0.25H$ at low axial loads and increases at higher axial load levels. It was also observed that for a given P/P_o, the L_p increased as the L/H increased. Biskinis and Fardis [42] proposed the L_p as a function of the H and L/H. Ning and Li [25] used a method to determine the L_p by a probabilistic approach. They proposed an equation for the L_p by taking into account the P/P_o, L, H, and the d_b.

4 Gene expression programming

GEP which combines genetic programming (GP) and genetic algorithm (GA) is a new technique in which individuals are encoded as fixed-length linear sequences expressed as nonlinear entities of different sizes and shapes [43]. Gene Expression Programming (GEP) algorithm consists of five components: function set, terminal set, fitness function, control parameters and termination condition. To summarize, the function set includes various mathematical operations and functions. The terminal set consists of variables, constants, and other basic elements that construct expressions on chromosomes. The fitness function plays an especially key role in evaluation by measuring the difference between the output and the desired output. Control parameters include factors such as

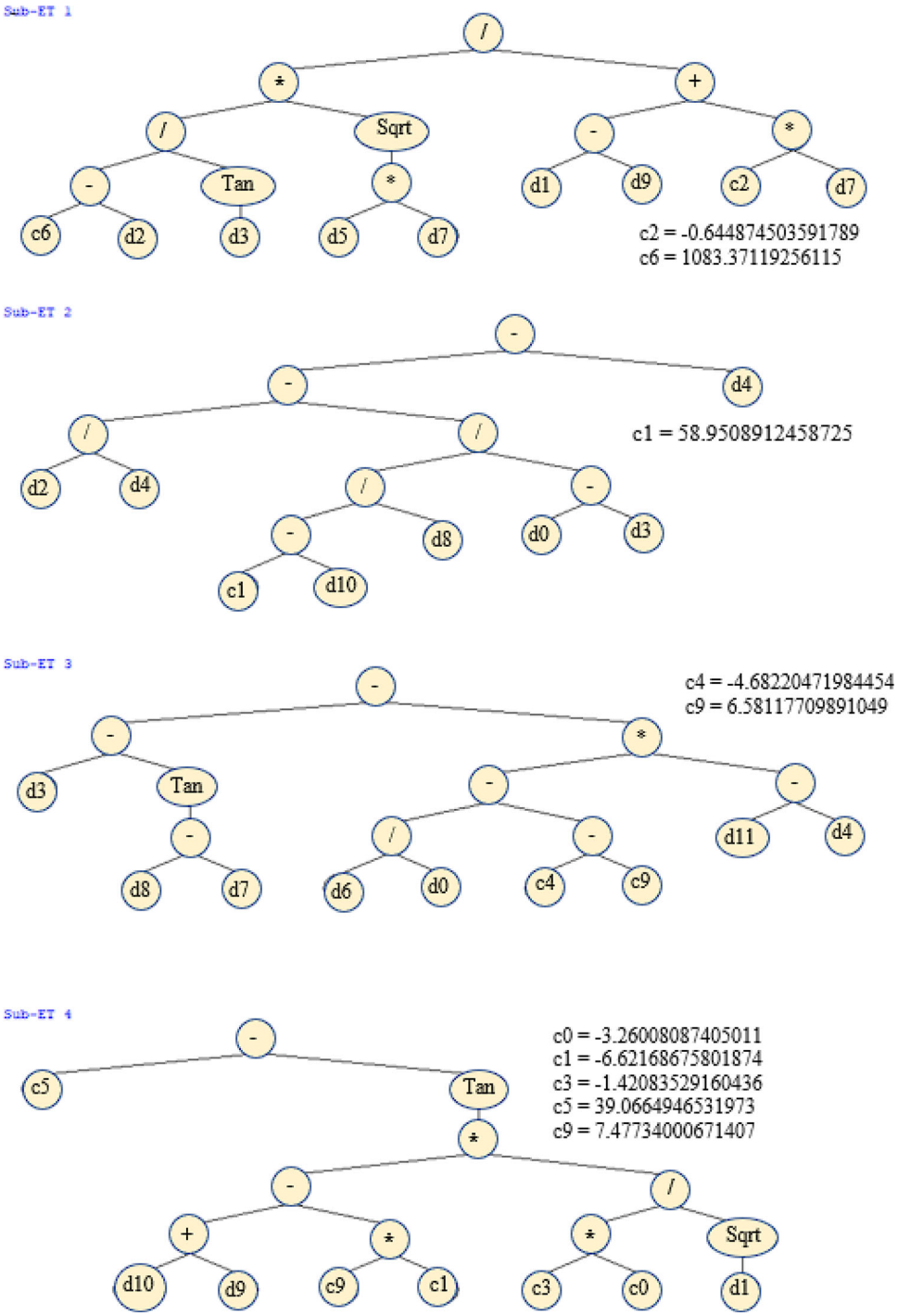


Fig. 4 Expression trees of GEP model for L_p

population size, mutation rate, transmission rate, which affect the manipulation of genetic material. The termination condition determines when its evolutionary process should stop. GEP may define complicated mathematical expressions that are tough to see in the data since it combines functions and terminals. It assumes that a suitable mathematical expression can be used to model the data and that it precisely represents how the gene functions on unobserved data [43–45].

The flowchart shown in Fig. 3a presents the main steps of GEP. The process begins with the creation of the chromosomes of the initial population and then, the fitness of each individual is evaluated. Individuals are then selected based on their fitness for reproduction. The process is repeated for a certain number of generations or until

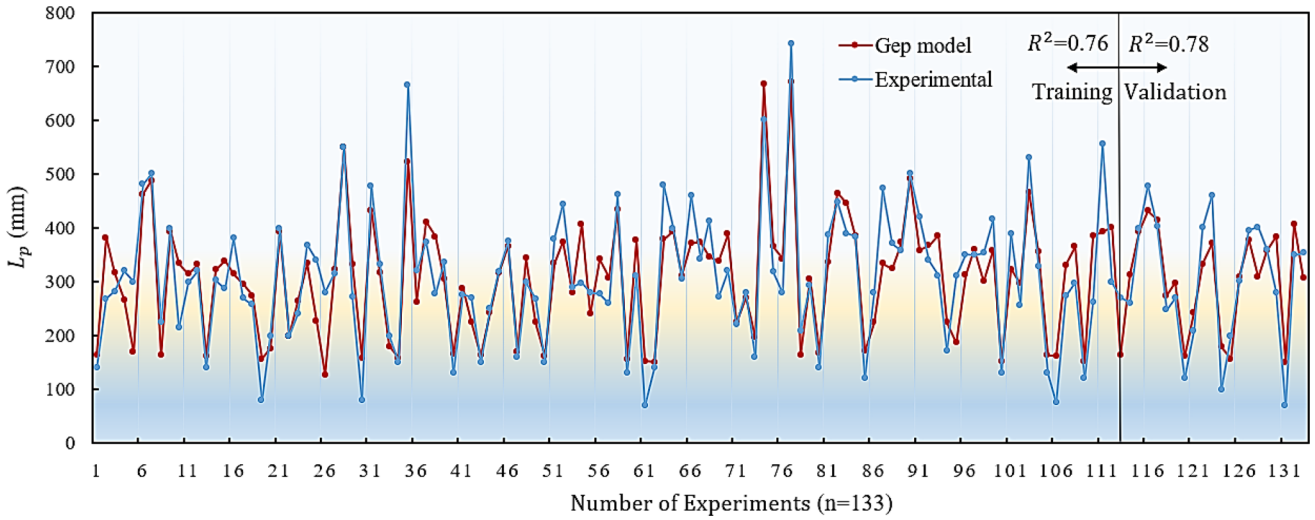


Fig. 5 Comparison of the performance of the GEP model and experimental results for training and validation sets

Table 4 Comparison of the statistical parameters for the L_p

Model	*M*	*SD*	*MAPE*	*RMSE*	R^2	*COV*
Proposed Model (all data)	1.07	0.29	19.67	60.17	0.76	0.27
Priestley and Park [36]	0.90	0.42	32.81	125.56	0.30	0.47
Paulay and Priestly [14]	1.20	0.56	37.22	101.30	0.32	0.46
Sheikh and Khoury [28]	1.03	0.41	26.64	114.20	0.28	0.40
Priestley et al. [37]	1.18	0.55	36.15	100.04	0.34	0.47
Panagiotakos and Fardis [38]	1.23	0.60	40.90	107.34	0.25	0.49
Ho [18]	1.21	0.60	36.17	124.86	0.44	0.49
Lu et al. [1]	1.01	0.46	31.46	108.01	0.34	0.45
EN 1998–3 EC8 [39]	0.87	0.41	34.00	136.30	0.17	0.48
Berry and Eberhard [40]	0.74	0.38	40.11	164.11	0.11	0.52
Berry et al. [41]	0.73	0.39	40.58	165.33	0.10	0.53
Bae and Bayrak [26]	0.75	0.39	40.00	156.35	0.20	0.52
Biskinis and Fardis [42]	0.63	0.28	42.69	174.86	0.26	0.44
Ning and Li [25]	1.13	0.43	30.86	96.02	0.39	0.38

a proper solution is found [43, 46]. All of the genes in GEP can code expression trees (ETs) with different shapes and sizes. Expression trees are put together by using the linking functions. The ET illustration of a chromosome is shown in Fig. 3b. The algebraic explanation ($\sqrt[3]{(a-b)} * \sqrt{(c+d)}$) and the Karva language ($3Rt. * . - .Sqrt.a.b. + .c.d$) of the ET are also seen in the same figure. The Karva language is formed by writing from left to right and top to bottom along the top line of the ET [43].

In addition, GEP uses chromosomes composed of genes structurally organized in a head and a tail. The head contains constant, parameters, functions, and mathematical operators such as {1, *a, b, Sqrt,3Rt, tan,* + , − , *, /}. The tail contains only terminals: constant and parameters such as {1, *a, b*}. In GEP applications such as symbolic regression or function finding, depending on whether the error chosen is absolute or relative, the fitness f_i of an individual program i is expressed by Eqs. (2)–(3).

$$f_i = \sum_{j=1}^{C_t} \left(M - \left|C_{(i,j)} - T_{(j)}\right|\right) \tag{2}$$

$$f_i = \sum_{j=1}^{C_t} \left(M - \left|\frac{C_{(i,j)} - T_{(j)}}{T_{(j)}}.100\right|\right) \tag{3}$$

where M is the range of selection, $C_{(i,j)}$ the value returned by the individual chromosome i for fitness case j, and T_j is the target value for fitness case j. Note that for a perfect fit $C_{(I,j)} = T_j$ and $f_i = f_{\max} = C_t\ M$ [43].

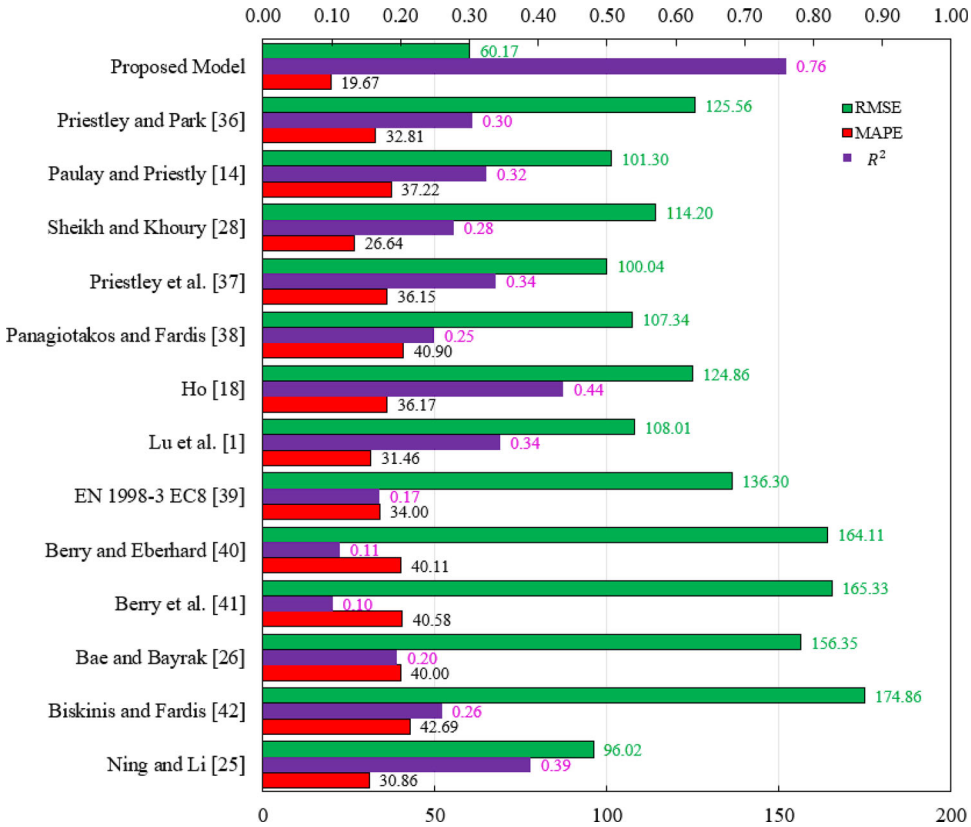


Fig. 6 Statistical comparison between GEP and other proposed models for L_p

4.1 Gene expression model and parameter selection

Most of the empirical models in the literature have been derived based on a limited number of experimental results. Experimental database was utilized in GeneXproTools 5.0 software to develop the GEP model in this study [47]. The program was run until the finest model was obtained. The experimental database in Table 1 is taken into account in the prediction of the L_p. The summary of the parameters used in GEP model is given in Table 3. In training and validation, B, H, L, f'_c, d_b, ρ_s, f_y, $s, \rho_v, f_{yv}, P/P_o$ and L/H are selected as input parameters, while L_p is used as output (target) parameter. The database parameters were randomly separated into two sets as training and validation. Thus, 85% of the whole experimental database (113 data) was used as training data whereas the remaining 15% (20 data) were employed as validation data for verifying the validity of the GEP model. The root mean square error (*RMSE*) constitutes a prevalent metric employed to quantify the disparities between predicted and target values. Root mean square error (*RMSE*) was considered as a fitness function in the GEP model.

In Genetic Expression Programming (GEP) models, sub-expression trees (Sub-ETs) contribute to the overall functionality of the model by providing a structured way to represent and process genetic information. In other words, it represents the basic building blocks of the model and captures the terminal input parameters (di for i = 0, 1, 2, …). These parameters represent variables or features that the model uses to make predictions or perform calculations. Constants (c0, c1, …) add constant values to expression trees, providing stability and diversity to genetic material. Mathematical operations such as addition, subtraction, multiplication, and division involve mathematical or logical functions that operate on genetic information, introducing complexity and variability. These functions enable the model to perform complex calculations and transformations on input parameters and constants. In this study, all models consist of four different sub-expression trees (Sub-ETs) are linked by addition function. For the derivation of the GEP model, input parameters di and the constants (ci) have been parsed into expression trees as shown in Fig. 4,

In the presented ETs, d0 represents B (mm), d1 is H (mm), d2 is L (mm), d3 is f'_c (MPa), d4 is d_b (mm), d5 is ρ_s, d6 is f_y(MPa), d7 is s (mm), d8 is ρ_v, d9 is f_{yv}(MPa), d10 is

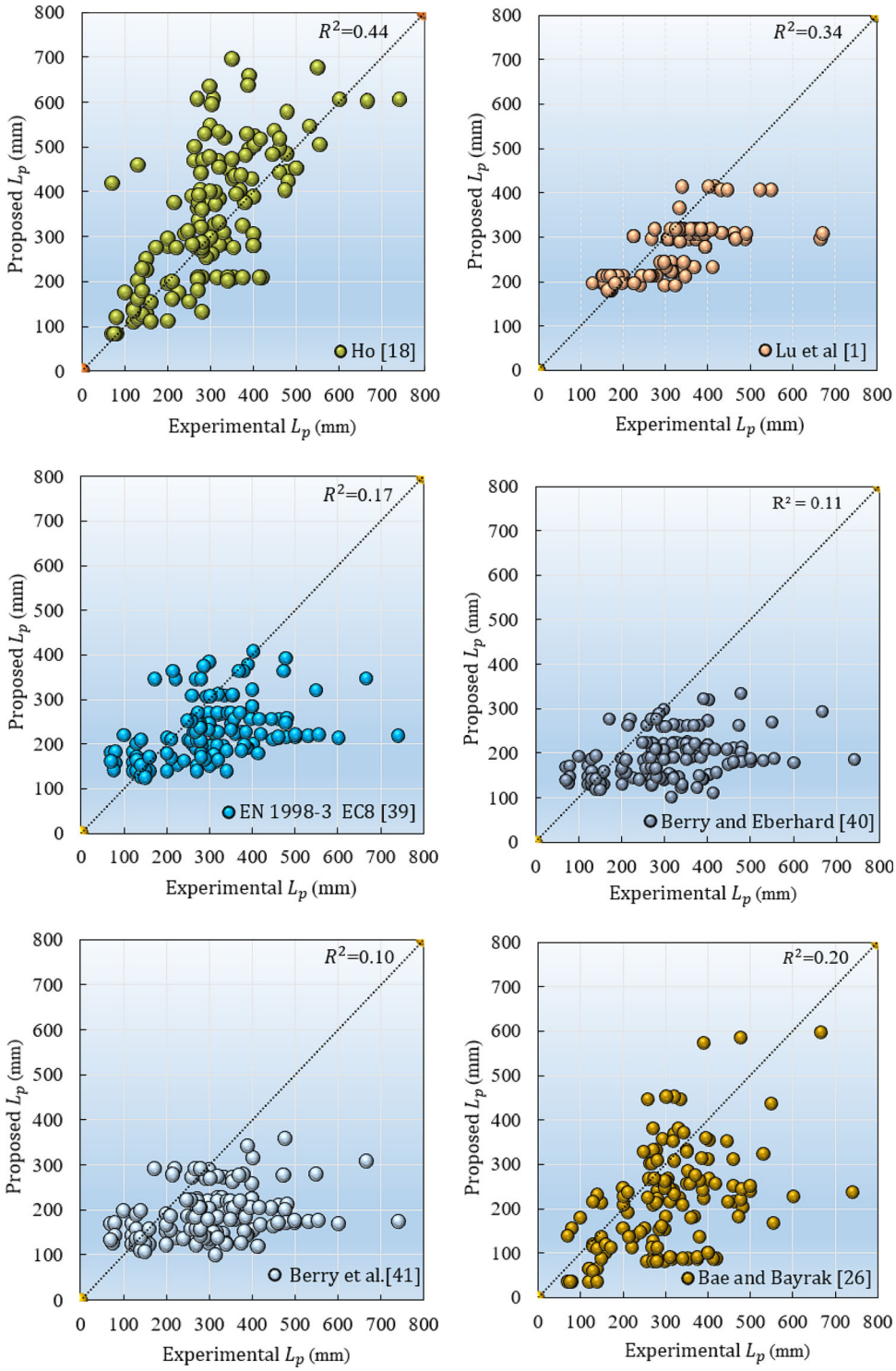


Fig. 7 Comparisons of the proposed models versus experimental L_p

P/P_o, and d11 is L/H, respectively. The GEP-based formulation of the L_p of RC columns in terms of B, H, L, f'_c, d_b, ρ_s, f_y, $s, \rho_v, f_{yv}, P/P_o, L/H$ is presented in Eq. 4. The model proposed in this study is valid for columns where the bending failure occurs, cross-sectional dimensions greater than 14 cm, and L/H ratio greater than 1.8.

Fig. 7 continued

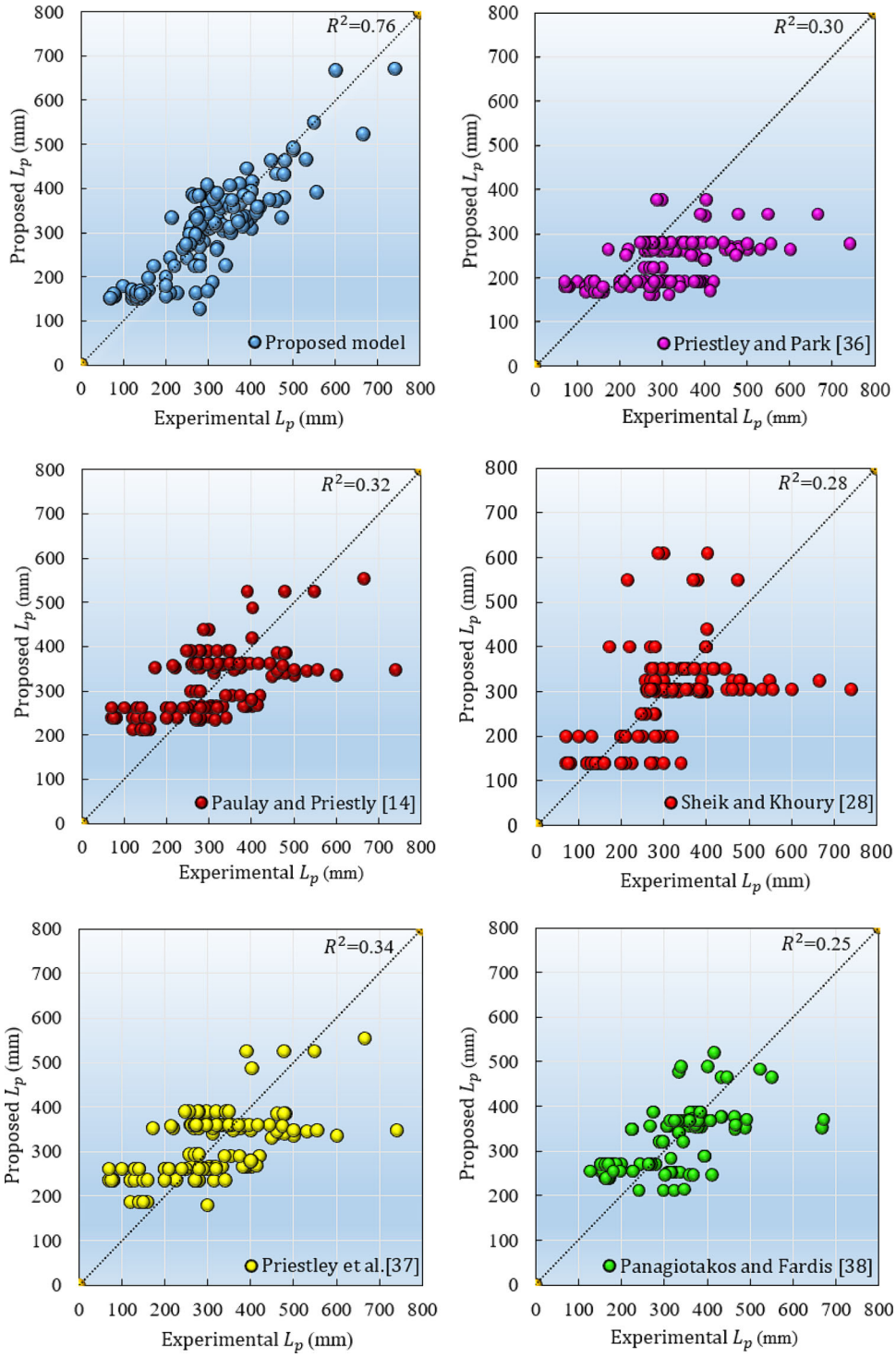
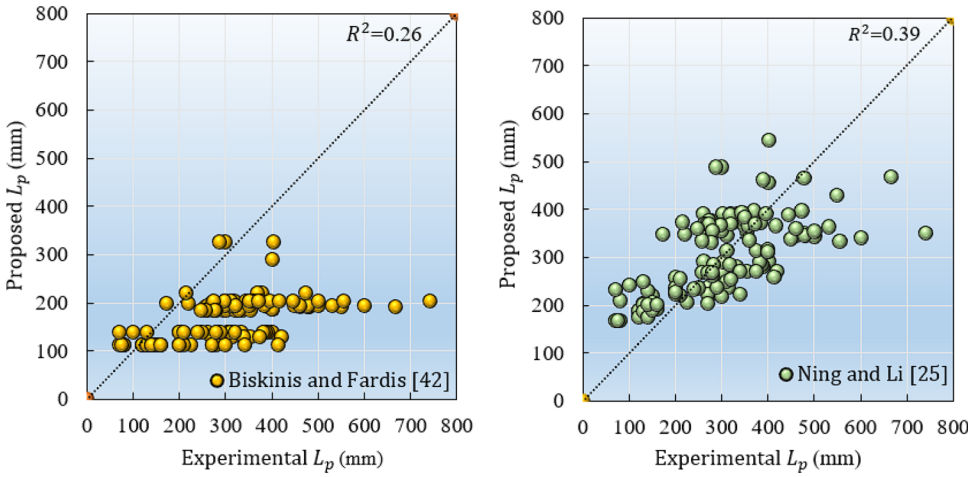


Fig. 7 continued



$$L_p = f'_c - d_b + \frac{(1083.37 - L)\sqrt{\rho_s . s}}{(H - f_{yv} - 0.645s)\tan f'_c} + \frac{L}{d_b} - \frac{\left(58.95 - \frac{P}{P_o}\right)}{\rho_v (B - f'_c)} - \left(\frac{f_y}{B} + 11.26\right)\left(\frac{L}{H} - \mathrm{d}_b\right) - \tan\left\{\frac{4.63}{\sqrt{H}}\left(\frac{P}{P_o} + f_{yv} + 49.51\right)\right\} - \tan(\rho_v - s) + 39.07 \quad (4)$$

4.2 The performance of the GEP-based formulation

In order to evaluate the prediction performance of the GEP model, the results of the experimental and GEP model for L_p were compared in Fig. 5. As seen in Fig. 5, the fluctuation in the GEP model and experimental results was compatible with each other. The R^2 values for the training and validation database were obtained as 0.76 and 0.78, respectively. Therefore, it can be said that the results of the GEP model were satisfactory and close to those of the experiments.

5 Performance evaluation of existing models

Table 4 shows the statistical criteria to evaluate the performance of existing models. These include mean (*M*), standard deviation (*SD*), the root mean square error (*RMSE*), mean absolute percentage error (*MAPE*), coefficient of variation (*COV*) and coefficient of determination (R^2). The description and mathematical formulation of each indicator are represented as follows:

$$M = \frac{\sum_{i=1}^{n} \frac{\text{model}}{\text{exp}}}{n} \quad (5)$$

$$\text{SD} = \sqrt{\frac{\sum_{i=1}^{n}\left(\frac{\text{model}}{\text{exp}} - M\right)^2}{n-1}} \quad (6)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\text{exp} - \text{model})^2}{n}} \quad (7)$$

$$\text{MAPE} = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{\text{exp} - \text{model}}{\text{exp}}\right| \quad (8)$$

$$\text{COV} = \frac{\text{SD}}{\text{M}} \quad (9)$$

$$R^2 = \frac{\left[\sum_{i=1}^{n}(\text{exp} - \overline{\text{exp}})(\text{model} - \overline{\text{model}})\right]^2}{\sum_{i=1}^{n}(\text{exp} - \overline{\text{exp}})^2 \sum_{i=1}^{n}(\text{model} - \overline{\text{model}})^2} \quad (10)$$

where n shows the number of specimens, exp and $\overline{\text{exp}}$ show the actual output and the mean of the real output values, respectively. The model and $\overline{\text{model}}$ also show the predicted output and the mean of the predicted outputs, respectively.

In Table 4 and Fig. 6, the results of the Eq. (4) derived by the GEP model were compared with those of existing models. Table 4 and Fig. 6 show that *MAPE*, *RMSE* and *COV* values of the GEP model are equal to 19.67, 60.17 and 0.27, respectively, and the R^2 value is equal to 0.76 for all data. As seen in, compared to other existing models, the GEP model has the lowest *MAPE*, *RMSE* and *COV* values, on the contrary, the highest R^2. This confirms that the developed GEP model can predict the L_p of RC column with good accuracy. Among the existing models, Ning and Li [25] model provided the least *RMSE* and *COV*, and Ho [18] model has the highest R^2, as seen in Table 4.

Thus, it can be said that Ning and Li [25] and Ho [18] models outperformed the other models in predicting the L_p. Also, Biskinis and Fardis [42] model have the highest

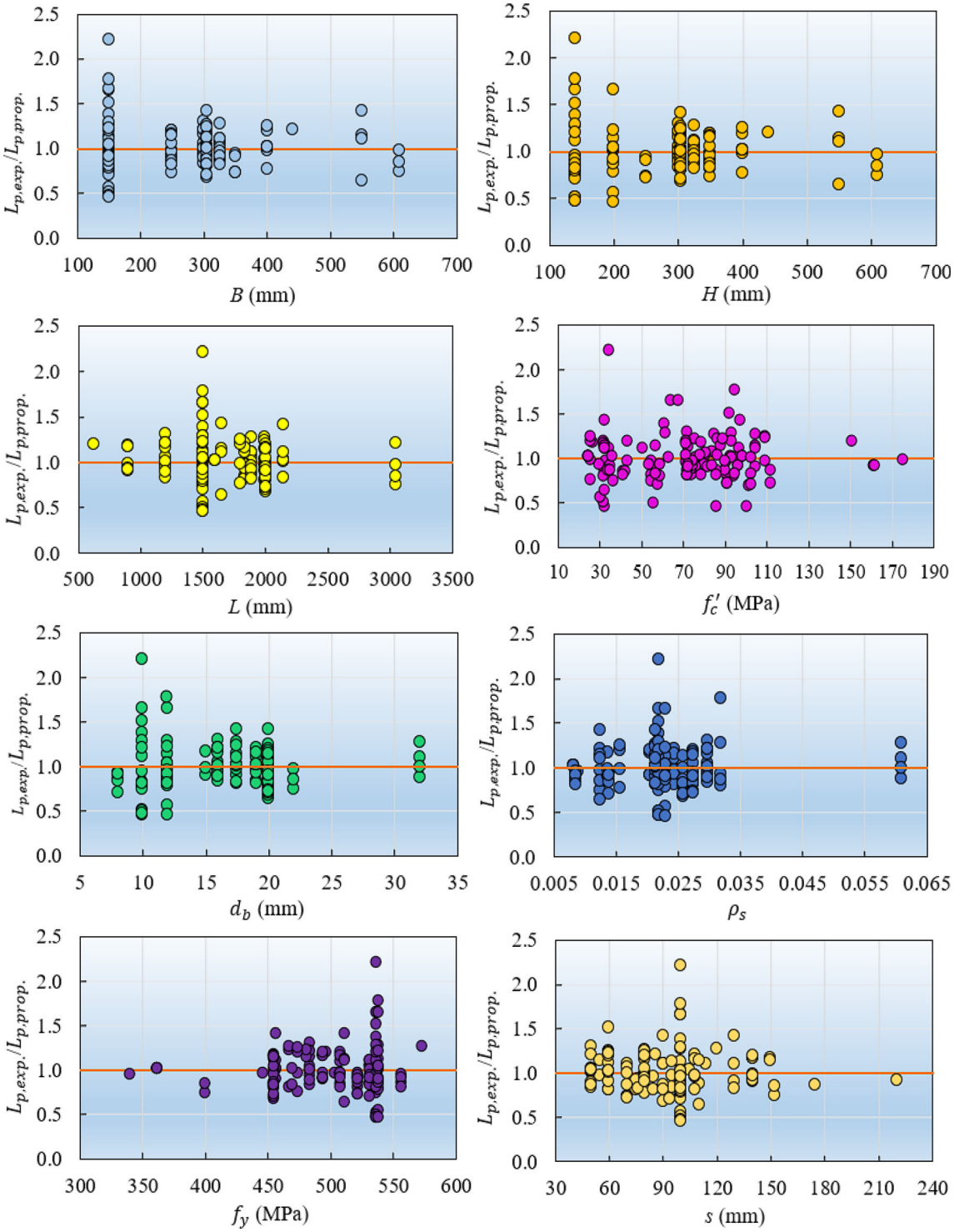


Fig. 8 Variation of the $(L_{p,\text{exp.}}/L_{p,\text{prop.}})$ with experimental input data

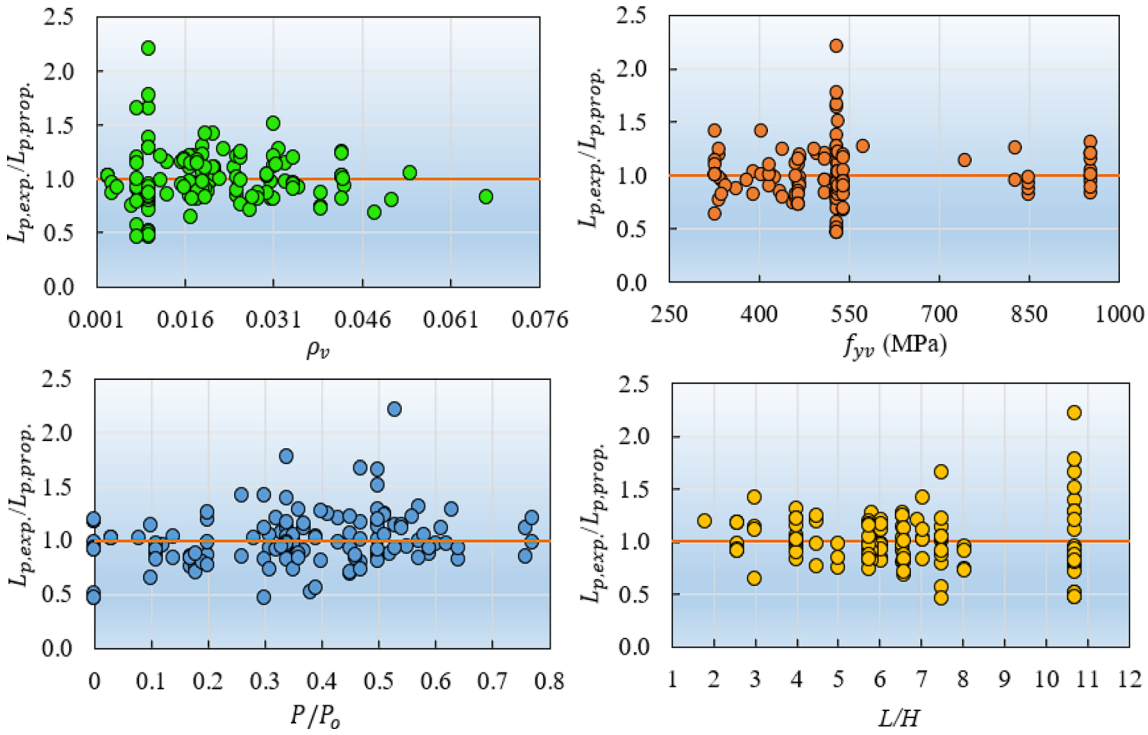


Fig. 8 continued

MAPE and *RMSE*. Berry et al. [41] model has the highest *COV* and the least R^2 as shown in Fig. 6 and Table 4. Thus, Table 4 and Fig. 6 show that the Biskinis and Fardis [42] and Berry et al. [41] models give the lowest performance compared to the existing models.

In addition, the comparisons of the proposed models versus experimental L_p were plotted in Fig. 7. As can be seen, there is a considerable discrepancy between the experimental data and the results of the models which are proposed by researchers. An evaluation was made by considering the R^2 value, which is one of the statistical criteria. The highest values of R^2 demonstrated that proposed and experimental L_p are close and have a good correlation. Among the existing models, Ho [18] model provided the highest R^2 of 0.44, followed by Ning and Li [25] model R^2 of 0.39 as shown in Fig. 7. Ho [18] showed the highest agreement with experiment results.

Conversely, as seen in Fig. 7, the Berry et al. [41] model shows the worst performance among all existing models with the value 0.1 of R^2. As shown in Fig. 7, the GEP model in this study demonstrates the highest accuracy between all existing models with the 0.76 value of R^2 for all database.

Besides, the performance of the proposed GEP model was assessed by plotting the ratio of the experimental input data to the $(L_{p,exp.}/L_{p,prop.})$ for each parameter (Fig. 8). The trend line reflects accurate and consistent predictions for all parameters.

6 Sensitivity and parametric analyses

A parametric analysis was also conducted in conjunction with sensitivity analysis. The predicted values of L_p obtained by GEP-based models as a function of each parameter to find the effect of each input parameter on the output parameter are presented in Fig. 9. The changes in L_p were noted against the change in the value of each input parameter from maximum to minimum, and the rest of all input parameters are kept at the mean value [48]. As shown in Fig. 9, the L_p constantly increase with the increase in L, f'_c, ρ_s and f_y, unlike decrease with increase in L/H and P/P_o. As B, H and ρ_v increase, the L_p results in an increment till a certain point, and then, it starts decreasing. Conversely, with the increasing of d_b, s and f_{yv}, L_p tends to decrease first and then increases.

A sensitivity analysis was carried out by Eqs. (11) and (12) to determine the relative contribution of the parameters on the L_p.

$$N_i = f_{\max}(x_j) - f_{\min}(x_j) \tag{11}$$

$$\mathrm{SA} = \frac{N_j}{\sum_n^{j=1} N_j} \tag{12}$$

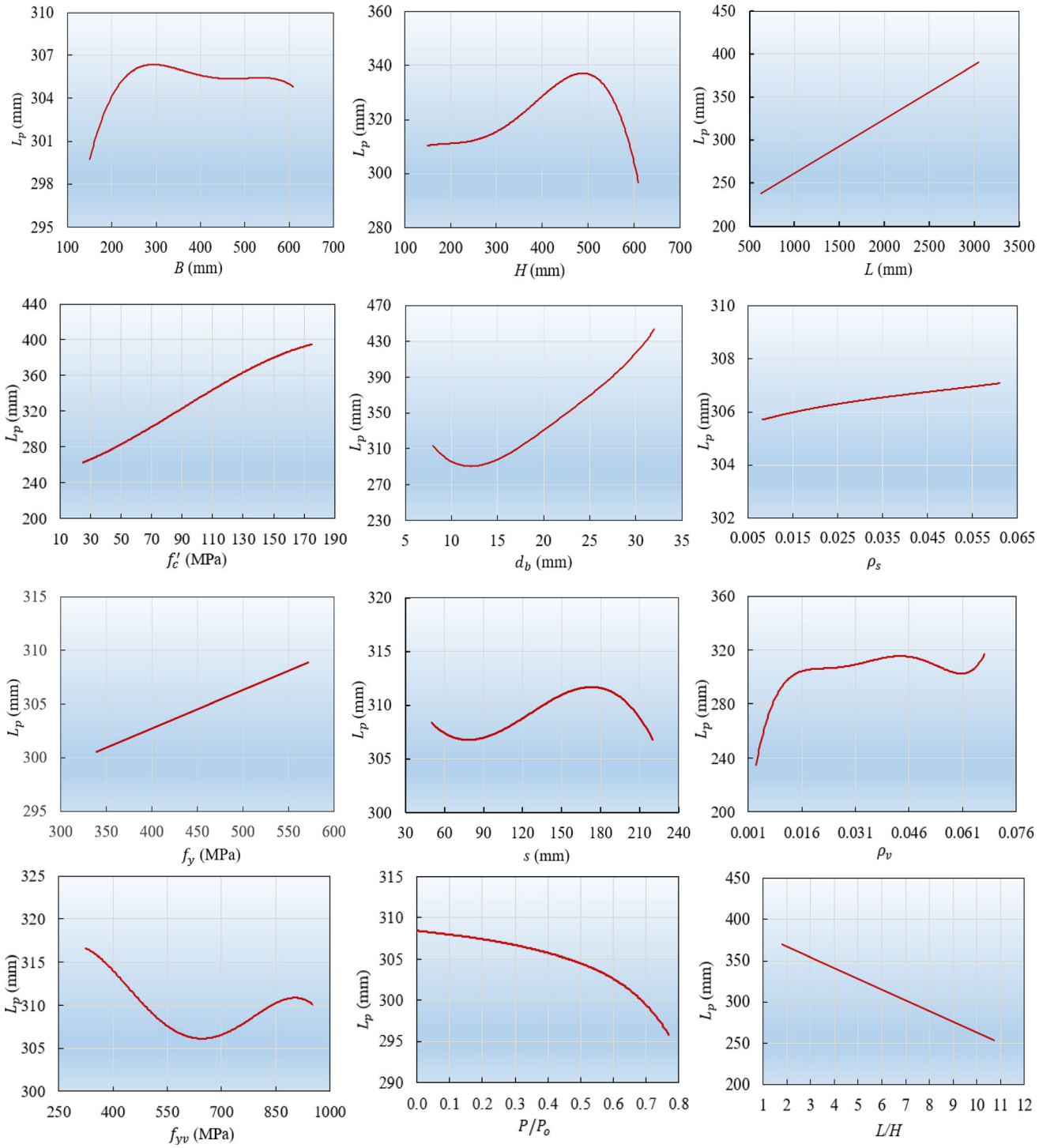


Fig. 9 Parametric analysis of inputs

where $f_{max}(x_j)$ and $f_{min}(x_j)$ represent the maximum and minimum predicted outputs according to the i th input domain, while other input parameters are kept constant their mean values. Sensitivity analysis is presented in Fig. 10. As can be seen in Fig. 10, the results of the sensitivity analysis revealed that f_c', d_b, L, L/H, ρ_v and H are the most important parameters on the L_p with a relative contribution of 23.05%, 18.56%, 18.52%, 14.2%, 10.43%, and 8.02%, respectively, while the parameter ρ_s shows the least contribution on the L_p, with the percentage of the impact of 0.17%. The relative contributions of the rest parameters are shown in Fig. 10.

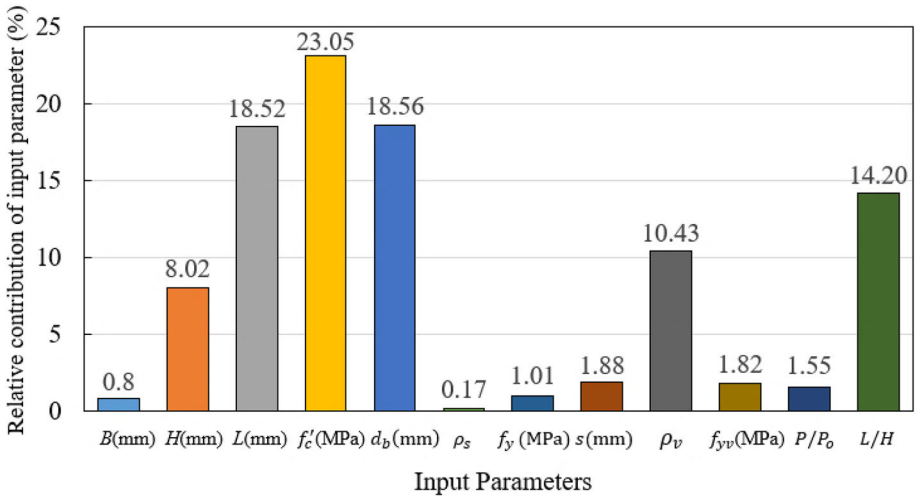


Fig. 10 Relative contribution of each input parameter on the proposed GEP model

The parameters s, $f_{yv}, P/P_o, f_y$ and B contribute about 1.88%, 1.82%, 1.55%, 1.01% and 0.8%, respectively.

7 Conclusions

Due to the large differences among existing equations in the literature, an equation for the plastic hinge length (L_p) of the RC columns using genetic expression programming (GEP) was proposed in this study. The following conclusions can be drawn from the present study:

When comparing the results of GEP model against the experimental results of the L_p, it has been concluded that R^2 values for the training and validation database are 0.76 and 0.78, respectively. This comparison demonstrates that the GEP model has sufficient sensitivity in predicting L_p.

Compared with the existing models in the literature, the GEP model exhibited *MAPE*, *RMSE*, and *COV* values of 19.67, 60.17, and 0.27, respectively. These results confirmed that the GEP model provided the best performance among existing models.

The results of the sensitivity and parameter analysis revealed that the compressive strength of concrete, diameter of longitudinal bars, shear span, aspect ratio, transverse reinforcement ratio and cross-sectional depth of the column are the most efficient parameters on the L_p, respectively, while the longitudinal reinforcement ratio is parameter with the least effect.

As a result, it can be said that GEP has more effective and reliable results compared to traditional regression analysis, so it can be used safely in the derivation of relations related to many problems in civil engineering in the future.

The GEP-based model proposed in this study and the obtained results are only valid for the database provided from the literature. In order to achieve more sensitive results, it is recommended to derive new equations for the L_p by increasing the number of experiments and input parameters.

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK).

Data availability statement The data used in this are available upon request to the corresponding author.

Declarations

Conflict of interest No potential conflict of interest was reported by the authors.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Lu Y, Gu X, Guan J (2005) Probabilistic drift limits and performance evaluation of reinforced concrete columns. J Struct Eng 131(6):966–978. https://doi.org/10.1061/(ASCE)0733-9445(2005)131:6(966)
2. Babazadeh A, Burgueño R, Silva PF (2016) Evaluation of the critical plastic region length in slender reinforced concrete bridge columns. Eng Struct 125:280–293. https://doi.org/10.1016/j.engstruct.2016.07.021

3. Ou Y-C, Kurniawan RA, Kurniawan DP, Nguyen ND (2012) Plastic hinge length of circular reinforced concrete columns. Comput Concr 10(6):663–681. https://doi.org/10.12989/cac.2012.10.6.663
4. Megalooikonomou KG, Tastani SP, Pantazopoulou SJ (2018) Effect of yield penetration on column plastic hinge length. Eng Struct 156:161–174. https://doi.org/10.1016/j.engstruct.2017.11.003
5. Pokhrel M, Bandelt MJ (2019) Plastic hinge behavior and rotation capacity in reinforced ductile concrete flexural members. Eng Struct 200:109699. https://doi.org/10.1016/j.engstruct.2019.109699
6. Elmenshawi A, Brown T, El-Metwally S (2012) Plastic hinge length considering shear reversal in reinforced concrete elements. J Earthquake Eng 16:188–210. https://doi.org/10.1080/13632469.2011.597485
7. Yuan F, Wu YF (2017) Effect of load cycling on plastic hinge length in RC columns. Eng Struct 147:90–102. https://doi.org/10.1016/j.engstruct.2017.05.046
8. Cansiz S, Aydemir C, Arslan G (2019) Comparison of displacement capacity of reinforced concrete columns with seismic codes. Adv Concr Constr 8(4):295–304. https://doi.org/10.12989/acc.2019.8.4.295
9. Baker ALL (1956) The ultimate load theory applied to the design of reinforced and prestressed concrete frames. Concrete Publications Ltd, London
10. Baker ALL, Amarakone AMN (1964) Inelastic hyperstatic frame analysis. Flexural mechanics of reinforced concrete, American Concrete Institute, pp 85–142. DOI: https://doi.org/10.14359/16715.
11. Corley WG (1966) Rotational capacity of reinforced concrete beams. J Struct Div ASCE 92(5):121–146
12. Mattock AH (1967) Discussion of rotational capacity of hinging regions in reinforced concrete beams. J Struct Div ASCE 93(2):519–522. https://doi.org/10.1061/JSDEAG.0001678
13. Verderame GM, Ricci P, Manfredi G, Cosenza E (2010) Ultimate chord rotation of RC columns with smooth bars: Some considerations about EC8 prescriptions. Bull Earthq Eng 8:1351–1373. https://doi.org/10.1007/s10518-010-9190-x
14. Paulay T, Priestley MJN (1992) Seismic design of reinforced concrete and masonry buildings. Wiley, New York. https://doi.org/10.1002/9780470172841
15. Tanaka H (1990) Effect of lateral confining reinforcement on the ductile behaviour of reinforced concrete columns. Ph.D. thesis, Dept. of civil engineering, Univ. of Canterbury, Christchurch, New Zealand
16. Ho JCM, Pam HJ (2003) Inelastic design of low-axially loaded high-strength reinforced concrete columns. Eng Struct 25(8):1083–1096. https://doi.org/10.1016/S0141-0296(03)00050-6
17. Légeron F, Paultre P (2000) Behavior of high-strength concrete columns under cyclic flexure and constant axial load. ACI Struct J 97(4):591–601. https://doi.org/10.14359/7425
18. Ho, J.: Inelastic design of reinforced concrete beams and limited ductile high-strength concrete columns. Ph.D. thesis. University of Hong Kong (2003)
19. Mortezaei A (2014) Plastic hinge length of rc columns under the combined effect of near-fault vertical and horizontal ground motions. Period Polytech Civil Eng 58(3):243–253. https://doi.org/10.3311/PPci.7329
20. Geng X, Zhou W (2018) Plastic hinge capacity prediction for existing RC rectangular columns to load reversals. KSCE J Civ Eng 22(10):3883–3893. https://doi.org/10.1007/s12205-018-0454-3
21. Barbagallo F, Bosco M, Floridia A, Marino EM, Panarelli D, Rossi PR, Spinell N (2022) Calibration of the length of the plastic hinge for numerical models of reinforced concrete members. Buildings 12(1603):20. https://doi.org/10.3390/buildings12101603
22. Wakjira TG, Alam MS, Ebead U (2021) Plastic hinge length of rectangular RC columns using ensemble machine learning model. Eng Struct 244:112808. https://doi.org/10.1016/j.engstruct.2021.112808
23. Feng D, Cetiner B, Kakavand MRA, Taciroglu E (2021) Data-driven approach to predict the plastic hinge length of reinforced concrete columns and its application. J Struct Eng 147(2):04020332. https://doi.org/10.1061/(ASCE)ST.1943-541X.0002852
24. Bayrak B, Akarsu O, Kaplan G, Aydın AC (2023) The plastic hinge length prediction of RC members by using ANN, Sådhanå 48:114. Indian Acad Sci. https://doi.org/10.1007/s12046-023-02182-4
25. Ning C-L, Li B (2016) Probabilistic approach for estimating plastic hinge length of reinforced concrete columns. J Struct Eng 142:04015164. https://doi.org/10.1061/(asce)st.1943-541x.0001436
26. Bae S, Bayrak O (2008) Plastic hinge length of reinforced concrete columns. ACI Struct J 105(3):290–300. https://doi.org/10.12989/cac.2012.10.6.663
27. Pam HJ, Ho JCM (2019) Length of critical region for confinement steel in limited ductility high-strength reinforced concrete columns. Eng Struct 31(12):2896–2908. https://doi.org/10.1016/j.engstruct.2009.07.015
28. Sheikh SA, Khoury SS (1993) Confined concrete columns with stubs. ACI Struct J 90:414–431. https://doi.org/10.14359/3960
29. Sheikh SA, Shah DV, Khoury SS (1994) Confinement of high strength concrete columns. ACI Struct J 91(1):100–111. https://doi.org/10.14359/4483
30. Yang K, Shi QX, Zhao JH (2013) Plastic hinge length of high strength concrete columns confined by high strength stirrups. Eng. Mech. 30(2):254–259. https://doi.org/10.6052/j.issn.1000-4750.2011.08.0524
31. Paultre P, Legeron F, Mongeau D (2001) Influence of concrete strength and transverse reinforcement yield strength on behavior of high-strength concrete columns. ACI Struct J 98(4):490–501. https://doi.org/10.14359/10292
32. Barrera AC, Bonet JL, Romero ML, Miguel PF (2011) Experimental tests of slender reinforced concrete columns under combined axial load and lateral force. Eng Struct 33(12):3676–3689. https://doi.org/10.1016/j.engstruct.2011.08.003
33. Bayrak O, Sheikh SA: Seismic performance of high strength concrete columns confined with high strength steel. Proc, 13th World Conf on Earthquake Engineering, Vancouver, BC, Canada (2004)
34. Ohno T, Nishioka T (1984) An experimental study on energy absorption capacity of columns in reinforced concrete structures. Proc JSCE, Struct Eng/Earthquake Eng 1(1):137–147
35. Bayrak, O.: Seismic performance of rectilinearly confined high strength concrete columns, Ph.D. thesis, Univ. of Toronto, Toronto (1998)
36. Priestley MJN, Park R (1987) Strength and ductility of RC bridge columns under seismic loading. Struct J ACI 84:285–336. https://doi.org/10.14359/2800
37. Priestley MJN, Seible F, Calvi G (1996) Seismic design and retrofit of bridges. Wiley, New York. https://doi.org/10.1002/9780470172858
38. Panagiotakos TB, Fardis MN (2001) Deformations of reinforced concrete members at yielding and ultimate. ACI Struct J 98:135–48. https://doi.org/10.14359/10181
39. EN 1998–3:2005.: Eurocode 8: design of structures for earthquake resistance-Part 3: assessment and retrofitting of buildings, vol. 1 (2005)

40. Berry MP, Eberhard MO (2007) Performance modelling strategies for modern reinforced concrete bridge columns. PEER report 2007/07, Berkeley, USA: pacific earthquake engineering research center, College of Engineering, University of California
41. Berry MP, Lehman DE, Lowes LN (2008) Lumped-plasticity models for performance simulation of bridge columns. ACI Struct J 105(3):270–279. https://doi.org/10.14359/19786
42. Biskinis D, Fardis MN (2010) Flexure-controlled ultimate deformations of members with continuous or lap-spliced bars. Struct Concr 11(2):93–108. https://doi.org/10.1680/stco.2010.11.2.93
43. Ferreira C (2001) Gene expression programming: a new adaptive algorithm for solving problems. Compl Syst 13(2):87–129
44. Ferreira C (2002) Gene expression programming in problem solving. Springer, London, pp 635–653
45. Gandomi AH, Alavi AH, Ryan C (2015) Handbook of genetic programming applications. Springer, New Jersey, p 450
46. Ferreira C (2006) Gene expression programming: mathematical modeling by an artificial intelligence. Second, revised and extended. Springer, London, p 478
47. Gepsoft GeneXproTools 5.: Data modeling & Analysis software, (2023). https://www.gepsoft.com
48. Gandomi AH, Alavi AH, Mirzahosseini MR, Nejad FM (2011) Nonlinear genetic-based models for prediction of flow number of asphalt mixtures. J Mater Civ Eng 23(3):248–263. https://doi.org/10.1061/(ASCE)MT.1943-5533.0000154

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