Engineering Failure Analysis 169 (2025) 109179

Contents lists available at ScienceDirect

Engineering Failure Analysis

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

Plastic hinge length of RC shear walls: Practical approximation via machine learning and probabilistic assessment

Abouzar Jafari [a], Masoomeh Mirrashid [b,*], Ryan D. Hoult [c], Ying Zhou [a]

[a] State Key Laboratory of Disaster Reduction in Civil Engineering, Tongji University, 1239 Siping Rd, Shanghai 200092, China
[b] College of Engineering, Abu Dhabi University, United Arab Emirates
[c] Institute of Mechanics, Materials and Civil Engineering, Université Catholique de Louvain, Lou-vain-la-Neuve 1348, Belgium

A R T I C L E I N F O

Keywords:
Plastic hinge length
RC shear walls
Machine learning models
Seismic failure
Monte Carlo simulations
Reliability analysis

A B S T R A C T

This study presents practical solutions for estimating the plastic hinge length of RC shear walls using machine learning (ML)-based approaches and incorporates probabilistic assessment to investigate the robustness of the developed solutions. A comprehensive dataset, comprising 234 experimental tests and 487 numerical simulation cases, was used to develop and refine the models, accounting for diverse wall configurations and material properties. Feature selection identified key influencing factors such as wall dimensions, reinforcement ratios, and material strengths. Two solutions were developed: a simpler, more straightforward model and a more complex model with enhanced performance. Both models outperform existing empirical methods in terms of accuracy and reliability. The SHapley Additive exPlanations (SHAP) analysis reveals that wall length and shear span are the most influential features. Using the Monte Carlo Simulation (MSC) method, the reliability of the existing and proposed solutions was analyzed to evaluate the robustness of their predictions. The results demonstrated that the predictions of the developed closed-form solutions are more reliable and robust than those of existing solutions in estimating plastic hinge length. While the second solution offers higher accuracy, the first one provides a simpler and more practical solution without significant compromise in performance.

1. Introduction

Recent research underscores the importance of robust structural design in mitigating seismic risks [1]. However, while structural wall systems play a vital role in enhancing building performance, poorly designed shear walls or the use of partial shear walls can lead to torsional irregularities that complicate seismic performance [2]. The term “shear wall,” historically rooted in force-based design approaches, refers to walls resisting lateral loads (e.g., wind, earthquakes), though many primarily rely on flexural rather than shear actions for their performance. Insights from the Van 2011 earthquake in Turkey emphasize that, although shear walls are crucial for lateral stability, adherence to modern seismic codes is essential for overall safety. A better understanding of their response and mechanics can further mitigate seismic failure [3]. Furthermore, the influence of shear span and reinforcement ratios on shear wall behavior points to critical design considerations [4] and highlights the need for a more comprehensive approach that accounts for as many influential parameters as possible in shear wall design. These findings inform our study’s objective to develop solutions for

* Corresponding author.
E-mail addresses: a_jafari@tongji.edu.cn (A. Jafari), mirrashid.research@adu.ac.ae (M. Mirrashid), ryan.hoult@uclouvain.be (R.D. Hoult), yingzhou@tongji.edu.cn (Y. Zhou).

https://doi.org/10.1016/j.engfailanal.2024.109179
Received 29 October 2024; Received in revised form 26 November 2024; Accepted 4 December 2024
Available online 6 December 2024

estimating plastic hinge lengths in RC shear walls by considering influential features and incorporating epistemic uncertainties to improve seismic failure mitigation.

Since Park and Paulay [5] introduced the plastic hinge concept, it has been extensively utilized by engineers to estimate the inelastic displacement capacity of reinforced concrete (RC) structural elements. Due to the simplicity and applicability of the plastic hinge concept, substantial research has been conducted to develop a comprehensive relationship capable of accurately predicting an equivalent height/length of the zone that encompasses the majority of plastic deformation in an RC structural element, known as the 'plastic hinge length'. Given that structural walls often act as relatively isolated elements within a building, the plastic hinge concept becomes more practical and appealing for estimating the plastic response of structural walls.

The initial approximation of the plastic hinge height of structural walls was proposed based on experimental results conducted on RC beams, recommending it be half the length of the wall, l_w, as a safe and conservative estimation [6–8]. The proposed plastic hinge height has been incorporated into many codes of practice [9–12], likely due to its simplicity. However, several studies have shown that this approximation is unreasonable [13] or overestimates the plastic hinge length, leading to an overestimation of the ductility and displacement of shear walls [14–16]. Some studies [13,14,17,18] have shown the significant effect of shear stress demand on the plastic hinge length, especially in non-planar shear walls [17,18], where the shear stress is much higher compared to planar walls. A careful literature survey reveals multiple expressions for calculating the plastic hinge length of planar and non-planar shear walls, whether they are lightly or heavily reinforced. The majority of available equations were derived through numerical and experimental investigations to estimate the plastic hinge length of planar shear walls that are moderately or heavily reinforced [6–8,10,12,13,19–21]. A few studies aimed at calculating the plastic hinge length of lightly reinforced shear walls (mostly with planar sections) [15,22,23], while others discussed the plastic hinge length of non-planar shear walls [17,18] and coupled shear walls [24]. In addition to the studies mentioned, numerous other research efforts have been conducted to estimate the plastic hinge length of shear walls, some of which were reviewed earlier by Bohl [14] and Hoult [25].

Although the plastic hinge length of shear walls has been extensively researched, the applicability of these results is mostly limited to the range of studied variables and defined scenarios. Furthermore, most available solutions have focused on planar shear walls, even though RC shear walls are often used as non-planar cores in buildings [26]. There is, therefore, a pressing need to derive a practical, accurate, and universal solution for the plastic hinge length of RC shear walls, whether for assessment or design purposes. Addressing this need through data mining has become more feasible, thanks to the extensive experimental data and recent numerical studies available [13,15,17,19]. In the most recent study by one of the co-authors [27], the available experimental and numerical data were collected, and using nonlinear regression, a simplified and universal relation was derived for estimating the plastic hinge length of shear walls.

Given recent advancements in data science, ML approaches or algorithms are now recognized for their robustness and superior performance compared to classic statistical analysis methods in pattern recognition tasks. Therefore, employing these advanced techniques for analyzing the collected experimental and numerical data and estimating plastic hinge length can result in more accurate, robust, and universally applicable relations. To the best of the authors' knowledge, the only ML prediction model for the plastic hinge length of shear walls was developed by Barkhordari and Jawdhari [28] using the data set provided by Hoult [27]. Despite the reported high accuracy of the proposed ML models, which is also accessible via a web application interface, a closed-form solution or an explainable ML model was not provided, highlighting the need for further research to develop more practical and straightforward solutions.

A literature review shows that the existing relationships used to predict plastic hinge length in shear walls are primarily developed for walls with planar configurations and moderate to heavy reinforcement. Additionally, although the relationships in use are practical and robust, they lack acceptable accuracy in predicting the plastic hinge length and may overestimate it. The significance of the forthcoming study lies in its potential to provide accurate and robust predictions of plastic hinge length by considering all influential parameters across the majority of wall configurations. The study first aims to scrutinize the influence of various parameters on plastic hinge length using collected experimental and numerical data, ensuring a comprehensive examination of the effective factors. Subsequently, closed-form solutions are developed using ML algorithms that not only promise superior predictive capabilities but also offer robustness and interpretability. The estimated plastic hinge lengths from the developed ML-based solutions and the frequently referenced or used empirical relationships are then compared to the experimentally and numerically estimated plastic hinge lengths to evaluate the efficiency and capability of the developed models. Finally, incorporating a probabilistic analysis into the developed solutions and empirical solutions adds a layer of robustness by assessing their performance in stochastic scenarios, thereby demonstrating the reliability of the proposed ML-based solutions compared to existing solutions. Given the ongoing threat of seismic events, this research offers a valuable tool for engineers and researchers to enhance the seismic resilience of RC shear walls.

2. Materials and methods

This section consists of the following subsections: a brief introduction of the plastic hinge concept, a summary of past studies on plastic hinge length, an explanation of the collected dataset and the selected machine learning approach, a description of the feature selection method, an evaluation of the performance of both the developed and existing solutions and a probabilistic assessment of these solutions.

2.1. Plastic hinge concept

A slender RC shear wall subjected to a lateral load F at the top is illustrated in Fig. 1a. The total lateral displacement of the shear wall comprises flexural, shear, and slide displacements. However, both shear and slide displacements are negligible for flexure-governed walls (i.e., high-rise walls or those with a large aspect ratio). As shown in Fig. 1a, the total displacement of the wall after yielding consists of two elements: yield displacement and plastic deformation. The moment distribution over the height of the wall due to the applied lateral load is depicted in Fig. 1b.

The schematic moment–curvature diagram of the shear wall is displayed in Fig. 1c. Using the moment–curvature of the wall, the wall displacement can be estimated. As shown in Fig. 1c, three limit states—cracking, yielding, and crushing—are demonstrated on the moment–curvature diagram. The wall response is elastic and linear up to the first limit state (cracking; M_{cr}, ϕ_{cr}) indicating the initiation of concrete cracking. When the applied force is increased, the flexural rigidity of the wall decreases due to crack propagation in the section. Subsequently, the steel reinforcement on the tension side begin to yield, indicating the second limit state (yielding), defined by M_y and ϕ_y. Further increasing the lateral load leads to the beginning of concrete crushing in the section, represented by M_u and ϕ_u, the crushing limit state (refer to Fig. 1c). This example assumes the wall is well-designed, with failure governed by compression rather than tension (and subsequent steel bar rupture). The curvature distribution over the height of the wall at the crushing limit state is demonstrated in Fig. 1d. The total displacement of the wall can be calculated from the curvature profile as follows:

$$\Delta_y + \Delta_p = \int_0^H \phi(y)y.dy \tag{1}$$

As depicted in Fig. 1d, the curvature distribution along the height of the wall can be idealized into elastic and inelastic curvature distributions. By utilizing these idealized distributions, Eq. (1) can be rewritten as follows:

$$\Delta_y = \frac{\phi_y H^2}{3} \tag{2}$$

As shown in Fig. 1d, the inelastic curvature, ϕ_p, can be estimated using the following equation:

$$\phi_p = \phi - \phi_y \tag{3}$$

Using the plastic hinge concept, the strain and curvature, over a region at the base of the wall with the height of L_p, were assumed constant and equal to their maximum values at the wall base, the inelastic rotation, θ_p, can be calculated using the following equation:

$$\theta_p = \phi_p L_p \tag{4}$$

Looking at Fig. 1e, by concentrating the inelastic rotation at the centroid of the plastic hinge, the inelastic displacement, Δ_p, can be estimated as:

$$\Delta_p = \theta_p \left(H - \frac{L_p}{2}\right) = (\phi - \phi_y) L_p \left(H - \frac{L_p}{2}\right) \tag{5}$$

As depicted in Fig. 1d, below the wall base, the curvature immediately drops to zero. This assumption leads to underestimating the total wall displacement due to strain penetration (anchorage deformation). Therefore, a length L_{SP} is defined, over which the curvature remains constant and is equal to the curvature at the wall base.

2.2. Empirical solutions for plastic hinge length

A summary of more recent, frequently referenced solutions, or those used in guidelines and codes of practice, was selected and

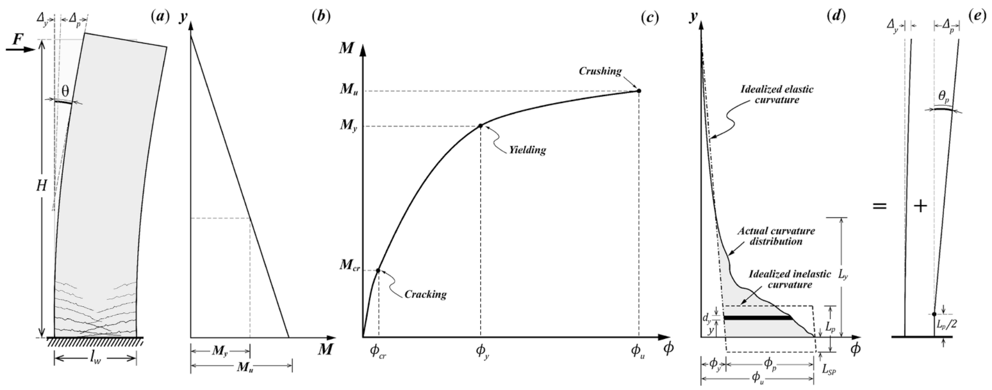

Fig. 1. A schematic of a shear wall subjected to a load F. (a) the wall under the load F, (b) the moment diagram of the wall, (c) the moment–curvature diagram of the wall, (d) the actual and idealized curvature distributions over the wall height, including plastic hinge length and stress-penetration length, and (e) classical plastic hinge formulation with a zero-length hinge.

listed in Table 1. The following relationships will later be used to compare the accuracy and reliability of the solutions developed in this study.

In Table 1, l_w, H and A_g represent the length, height and the gross cross-section of the wall, respectively. f_y and f_u denote the yield and ultimate strengths of the steel reinforcements. H_e signifies the wall shear span, calculated as the ratio of moment to shear. f_{ye} represents the expected yield strength of longitudinal reinforcements, set at $1.1f_y$. d_{bl} refers to the diameter of the longitudinal reinforcements. f_c' and V_{max} stand for the compressive strength of concrete and the maximum lateral capacity of the wall, respectively. d indicates the effective flexural depth of the wall. ALR denotes the axial load ratio of the wall, which is equal to $P_0/(A_g f_c')$, where P_0 represents the applied axial load. ρ_h, ρ_{wv} represent the horizontal and longitudinal reinforcement ratios, respectively. $\rho_{wv,min}$ presents the minimum longitudinal reinforcement ratio, which can be calculated using the following expiration:

$$\rho_{wv,min} = \frac{(t_w - n_t d_{bt}) f_t}{t_w f_u} \tag{22}$$

Where t_w represents the thickness of the wall, n_t and d_{bt} denote the number of layers and diameter of the transverse reinforcement, respectively. f_t is the flexural tensile strength of concrete ($\approx 0.6\sqrt{f_c'}$). The plastic hinge length solutions listed in Table 1 are later utilized to assess the accuracy and reliability of the derived closed-form approximation in this research.

2.3. Database

The authors previously collected the dataset used in this research to derive solutions for the plastic hinge length of walls using statistical regression methods [27]. However, a brief explanation of the dataset is provided here. The dataset comprises numerical and experimental results on RC walls exhibiting flexure-controlled responses. It includes 234 experimental results from full-scale and scaled RC walls, along with 487 numerical results compiled from various sources. Specifically, the compiled numerical data consist of 240 rectangular walls simulated by Kazaz [9], 17 rectangular walls simulated by Bohl and Adebar [15], 197 rectangular walls modeled by Hoult *et al.* [11], and 33 U-shaped walls simulated by Hoult *et al.* [13]. It is noteworthy that the collected data also includes experimentally tested non-planar walls, such as barbell-shaped, I-shaped, L-shaped, H-shaped, T-shaped, and U-shaped walls.

All influential variables affecting the response of a shear wall, including the geometry of the wall, mechanical properties of concrete and steel reinforcements, longitudinal and transverse reinforcement ratios, size of longitudinal reinforcements, axial load ratio (ALR), wall shear span, and peak lateral capacity, are considered input parameters. These parameters are later ranked and shortlisted to eliminate less important features, with the wall's plastic hinge length being considered the output. The characteristic summary of selected inputs and output variables is listed in Table 2. The distribution of the inputs and the output (L_p) are illustrated in Fig. 2.

To provide insight into the inputs (independent variables) and their influence on the output (L_p), Fig. 3 demonstrates the variation of the output parameter against each input parameter.

To address the issues with low learning rates of ML algorithms at the upper and lower bounds of the database and to reduce the intervals of the variables, all input and output parameters are normalized to the range [-1, 1] using the following equation:

Table 1
Summary of past studies on plastic hinge length solutions for shear walls.

No	Researchers	Proposed solutions	Eq.
1	Paulay and Uzumeri [29]	$L_p = 0.2l_w + 0.075H$	(6)
2	Oesterle *et al.* [30]	$L_p = l_w$	(7)
3	Paulay [6]	$0.5l_w \le L_p \le l_w$	(8)
4	Wallace and Moehle [8]	$0.5l_w \le L_p \le l_w$	(9)
5	Paulay and Priestley [31]	$L_p = \left(0.2\left(f_u/f_y - 1\right) \le 0.08\right)H_e + 0.2l_w + 0.022f_{ye}d_{bl}$	(10)
6	Paulay and Priestley [32]	$L_p = 0.2l_w + 0.07H_e$	(11)
7	Sasani and Kiureghian [33]	$L_p = 0.43d + \frac{0.077\sqrt{H_e}}{d}$	(12)
8	Thomsen and Wallace [21]	$0.33l_w \le L_p \le 0.5l_w$	(13)
9	Eurocode 8 [34]	$L_p = H_e/30 + 0.2l_w + 0.11\left(d_{bl}f_y/\sqrt{f_c'}\right)$	(14)
10	Bohl and Adebar [19]	$L_p = (0.2d + 0.05H_e)(1 - 1.5ALR) \le 0.8l_w$	(15)
11	Wibowo *et al.* [22]	$L_p = 15d_{bl}$	(16)
12	Kazaz [13]	$L_p = 0.27l_w(1 - ALR)\left(1 - \frac{f_y\rho_h}{f_c'}\right)\left(\frac{H_e}{l_w}\right)^{0.45}$	(17)
13	Altheeb [23]	$L_p = \frac{\left(f_u - f_y\right)d_{bl}^{1.2}}{4\sqrt{f_c'}}$	(18)
14	Hoult *et al.* [15]	$L_p = \begin{cases} (0.1l_w + 0.075H_e)(1 - 6ALR) \le 0.5l_w & \rho_{wv} \ge \rho_{wv,min} \\ 100mm & \rho_{wv} < \rho_{wv,min} \end{cases}$	(19)
15	Hoult [27]	$L_p = \left(\frac{\rho_{wv}}{\rho_{wv,min}} \le 1.0\right)(0.13l_w + 1.8H_e^{0.6})(1 - 0.3ALR)\left(1 - 0.13\ln\left(\frac{V_{max}}{A_g\sqrt{f_c'}}\right)\right)$	(20)
16	Hoult [27]	$3t_w \le L_p = (0.1l_w + 0.02H_e) \le 0.8l_w$	(21)

Table 2
Characteristics of the inputs and output variables in the collected dataset.

Parameter			Unit	$\mu^{\dagger}$	σ	*Min*	*Max*
Input parameters	l_w	The length of the wall	mm	4263.25	2717.32	507.00	9000.00
	t_w	The thickness of the wall	mm	200.81	59.74	50.00	250.00
	ρ_{wv}	The longitudinal reinforcement ratio	%	0.0141	0.0114	0.0011	0.0450
	ρ_h	The transverse reinforcement ratio	%	0.0045	0.0040	0.0015	0.0314
	ALR	Axial load ratio	–	0.090	0.076	0.000	0.430
	f_c'	The concrete compressive strength	MPa	33.09	12.58	13.50	109.10
	f_t	The tensile strength of concrete	MPa	2.89	1.055	1.50	6.35
	f_y	The yield strength of longitudinal reinforcements	MPa	477.96	85.77	260.68	900.00
	f_u	The ultimate strength of longitudinal reinforcements	MPa	620.09	79.50	312.82	1036.45
	H_e	The wall shear span	mm	11317.09	9976.87	500.00	44100.00
	H	The height of the wall	mm	15694.92	14879.27	500.00	63000.00
	V_{max}	The peak applied force	kN	1736.33	2453.41	37.69	16450.00
	d_{bl}	The diameter of the longitudinal reinforcements	mm	16.00	4.16	3.00	43.00
Output parameter	L_p	The plastic hinge length	mm	1150.83	845.10	21.60	4889.53

† μ: mean; σ: standard deviation; *Min*: minimum and *Max*: maximum.

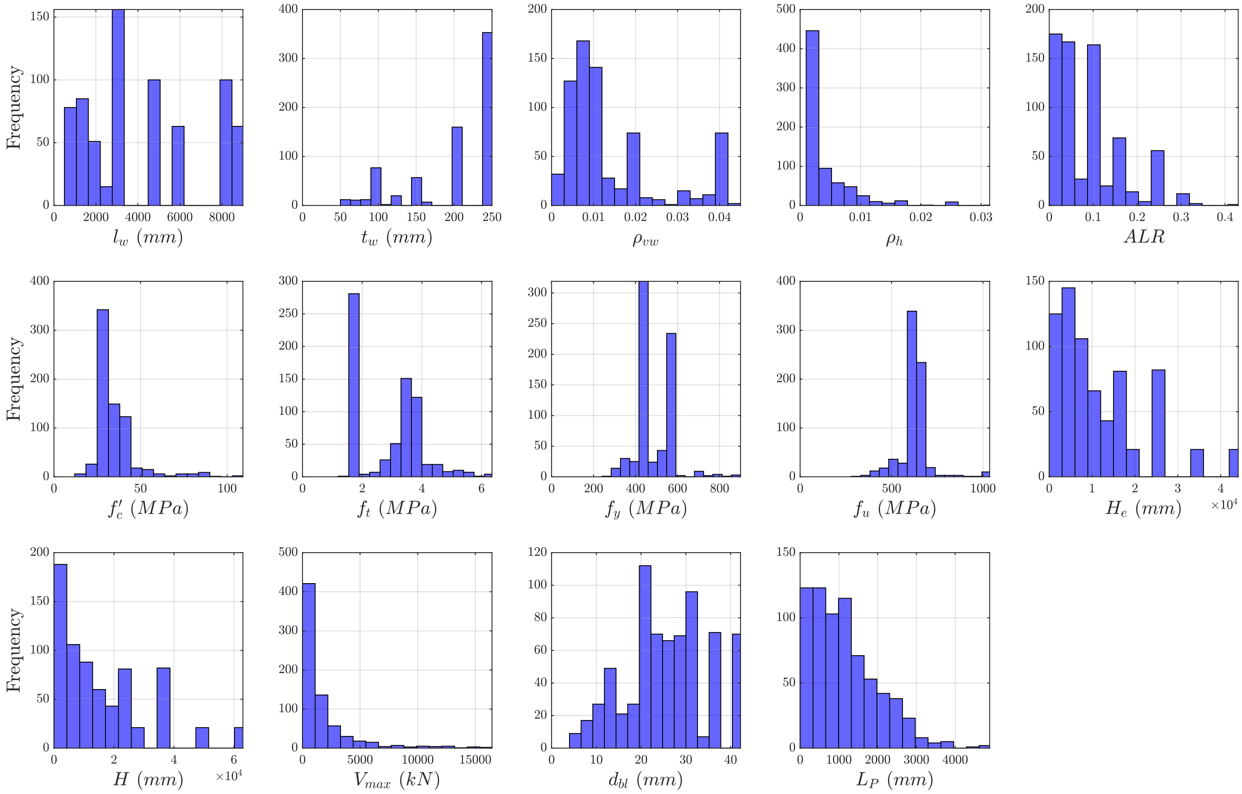

Fig. 2. Frequency distributions of the input parameters and the output.

$$\tilde{x} = 2\frac{x_{orig} - x_{min}}{x_{max} - x_{min}} - 1 \tag{23}$$

Where $\tilde{x}$ and x_{orig} represent the normalized and original values of a feature, respectively, and x_{max} and x_{min} denote the maximum and minimum values of the feature, respectively.

2.4. Machine learning methods

Machine learning is revolutionizing civil engineering by analyzing extensive databases. These algorithms optimize designs, predict structural behaviors, and enhance decision-making [35]. Nowadays, neural networks (NNs), a popular and well-developed machine

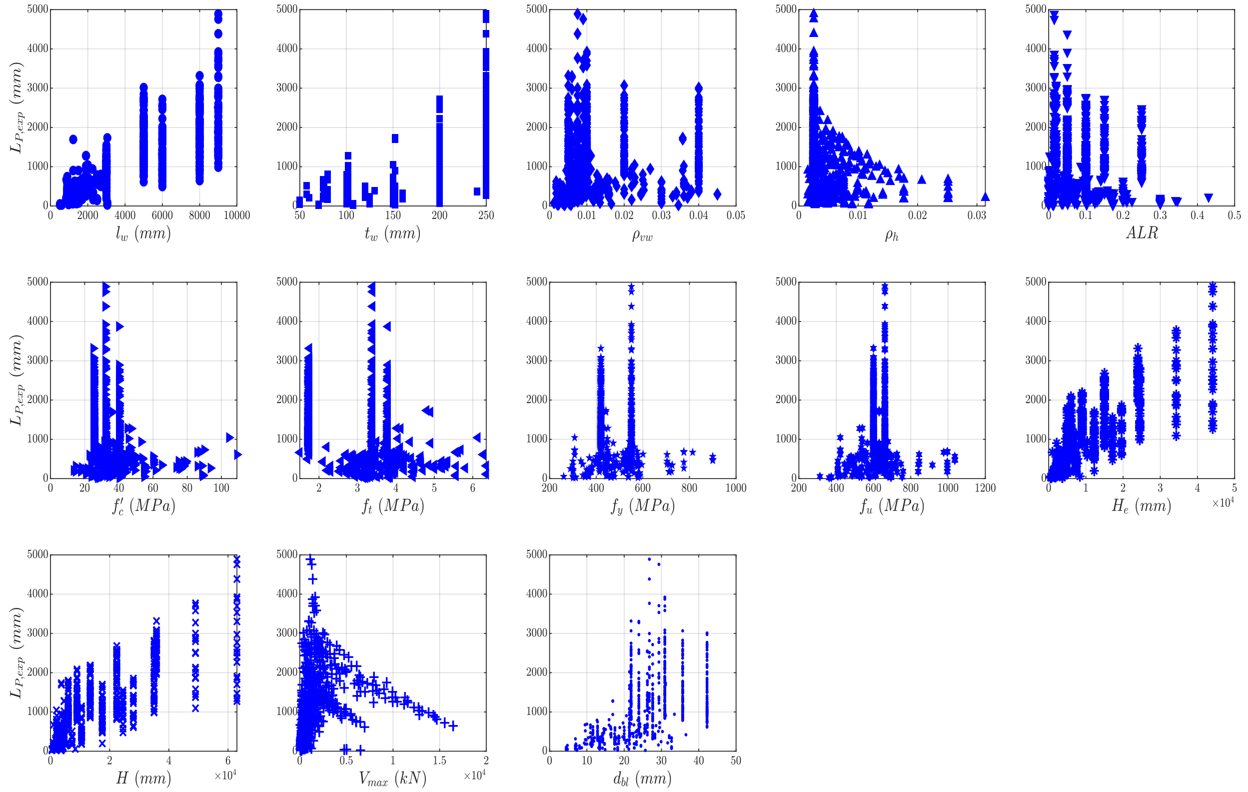

Fig. 3. Variation of input parameters versus the output parameter (L_p).

learning (ML) algorithm, are widely used to monitor structures for damage and predict or schedule maintenance needs using real-time data. By handling complex data relationships, NNs improve efficiency and sustainability in engineering practices, making infrastructure more resilient for the future [36].

2.4.1. Neural networks

This study employed a widely-used feed-forward NN framework known as the multilayer perceptron (MLP) for data training using error back-propagation to develop a closed-form approximation of the plastic hinge length for RC shear walls. An MLP framework comprises neurons structured across three main components: (i) an input layer that receives input parameters, (ii) one or more hidden layers that process information, and (iii) an output layer that generates predicted results. Neurons within this framework are interconnected, with each connection assigned a weight and every neuron incorporating a bias and an activation function.

Given an input vector $\mathbf{x} = [x_1, x_2, \cdots, x_n]$ for a neuron, the weighted sum, $z \in \mathbb{R}$, can be described as below:

$$z = \sum_{i=1}^{d} w_i x_i = \mathbf{w}^{\mathrm{T}}\mathbf{x} + b \tag{24}$$

Where $b \in \mathbb{R}$ is the bias parameter, and $\mathbf{w} = [w_1, w_2, \cdots, w_{\mathrm{d}}] \in \mathbb{R}^{\mathrm{d}}$ is a weight vector with d real-valued components. To capture the nonlinear relationship between the input layer and output layer, the weighted sum needs to undergo nonlinear processing defined as:

$$y = f(z) \tag{25}$$

Where y represents the neuron's activation value, and f represents the activation function, which ensures a differentiable and smooth transition throughout the network training process. This study uses the *linear* and *logsig* activation functions as defined below:

$$y = x \tag{26}$$

$$y = logsig(x) = \frac{1}{1 + e^{-x}} \tag{27}$$

The *linear* activation function scales the output from $-\infty$ to $+\infty$, while the *logsig* (log-sigmoid) activation function scales the output from 0 to 1.

The feed-forward back-propagation approach begins by feeding input data to the input layer during the feed-forward phase. The information then propagates from one neuron to others via connections in the network. Upon obtaining the output from the back-propagation phase, the next step involves evaluating the output or predicted value (p_i) by comparing it with the target values (t_i) using the mean square error (MSE), as shown in the following equation:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(t_i - p_i)^2 \tag{28}$$

Where n denotes the number of samples. The back-propagation then iterates to adjust weights and biases, minimizing MSE by assessing the errors associated with each weight and bias until convergence. Using the NNs, feature selection is performed in the following section.

2.4.2. Feature selection analysis

To perform feature selection, the features considered in the existing empirical solutions are first outlined in Table 3, showing their usage frequency and dependencies. Notably, the features outlined in Table 3 include all those that can be extracted from the collected data.

According to the conducted literature review, the plastic hinge of shear walls can generally be expressed as follows [20]:

$$L_p = \alpha H_e + \beta l_w + L_{SP} \tag{29}$$

As can be seen, the wall shear span and length are key features that must be considered for deriving the plastic hinge length approximation. As explained earlier (refer to Fig. 1), L_{SP} denotes the strain penetration length, which plays a significant role in lightly reinforced RC shear walls [15,17,22,23,37]. According to the proposed solutions for these walls, L_{SP} is a function of the yield and ultimate strength and the diameter of the longitudinal reinforcements, as well as the compressive strength of concrete.

In this study, although the diameter of longitudinal reinforcements is influential in lightly reinforced RC shear walls due to the direct correlation between the reinforcement ratio (ρ_{wv}) and d_{bl}, and the greater importance of ρ_{wv} on the response of RC structural elements, ρ_{wv} was selected for approximating the plastic hinge length rather than d_{bl}. The transverse reinforcement ratio, ρ_h, was also selected as a modeling feature, considering the significant role of transverse reinforcements in the lateral capacity of RC shear walls (see Fig. 3). It should be noted that both ρ_{wv} and ρ_h are directly correlated with the wall thickness (t_w); therefore, it is a feature that could potentially be redundant. Furthermore, eliminating t_w seems reasonable considering the role of ALR in estimating the plastic hinge length (refer to Table 3) and its inverse correlation of t_w.

Both the compressive strength of concrete (f_c') and the tensile strength of concrete (f_t) were initially selected as features for approximating the plastic hinge length, even though f_t can be estimated as a function of f_c'. This assumption was later tested through sequential backward elimination (SBE) to ensure that both features are necessary, contribute uniquely to the plastic hinge length and cannot be eliminated. The next feature to be eliminated is H, considering the significant contribution of the shear span (H_e) to the plastic hinge length of RC shear walls compared to the wall height. Furthermore, H was only used in one of the initial proposed solutions, while all later-developed solutions adopted H_e. V_{max} is a function of other features and can be calculated through section analysis. Notably, V_{max} was considered a key factor in the most recently developed solution, which achieved high accuracy in predicting the plastic hinge length. Although this feature is dependent, it was initially included as a distinctive feature to undergo the SBE process to determine whether it should be retained or eliminated.

Therefore, the initially eliminated features were d_{bl}, H and t_w. To further reduce the remaining features, the SBE process was performed. Using a trial-and-error approach, a network with one input layer, one output layer, and one hidden layer with five neurons, utilizing the log-sigmoid activation function, has been developed. The collected dataset was randomly divided into three subsets: 70 % for training, 15 % for testing, and 15 % for validation. These subsets were used to train, test and validate the network to approximate

Table 3
Characteristics of $L_{p,pred}/L_{p,exp}$ for the existing solutions and selected features.

No	Researchers	Design Parameters													$L_{p,pred}/L_{p,exp}$			
		l_w	t_w	ρ_{wv}	ρ_h	ALR	f'_c	f_t	f_y	f_u	H_e	H	V_{max}	d_{bl}	μ	σ	Min	Max
1	Paulay and Uzumeri [29]	x										x			2.18	2.16	0.31	34.93
2	Oesterle *et al.* [30]	x													5.22	6.71	0.72	120.44
3	Paulay [6]	x													3.92	5.03	0.54	90.33
4	Wallace and Moehle [8]	x													3.92	5.03	0.54	90.33
5	Paulay and Priestley [31]	x							x	x	x			x	2.03	2.26	0.47	36.86
6	Paulay and Priestley [32]	x									x				1.71	1.80	0.24	30.45
7	Sasani and Kiureghian [33]	x									x				1.91	2.45	0.26	44.02
8	Thomsen and Wallace [21]	x													2.17	2.79	0.30	49.98
9	Eurocode 8 [34]	x					x		x		x			x	1.70	1.95	0.37	33.94
10	Bohl and Adebar [19]	x				x					x				1.34	1.61	0.27	29.16
11	Wibowo *et al.* [22]													x	0.44	0.66	0.05	9.03
12	Kazaz [13]	x			x	x	x		x		x				1.68	1.93	0.26	33.48
13	Altheeb [23]						x		x		x			x	0.29	0.43	0.03	7.17
14	Hoult *et al.* [15]	x	x	x		x		x		x	x			x	0.70	1.06	0.00	21.10
15	Hoult [27]	x	x	x		x	x	x		x	x		x	x	1.34	1.34	0.27	22.71
16	Hoult [27]	x	x								x				1.13	1.44	0.24	18.02

the plastic hinge length, incorporating all ten remaining features. The network underwent iterative training, validation, and testing to converge to a minimum MSE over ten repetitions, with the average MSE serving as the final evaluation metric for the model. Each feature was then sequentially excluded one at a time, and the MSE was estimated with the remaining features. The calculated MSEs of the networks with all features and without each feature (10 models) are illustrated in Fig. 4.

As shown in Fig. 4, the calculated MSE of the model with all features is smaller than the other ten developed models, which means excluding any remaining features could not increase or result in a similar performance. Therefore, all the initially selected inputs were used to develop the approximation for the plastic hinge length.

2.4.3. Structure of the proposed NN models

The black-box nature of neural networks can be a major drawback of the approach, as it produces predictions or decisions without providing insight into how it arrived at those conclusions. This study aimed to enhance the transparency and interpretability of neural network algorithms in predicting the plastic hinge lengths of RC shear walls. This was achieved by developing two models, NN-1 and NN-2, using previously selected features, with one neuron and two neurons, respectively. The architecture of these models (NN-1 and NN-2) and their neuron structures, employing the log-sigmoid activation function, are illustrated in Fig. 5.

The explicit solutions using the developed models were derived based on the mathematical framework of neural networks explained below:

$$[z]_{1\times1} = [w_{HL}]_{1\times10}\cdot[\widetilde{x}]_{10\times1} + [b_{HL}]_{1\times1} \tag{30}$$

Where z is the weighted sum of the hidden layer (HL), w_{HL} and b_{HL} are the weight and bias of HL, which were adjusted during training and finalized through testing of the developed model (NN-1), and $\widetilde{x}$ is the normalized selected features in the previous section. As displayed in Fig. 5, the model employs a neuron in a hidden layer with a log-sigmoid activation function. Therefore, the weighted sum of the hidden layer can be estimated as follows:

$$[y_{NN-1}]_{1\times1} = 1/1 + e^{-[z]_{1\times1}} \tag{31}$$

Using the weight and bias of the output layer (OL) obtained after training and testing the model (NN-1), the normalized plastic hinge length ($\widetilde{L}_{P,NN-1}$)—which represents the weighted sum of the output layer—can be calculated as follows:

$$\widetilde{L}_{P,NN-1} = [w_{OL}]_{1\times1}\cdot[y_{NN-1}]_{1\times1} + [b_{OL}]_{1\times1} \in [-1,1] \tag{32}$$

Using $\widetilde{L}_{P,NN-1}$ and Eq. (23), along with the maximum and minimum observed values for plastic hinge length (see Table 2), the normalized plastic hinge length was then denormalized.

Considering employing two neurons in the hidden layer of the NN-2 model, $\widetilde{L}_{P,NN-2}$ can be calculated using the following equations:

$$[z]_{2\times1} = [w_{HL}]_{2\times10}\cdot[\widetilde{x}]_{10\times1} + [b_{HL}]_{2\times1} \tag{33}$$

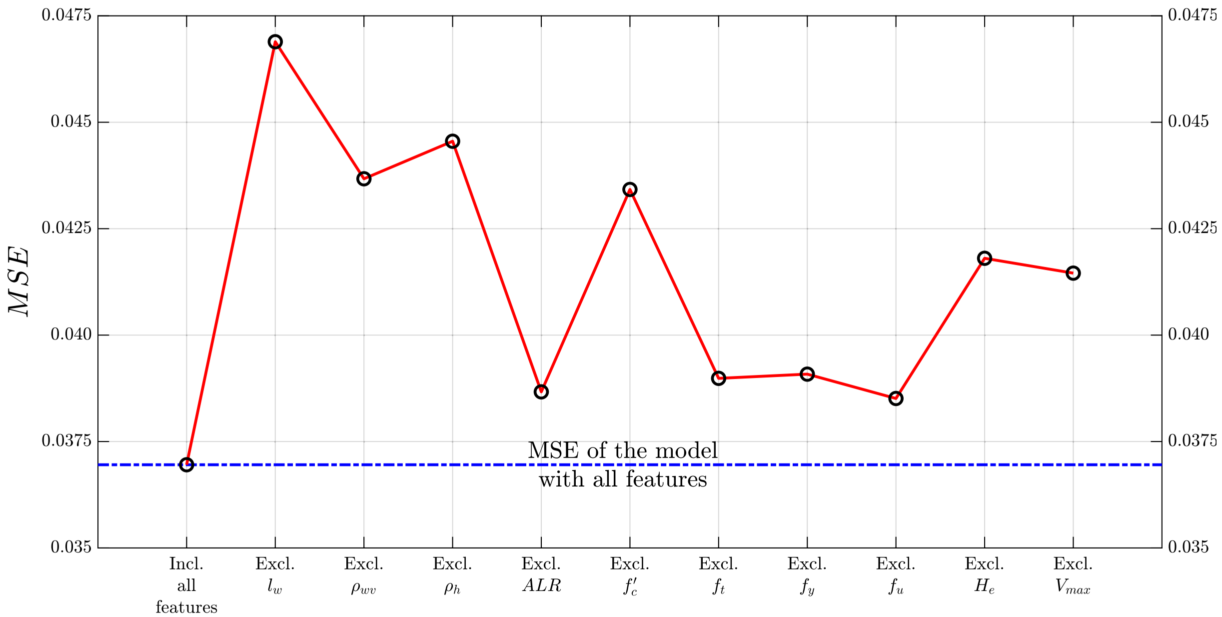

Fig. 4. The MSE of models with all features versus excluding each feature.

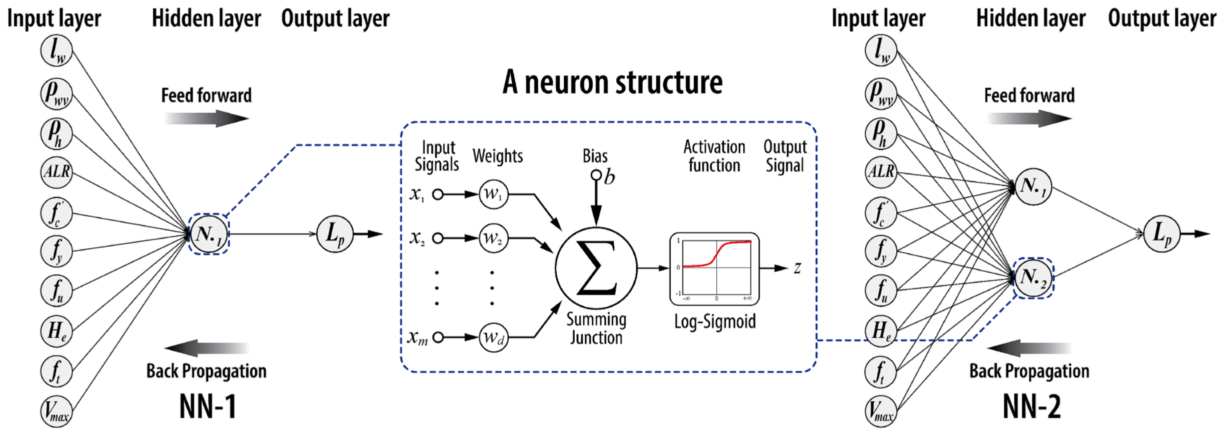

Fig. 5. The architecture of the developed models and a neuron structure schematic.

$$[y_{NN-2}]_{2\times1} = 1/1 + e^{-[z]_{2\times1}} \tag{34}$$

$$\widetilde{L}_{P,NN-2} = [w_{OL}]_{1\times2}.[y_{NN-2}]_{2\times1} + [b_{OL}]_{1\times1} \in [-1, 1] \tag{35}$$

The collected data was randomly divided into three subsets: training, testing, and validation, with each subset comprising 70 %, 15 %, and 15 % of the collected data, respectively. The training and testing subsets were then used to train and test the designed networks, enabling the calculation of weights and biases for the output and hidden layers in both models (NN-1 and NN-2). Using the obtained weights and biases of the output and hidden layers of the NN-1 model after training and testing stages, the closed-form approximation of the normalized plastic hinge length can be expressed as follows:

$$z_{NN-1} = -0.95l_w - 0.11\rho_{wv} - 0.027\rho_h + 0.58ALR - 0.17f_c' + 0.64\sqrt{f_c'} + 1.19f_y - 1.65f_u - 0.96H_e + 0.69V_{max} + 1.98 \tag{36}$$

$$\widetilde{L}_{P,NN-1} = -1.566y_{NN-1} + 0.585 \in [-1, 1] \tag{37}$$

It is noteworthy that y_{NN-1} needs to be calculated using Eq. (31). Furthermore, to simplify the derived equation, Eq. (36), f_t was approximated by its value ($\approx 0.6\sqrt{f_c'}$).

To estimate $\widetilde{L}_{P,NN-2}$, Eq. (33) was utilized along with the weights and bias values of the NN-2 model's hidden layer, as outlined in Table 4. To this end, initially, the weighted sum of HL, y_{NN-2}, needs to be calculated using Eq. (34).

The closed-form approximation of the normalized plastic hinge length using the NN-2 model can be expressed as follows:

$$\widetilde{L}_{P,NN-2} = [11.198 \quad 10.147].[y_{NN-2}]_{2\times1} - 11.071 \in [-1, 1] \tag{38}$$

To assist practical users in applying the developed ML-based solutions for estimating the plastic hinge length, a flowchart is provided below. This flowchart outlines the step-by-step process, ensuring clarity and ease of implementation. Fig. 6 illustrates a flowchart outlining the step-by-step process for using the developed ML-based solutions to estimate the plastic hinge length of RC shear walls.

Using Eqs. (37) and (38), both $\widetilde{L}_{P,NN-1}$ and $\widetilde{L}_{P,NN-2}$ were calculated and denormalized. For comparison purposes, the predicted and observed values of L_p are displayed in Fig. 7.

Fig. 7 demonstrates that the developed models predict the plastic hinge length of shear walls with acceptable accuracy. Although Fig. 7 does not quantitatively assess the models' performance, it indicates that the estimation error for the plastic hinge length is generally within ± 20 %. Model NN-2 performs significantly better than model NN-1, attributed to the additional neuron in its hidden layer. Both models show larger discrepancies between the predicted and observed plastic hinge lengths at lower values, which diminish as the plastic hinge length increases.

Table 4
Weights and biases values of the hidden layer of the NN-2 model.

Neuron	w_{HL}										b_{HL}
	l_w	ρ_{wv}	ρ_h	ALR	f_c'	$\sqrt{f_c'}$	f_y	f_u	H_e	V_{max}	
$N_{.1}$	−0.138	0.097	−1.006	−2.461	−1.222	−1.00	−4.687	4.233	0.551	0.625	−5.118
$N_{.2}$	0.344	−0.097	1.061	2.698	1.378	1.076	5.188	−4.674	−0.384	0.791	5.628

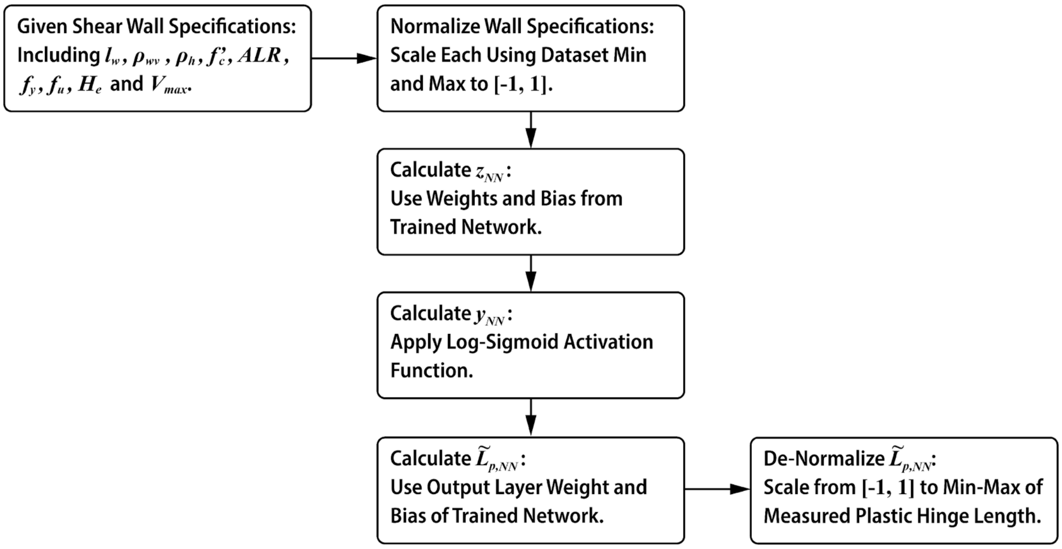

Fig. 6. Flowchart for applying the ML-based solutions to estimate the plastic hinge length of RC shear walls.

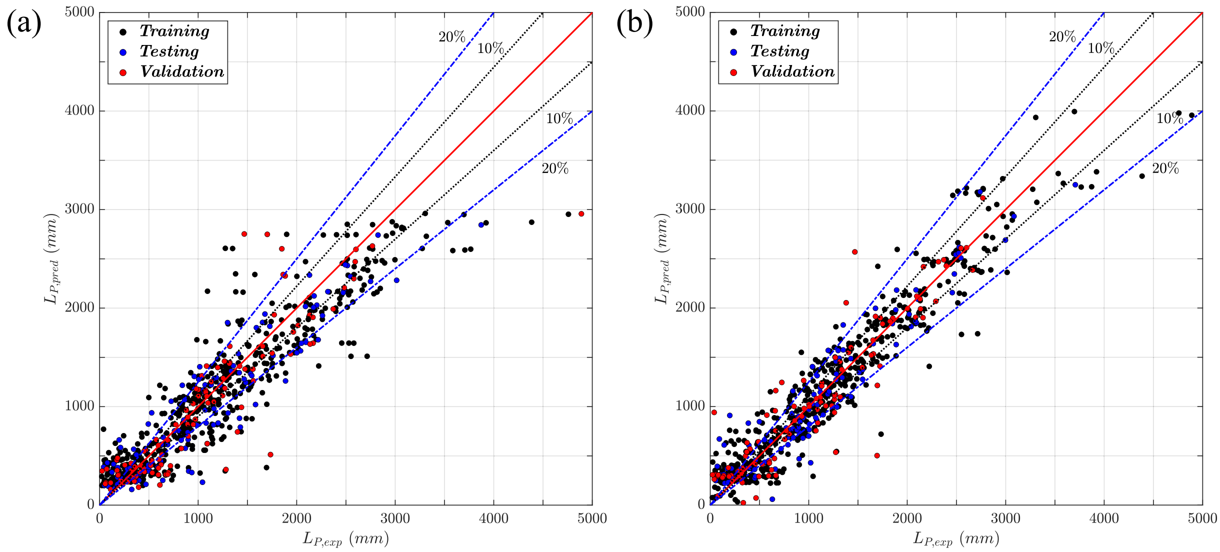

Fig. 7. Predicted ($L_{p,pred}$) versus experimental/numerical ($L_{p,exp}$) values during training, testing, and validation stages with ± 10 % and ± 20 % error bounds for models (a) NN-1 and (b) NN-2.

3. Results and discussion

The following subsections evaluate the developed solutions' performance and compare it to existing solutions available in the literature. Additionally, the explainability of the developed solutions is discussed, and their robustness and reliability are assessed through a reliability analysis with comparisons to existing solutions.

3.1. Performance evaluation

This section evaluates the performance of the developed models by estimating their accuracy and explainability. Initially, the solutions' accuracy was assessed using four widely accepted statistical indices: mean error (*ME*), root mean square error (*RMSE*), the correlation coefficient (*r*), and the coefficient of determination (R^2). These statistical indices are defined mathematically by the following equations:

$$ME = 1 \Big/ n\sum_{i=1}^{n}(t_i - p_i) \tag{39}$$

$$RMSE = \sqrt{1 \Big/ n\sum_{i=1}^{n}(t_i - p_i)^2} \tag{40}$$

$$r = \frac{n\sum t_i p_i - \sum t_i \sum p_i}{\sqrt{\left(n\sum t_i^2 - \left(\sum t_i\right)^2\right)\left(n\sum p_i^2 - \left(\sum p_i\right)^2\right)}} \tag{41}$$

$$R^2 = \frac{\left(n\sum t_i p_i - \sum t_i \sum p_i\right)^2}{\left(n\sum t_i^2 - \left(\sum t_i\right)^2\right)\left(n\sum p_i^2 - \left(\sum p_i\right)^2\right)} \tag{42}$$

Although a closed-form approximation of plastic hinge length was derived using NNs in this research, SHapley Additive exPlanations (SHAP) [38] was used to further investigate the explainability of the developed solutions.

In game theory, the Shapley value is a method for distributing the total gain generated by a coalition of players among the players themselves. Given a game with a set of players N and a value function $\upsilon : 2^N \longrightarrow \mathbb{R}$ that assigns a value to each coalition $S \subseteq N$, the Shapley value for a player i is defined as:

$$\phi_i(\upsilon) = \sum_{S\subseteq N\setminus\{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!}[\upsilon(S \cup \{i\}) - \upsilon(S)] \tag{43}$$

Where N is the set of all players, S is a subset of N not containing player i, $|S|$ and $|N|$ are the size of subset S and number of all players, respectively, and $\upsilon(S)$ is the value function representing the worth of coalition S. Equation (43) considers all possible orders in which players can join the coalition and computes the average marginal contribution of each player.

To apply Shapley values to machine learning, each feature can be treated as a player in a coalition, and the prediction function, f, as the value function. For a particular instance x, the SHAP value for feature i is given by:

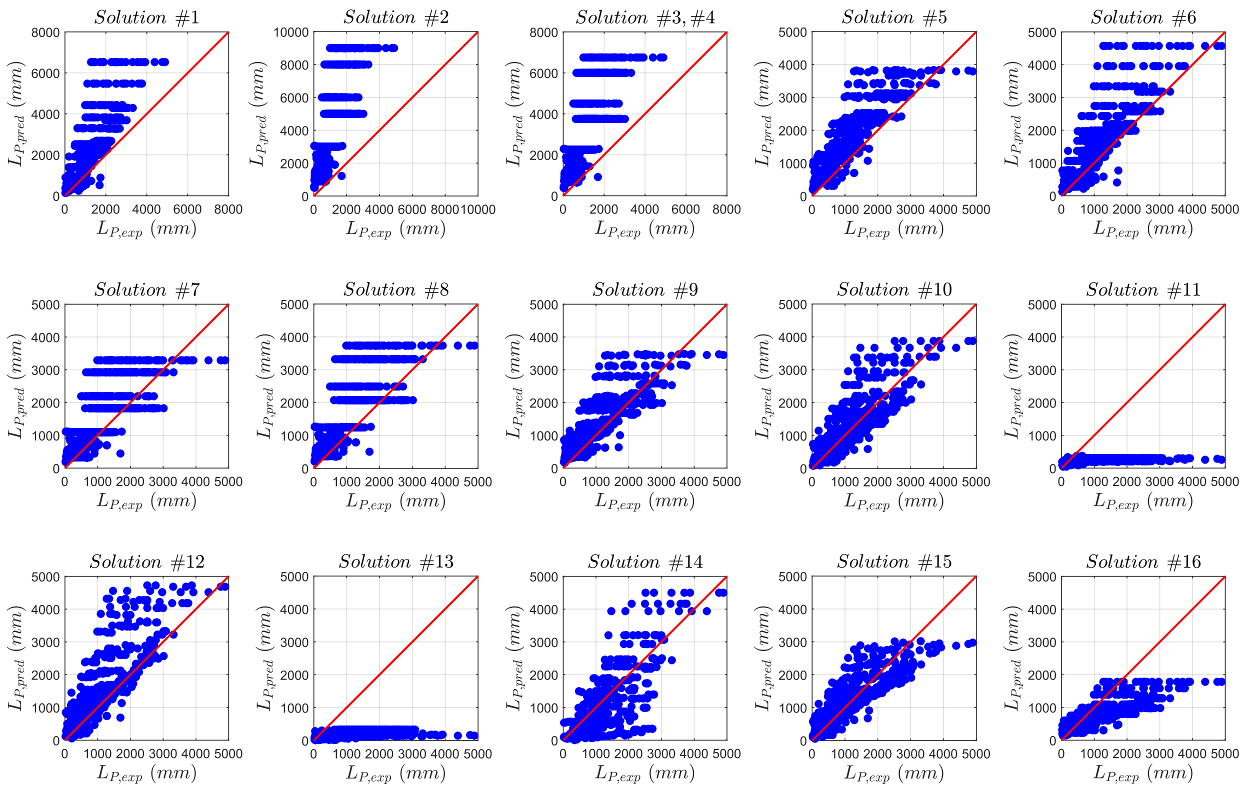

Fig. 8. Predicted ($L_{p,pred}$) versus experimental/numerical ($L_{p,exp}$) of plastic hinge length using the existing solutions (refer to Table 1).

$$\phi_i(f,x) = \sum_{S\subseteq F\setminus\{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!}[f(x_S\cup\{x_i\}) - f(x_S)] \tag{44}$$

Where F is the set of all features, S is a subset of features not containing i, x_S denotes the instance x with only the features in subset S, $f(x_S\cup\{x_i\})$ is the prediction model using the subset S and adding feature i and $f(x_S)$ is the prediction model using only the subset S. Equation (44) evaluates the difference in the prediction when feature i is added to the subset S, averaged over all possible subset S.

3.1.1. Accuracy evaluation of existing and proposed solutions

To intuitively investigate the accuracy of the empirical solutions, the experimentally and numerically measured plastic hinge lengths ($L_{p,exp}$) are initially compared with the predicted plastic hinge lengths ($L_{p,pred}$) using the existing solutions outlined in Table 1, as shown in Fig. 8.

As shown in Fig. 8, solutions #1 to #9 and #12 approximately overestimate the plastic hinge length, whereas solutions #11, #13 and #16 underestimate it. Despite some discrepancies, solutions #10, #14 and #15 can outperform the existing solutions. As shown in Fig. 8, the initially derived solutions overestimate the plastic hinge lengths. Additionally, solutions #11 and #13 are specifically derived for lightly reinforced shear walls; therefore, they are expected to underestimate the plastic hinge length. Although higher accuracy in approximating the plastic hinge length is desirable, a solution that underestimates this length may be considered superior. A shorter plastic hinge implies a reduced plastic deformation capacity of shear walls, which is more conservative in the RC shear wall design context. To this end, the mean error (*ME*) and σ (standard deviation) of the existing solutions and the developed models, NN-1 and NN-2, hereafter referred to as proposed solutions (PS) #1 and #2, respectively, are displayed in Fig. 9.

As shown in Fig. 9, the mean error of solutions #1 to #9 and #12 is negative (indicating overestimation), while the mean error of solutions #11, #13, #14 and #16 is positive (indicating underestimation). It should be noted that solutions #11 and #13 are highly conservative, while solutions #10 and PS #2, despite overestimating the plastic hinge length, are significantly accurate due to their small mean error values. In contrast, solutions #15 and PS #1 underestimate the plastic hinge length but are significantly accurate, as indicated by their small mean error values. As shown in Fig. 9, the standard deviations (σ) of PS #1 and PS #2 are smaller than those of all empirical solutions. This indicates that, for similar mean errors, the proposed solutions demonstrate higher consistency and precision in estimating the plastic hinge length, with predictions tightly clustered around the estimated mean error. To further investigate and compare the accuracy of the existing and proposed solutions, the root mean square error (*RMSE*) for both types of solutions is calculated and illustrated in Fig. 10.

As shown in Fig. 10, both proposed solutions #1 and #2 have smaller *RMSE* values compared to those obtained from the existing solutions. This suggests that the proposed solutions provide more accurate estimations of the plastic hinge length, with predictions closer to the observed values. Among the existing solutions, #15, #10, and #9 exhibit smaller *RMSE* values relative to the others. Although proposed solutions #1 and #2 use only one and two neurons, respectively, in the hidden layer to derive closed-form approximations, their calculated *RMSE* values indicate that these solutions considerably outperform others. This validates the efficiency of the models and confirms the effectiveness of using a low number of neurons. To provide a comprehensive assessment, the coefficient of determination (R^2) for existing solutions and proposed solutions are estimated and illustrated in Fig. 11.

Fig. 11 shows that the R^2 of the proposed models is higher than that of the existing solutions, indicating that the proposed solutions perform better in explaining and predicting the plastic hinge length. This suggests that the proposed solutions more effectively capture the relationship between the selected features and the plastic hinge length. Based on the results from the calculated mean errors, standard deviations of mean errors, root mean squared errors, and coefficients of determination, solutions #5 to #10, #12, and #14 to #16 are selected for further comparison with the proposed solutions. Fig. 12a demonstrates the calculated plastic hinge lengths for the

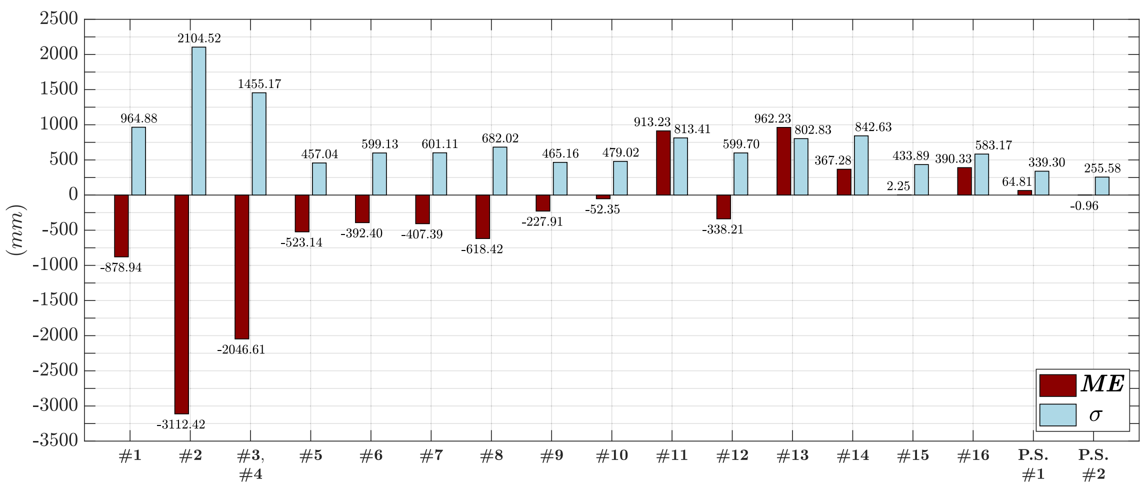

Fig. 9. *ME*And σ comparison of existing and proposed solutions.

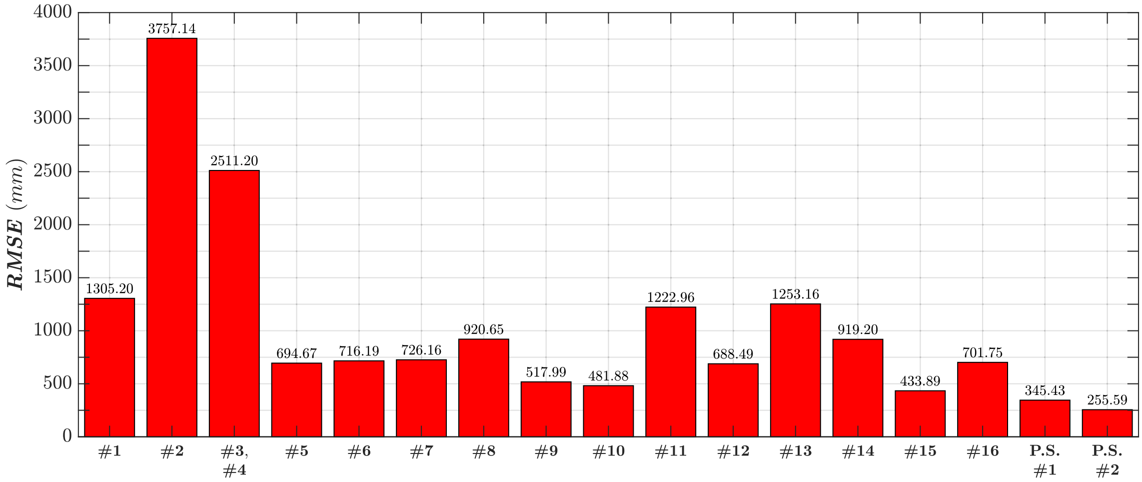

Fig. 10. *RMSE*Comparison of existing and proposed solutions.

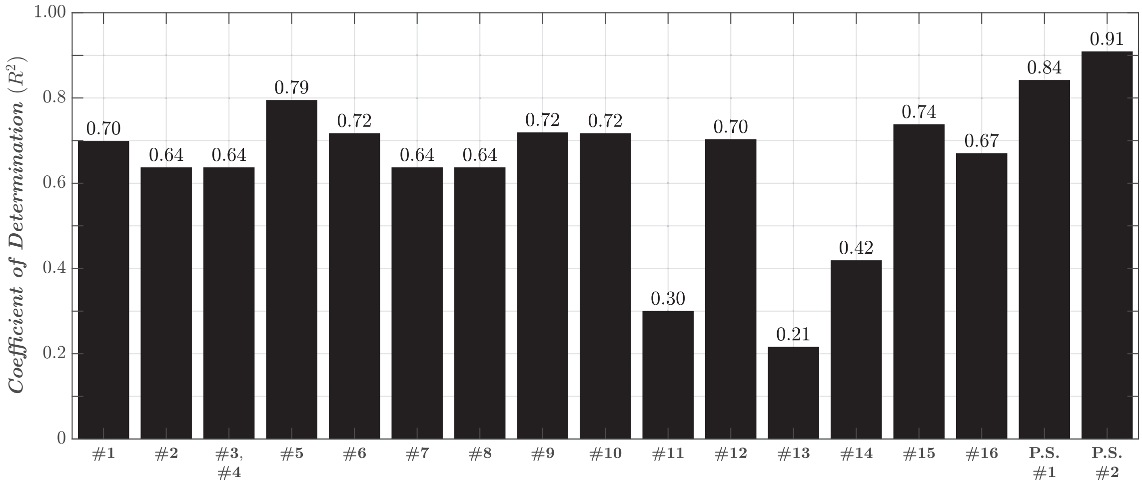

Fig. 11. The coefficient of determination (R^2) for existing and proposed solutions.

selected solutions alongside the observed values (both numerical and experimental). Additionally, Fig. 12b uses a Taylor diagram to illustrate the relationship between the correlation coefficient (r) and standard deviation (σ) of the collected (numerical and experimental) and predicted values.

Fig. 12a shows that the proposed solutions have less discrepancy and higher accuracy than existing solutions. Additionally, proposed solution #2 exhibits smaller discrepancies and greater accuracy than proposed solution #1, likely due to the larger number of neurons in its hidden layer. Fig. 12b shows the red dashed line representing the standard deviation of the observed values (experimental and numerical). The proposed solutions #1 and #2 have high correlation coefficients (> 0.916), indicating a stronger correlation with the collected data and greater accuracy than existing solutions. Moreover, the standard deviations of the proposed solutions are closer to the standard deviation of the collected data, demonstrating their consistency in predicting the plastic hinge length. To examine in greater detail, Fig. 13 illustrates the distribution of the ratio of predicted to observed plastic hinge lengths for proposed and existing solutions.

As shown in Fig. 13, proposed solution #1 demonstrates higher accuracy ($\mu = 1.10$) compared to other studied solutions. The lower mean value and standard deviation of the predicted-to-observed plastic hinge length for solution #1 indicate its stability, reliability, and accuracy. Solutions #10, #15, #16, and proposed solution #2 also outperform other solutions. Based on the calculated mean errors, standard deviations of mean errors, root mean square errors, correlation coefficients, and coefficients of determination, solutions #9, #10 and #15, and proposed solutions #1 and #2 are shortlisted for further probabilistic assessment in the following subsections.

3.1.2. *Explainability of the proposed solutions*

Although the proposed solutions are derived explicitly using NNs, specifically Eqs. (31), (36) and (37) for proposed solution #1 and

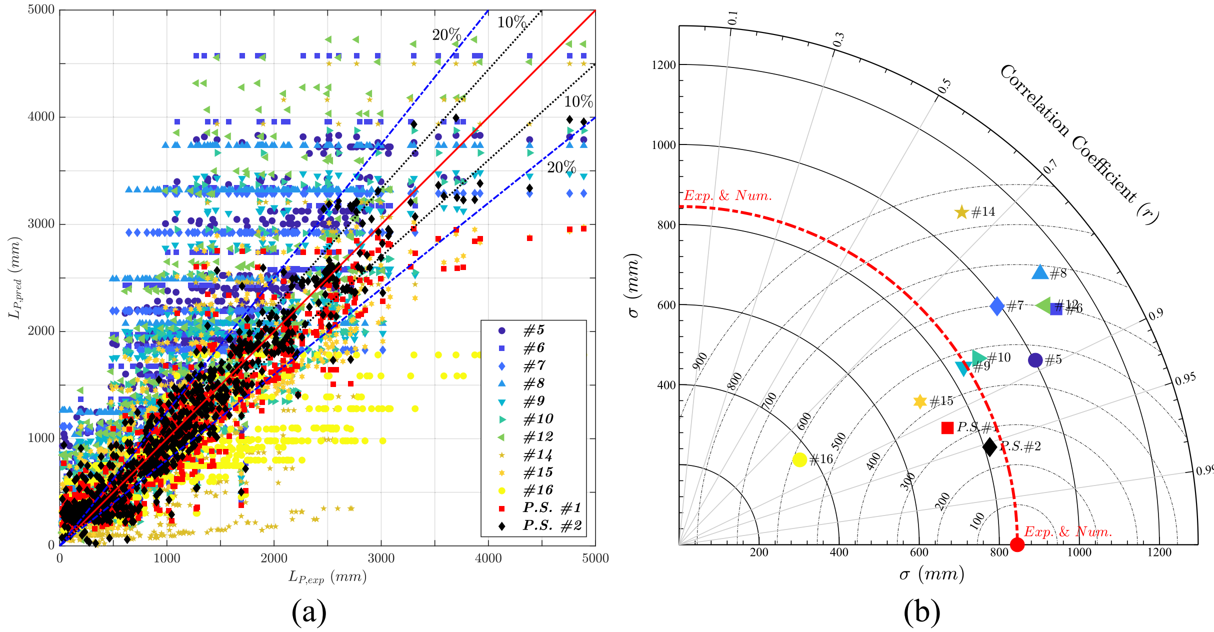

Fig. 12. Predicted ($L_{p,pred}$) versus observed ($L_{p,exp}$) values of the selected and proposed solutions. (a) scatter plot with ± 10 % and ± 20 % error bounds and (b) Taylor diagram.

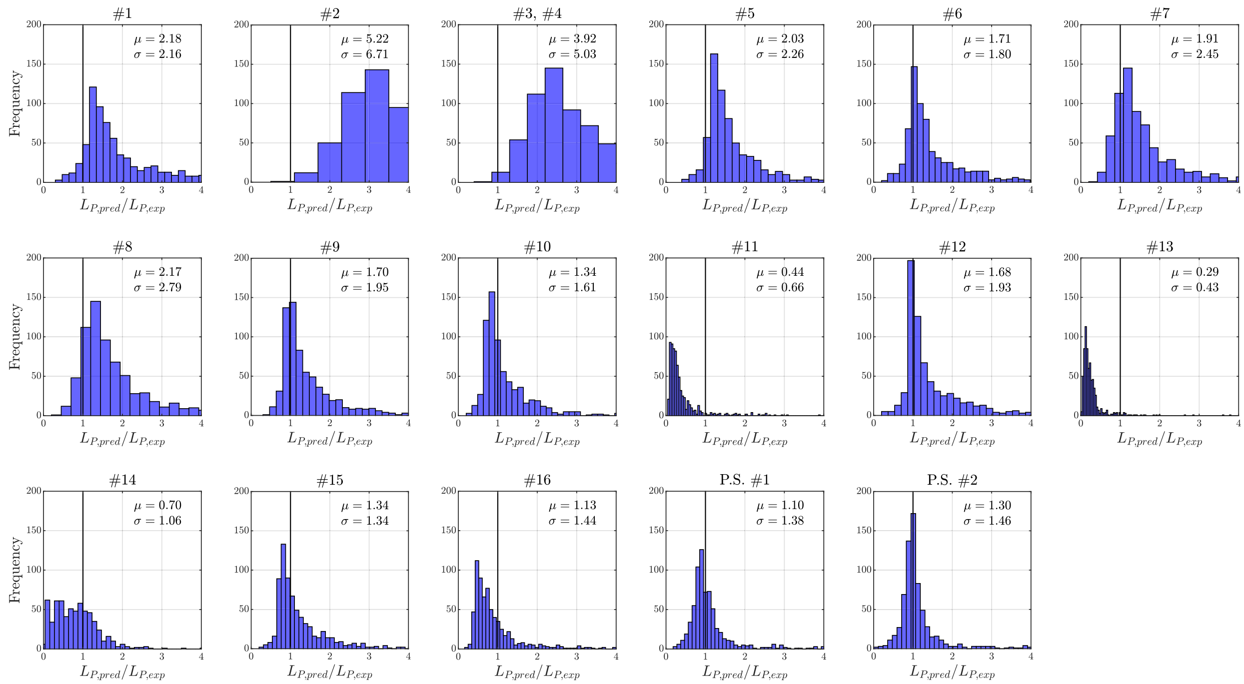

Fig. 13. Histograms of the predicted to observed plastic hinge lengths ratio for existing and proposed solutions.

Eqs. (34) and (38) for proposed solution #2, SHAP analysis was utilized to thoroughly investigate the interpretability of these solutions for plastic hinge length in shear walls. The SHAP analysis elucidates the influence and importance of various features within the two proposed models. The SHAP summaries of the proposed models are demonstrated in Fig. 14.

Fig. 14a and b illustrate how different features influence the predictions of two proposed models for plastic hinge length. Each dot represents a feature value's contribution to a single prediction. In Fig. 14a, for proposed solution #1, the top three contributing features are l_w, f_t and H_e, with SHAP values ranging from −0.2 to 0.3. Higher values of l_w, f_t and H_e generally increase the predicted

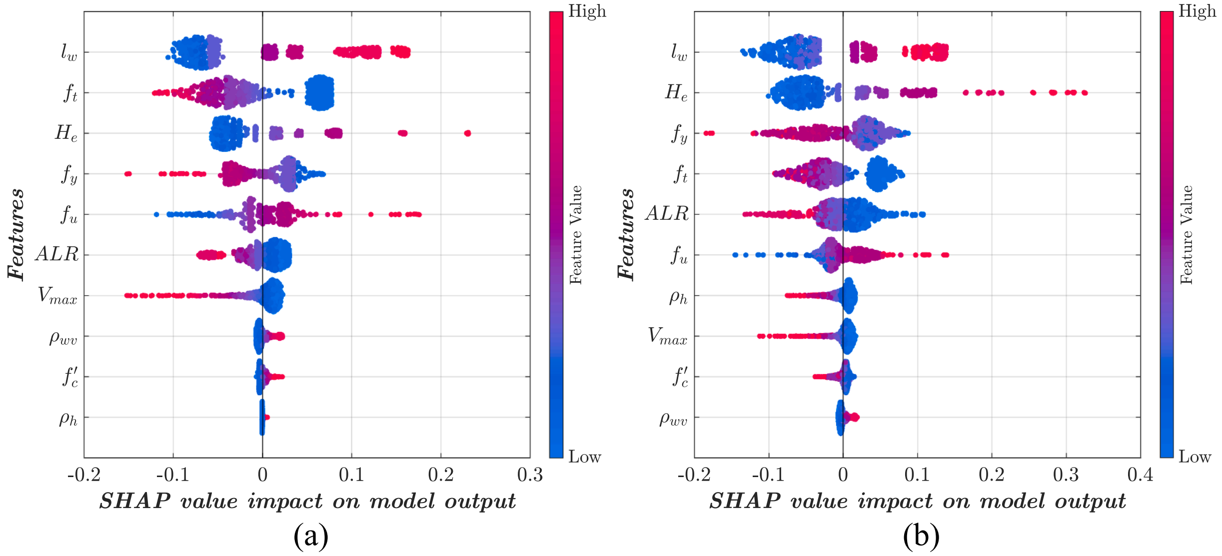

Fig. 14. SHAP summary plots for the proposed solutions predicting plastic hinge length: (a) proposed solution #1 and (b) proposed solution #2.

plastic hinge length, while features such as f_u and f'_c exhibit mixed contributions but tend to increase the hinge length overall. Fig. 14b shows that for proposed solution #2, the dominant features are l_w, H_e, f_t and ALR, with SHAP values spanning approximately from -0.2 to 0.4. Similar trends are observed in Fig. 14b as in Fig. 14a, but the impact of f_t and H_e appears stronger, and ALR also shows notable influence. Overall, both models consistently identify l_w and H_e as the most influential features in predicting plastic hinge length. Fig. 15 shows bar plots indicating feature importance based on mean absolute SHAP values for both proposed solutions.

This visualization ranks the features by their average contribution magnitude, indicating their relative importance in the predictive models. As can be seen in Fig. 15a, the most critical feature is l_w and another important feature is H_e; it is noteworthy that both f_t and f_y are less influential compared to l_w and H_e. Considering Fig. 14a and b and Fig. 15a indicate that l_w and H_e are the predominant features driving the model predictions, other features such as ρ_{wv} and f'_c exhibit lower overall importance. In Fig. 15b, l_w and H_e remain the most influential features. Additionally, f_y and f_t show substantial importance following l_w and H_e, whereas features such as ALR and V_{max} contribute meaningfully but are less significant compared to the top features.

3.2. Probabilistic assessment

The existing and proposed solutions for calculating the plastic hinge length of shear walls are straightforward. However, due to the probabilistic nature of the design parameters (refer to Table 3), the results are influenced by inherent uncertainties in these variables. Therefore, to understand the effect of these uncertainties on the calculated plastic hinge length, this section comprehensively investigates the reliability of the shortlisted solutions. To this end, a limit state function was first defined for each solution. Next, by

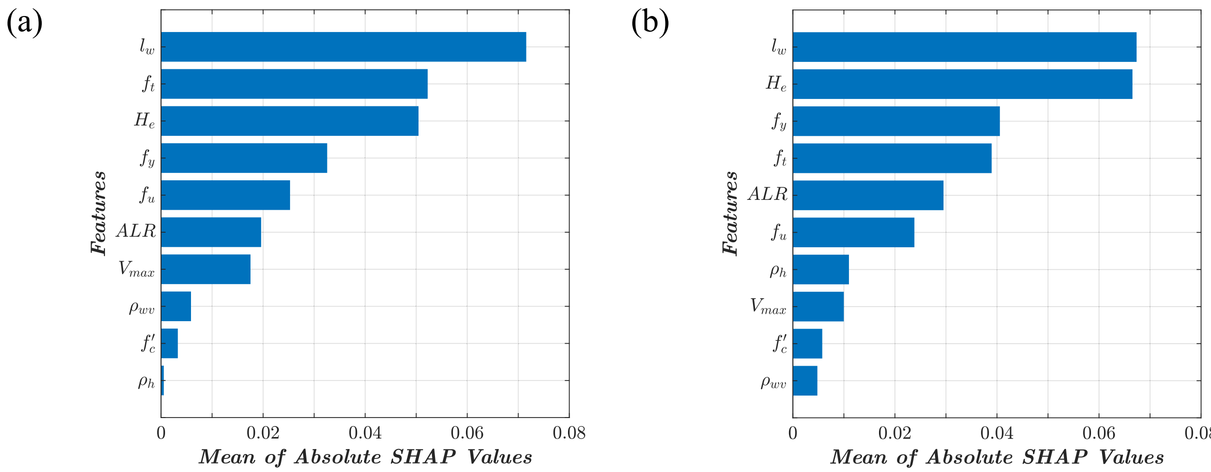

Fig. 15. Overall importance of features contributing to plastic hinge length using SHAP: (a) proposed solution #1 and (b) proposed solution #2.

considering the input parameters of the shortlisted solutions as random variables, the Monte Carlo Simulation (MCS) method was employed to generate probable realizations of these variables. Using the defined limit state functions of the shortlisted solutions, the exceedance probability of each solution was estimated, transforming the solutions from a deterministic to a probabilistic framework. Finally, the estimated exceedance probabilities were used to calculate and further investigate the reliability indexes of the solutions.

3.2.1. Limit state function

If the approximated plastic hinge length by the shortlisted solutions is denoted as $L_p(\mathbf{x})$, where $\mathbf{x}$ represents the design variables influencing the plastic hinge length, and $L_p^T(\mathbf{x})$ represents the threshold of the plastic hinge length, the limit state function, $g(\mathbf{x})$, can be defined as follows:

$$g(\mathbf{x}) = L_p^T(\mathbf{x}) - L_p(\mathbf{x}) \tag{45}$$

where $\mathbf{x}$ is a realization of random variables. According to the limit state function definition, g, a failure is indicated by whether the estimated plastic hinge length exceeds the threshold value. Assuming that N vectors, each containing n random variables, are generated using MCS method. The joint probability density function (PDF) of random variables is given by:

$$f_{X_1X_2\ldots X_n}(x_1.x_2.\cdots.x_n) = f_{X_1}(x_1).f_{X_2}(x_2).\cdots.f_{X_n}(x_n) \tag{46}$$

The failure probability, P_f, can be estimated as:

$$P_f = \int_D f_{\mathbf{x}_1\mathbf{x}_2\ldots\mathbf{x}_n}(x_1.x_2.\cdots.x_n).I\{g(\mathbf{x}) \geq 0\}dx_1dx_2\ldots dx_n \tag{47}$$

Where D represents the domain over which the joint PDF is defined, $g(\mathbf{x})$ is the limit state function for the realization $\mathbf{x}$, and $I\{.\}$ is the indicator function, which equals 1 if the condition inside the braces is satisfied, and 0 otherwise.

To solve Eq. (47), methods such as the first-order reliability method (FORM) and the first-order second moment (FOSM) can be implemented. However, the developed limit state function is highly nonlinear due to using NNs trained to estimate the plastic hinge length of RC shear walls. Furthermore, both FORM and FOSM are based on the linearization of the performance function via the Taylor expansion. Consequently, for the proposed limit state function in Eq. (45), using FORM and FOSM results in an inaccurate approximation of the failure probability. To overcome the numerical difficulty associated with the nonlinearity of the defined limit state functions, more sophisticated reliability methods can be employed. One such approach is the MCS method, which, although computationally expensive, does not rely on the linearization of the performance function and can accurately capture the nonlinear behavior. Moreover, given the high accuracy of the proposed solutions in estimating the plastic hinge length and its inherent analytical closed-form structure, using the MCS method results in a reasonably accurate and efficient solution for Eq. (47).

By using a discrete approximation of the failure probability with the MCS method, Eq. (47) can be estimated with the following equation:

$$P_f = \frac{1}{N}\sum_{j=1}^{N} I\{g(\mathbf{x}^j) \geq 0\} \tag{48}$$

where N is the number of simulations employing the MCS method. Using the estimated failure probability of a solution, its reliability index (β) can be determined with the following equation:

$$\beta = -\Phi^{-1}(P_f) \tag{49}$$

Where Φ is the cumulative distribution function of the standard normal distribution.

To estimate the limit state function, the threshold of the plastic hinge length, $L_p^T(\mathbf{x})$, needs to be determined first. For accurate estimation of the limit state function value for a realization of random variables, $\mathbf{x}$, the threshold value is formulated as a function of $\mathbf{x}$. In this study, rather than setting the threshold proportional to the wall length (l_w)—as in l_w [6,8,30], $0.5l_w$ [15,21], or $0.8l_w$ [19,27]—the threshold is expressed proportionally to the wall shear span (H_e). For flexure-governed walls, expressing the threshold in proportion to the wall's shear span rather than its length is more appropriate, given the intrinsic similarity between the plastic hinge length and the wall shear span, as well as the shear span's prominent contribution. Therefore, assuming a normal distribution for the estimated ratio, the threshold is set as follows:

$$L_p^T = \mu_{L_p/H_e} + 3\sigma_{L_p/H_e} \tag{50}$$

where μ and σ are the calculated mean and standard deviation of the estimated ratios (L_p/H_e) for the experimental data. Using Eq. (50), the threshold for the plastic hinge length was obtained as $H_e/3$. It is noteworthy that using Eq. (50) accurately provides the upper limit of the estimated ratio because it includes nearly all possible data points (≈ 99.7 %). This estimated threshold was later used to calculate the exceedance probability of the shortlisted solutions.

3.2.2. Generating realizations

Influential parameters affecting the shortlisted solutions were selected as random variables (RVs) to generate possible realizations.

These include l_w, t_w, ρ_{wv}, ρ_h, ALR, f'_c, f_t, f_y, f_u, H_e, V_{max} and d_{bl}. However, some of these variables (specifically f_t, V_{max} and d_{bl}) are dependent variables derived from independent RVs. Thus, realizations were generated using only the independent RVs. To this end, the PDFs of these independent RVs were determined by fitting curves to their frequency distributions based solely on the collected experimental data. Since the wall shear span (H_e) can be expressed as a function of the wall's height (H), an independent variable, both H_e and H were chosen for generating realizations. The determined PDFs of the independent RVs, along with their corresponding location, scale, and shape parameters, are listed in Table 5.

To preserve the dependence structure among the selected RVs when generating possible realizations, correlations between the RVs were carefully considered. The correlation matrix was calculated based on collected experimental data and is illustrated in Fig. 16.

Using the determined PDFs of selected RVs, their calculated correlation matrix, and distribution parameters obtained from experimental and numerical data, possible realizations were generated employing the MCS method. Initially, 12,000 realizations of the selected RVs were generated in this research. These realizations were then utilized to assess the variation in exceedance probability for the shortlisted solutions. Defined limit state functions were employed to ensure the convergence of the function and to verify the sufficiency of the generated scenarios.

To calculate the maximum lateral capacity of the wall (V_{max}), a computational code based on Bernoulli–Euler's theory was developed to compute strain along the shear wall cross-section. Strains in concrete and steel reinforcements were calculated using curvature and coordinates, utilizing stress–strain relationships defined by Mander *et al.* [39] for concrete and Ramberg and Osgood [40] for steel reinforcements. In estimating V_{max} using the generated RVs, 10 % of the wall's length was designated as boundary elements, with 10 % of the total longitudinal reinforcements assigned to each boundary element. Subsequently, to estimate d_{bl} (the diameter of the longitudinal reinforcements), the spacing of longitudinal reinforcements outside the boundary elements was set to 200 mm when the longitudinal reinforcement ratio (ρ_{wv}) exceeded 0.015; otherwise, it was set to 100 mm.

3.2.3. Reliability analysis of solutions

Given the sample-based nature of the MCS method, its accuracy depends on the number of simulations (generated realizations) used to calculate the exceedance probability. While a higher number of simulations generally improves accuracy, excessively large numbers lead to higher computational costs. To determine a reasonable/optimal number of simulations, the variation in exceedance probability was analyzed to assess convergence. The variation in exceedance probability across different simulation numbers was calculated and illustrated in Fig. 17.

As shown in Fig. 17, the probability of exceedance for solutions converges as the number of simulations increases. Specifically, beyond 10,000 simulations, the variation in the exceedance probability of the shortlisted solutions stabilizes and aligns with the values displayed in Fig. 17. Therefore, the initially selected number of simulations (12,000) is sufficient for obtaining accurate results, and increasing the number of simulations beyond this point is unnecessary. It is important to note that, given the number of simulations, the maximum reliability index obtained is 3.765, corresponding to $-\Phi^{-1}(1/12000)$. Fig. 17 further illustrates that using the defined limit state function, Eq. (45), and setting the threshold for the plastic hinge length at $H_e/3$, proposed solution #2 yields the lowest probability of exceedance. This indicates that proposed solution #2 has the highest reliability among the shortlisted scenarios. Moreover, both proposed solution #1 and solution #15 [27] exhibit a similar level of reliability. On the other hand, solution #9 [34]

Table 5
Probabilistic characteristics of the selected features.

Features	Distribution	Probability function	Parameters
l_w	Lognormal	$f(x) = \frac{1}{(x-\gamma)\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{\ln(x-\gamma)-\mu}{\sigma}\right)^2\right)$	$\sigma = 0.39$ $\mu = 7.06$ $\gamma = 164.3$
t_w	Gumbel max	$f(x) = \frac{1}{\sigma}\exp\left(-\left(\frac{x-\mu}{\sigma}\right)-\exp\left(-\frac{x-\mu}{\sigma}\right)\right)$	$\sigma = 37.86$ $\mu = 107.62$
ALR	Gamma	$f(x) = \frac{x^{\alpha-1}}{\beta^\alpha\Gamma(\alpha)}\exp\left(-\left(\frac{x}{\beta}\right)\right)$	$\alpha = 1.54$ $\beta = 0.067$
f'_c	Lognormal	$f(x) = \frac{1}{(x-\gamma)\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{\ln(x-\gamma)-\mu}{\sigma}\right)^2\right)$	$\sigma = 0.508$ $\mu = 3.352$ $\gamma = 8.113$
ρ_{wv}	Lognormal	$f(x) = \frac{1}{(x-\gamma)\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{\ln(x-\gamma)-\mu}{\sigma}\right)^2\right)$	$\sigma = 1.02$ $\mu = -4.810$ $\gamma = 0.000673$
ρ_h	Lognormal	$f(x) = \frac{1}{(x-\gamma)\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{\ln(x-\gamma)-\mu}{\sigma}\right)^2\right)$	$\sigma = 0.85$ $\mu = -5.531$ $\gamma = 0.00109$
f_y	Gumbel max	$f(x) = \frac{1}{\sigma}\exp\left(-\left(\frac{x-\mu}{\sigma}\right)-\exp\left(-\frac{x-\mu}{\sigma}\right)\right)$	$\sigma = 91.61$ $\mu = 432.95$
H_e	Gamma	$f(x) = \frac{x^{\alpha-1}}{\beta^\alpha\Gamma(\alpha)}\exp\left(-\left(\frac{x}{\beta}\right)\right)$	$\alpha = 2.99$ $\beta = 1152.6$
f_u	Gumbel max	$f(x) = \frac{1}{\sigma}\exp\left(-\left(\frac{x-\mu}{\sigma}\right)-\exp\left(-\frac{x-\mu}{\sigma}\right)\right)$	$\sigma = 131.27$ $\mu = 609.53$
H	Lognormal	$f(x) = \frac{1}{(x-\gamma)\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{\ln(x-\gamma)-\mu}{\sigma}\right)^2\right)$	$\sigma = 0.358$ $\mu = 8.16$ $\gamma = -687.17$

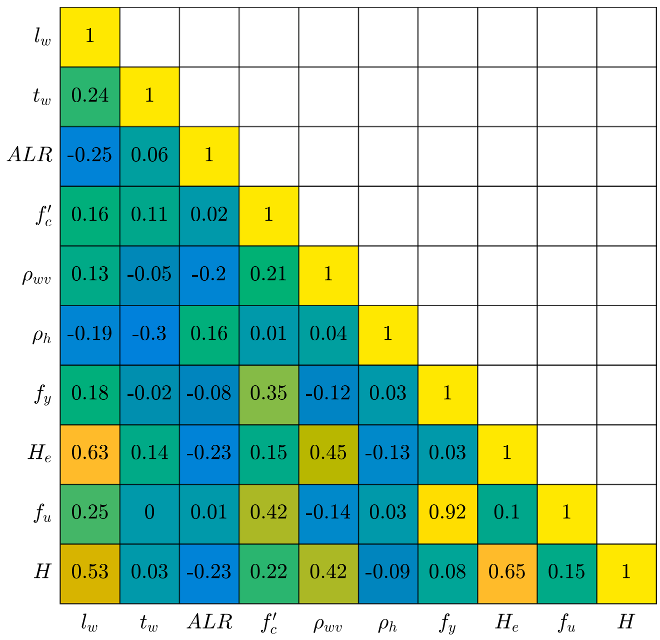

Fig. 16. Random features correlation matrix.

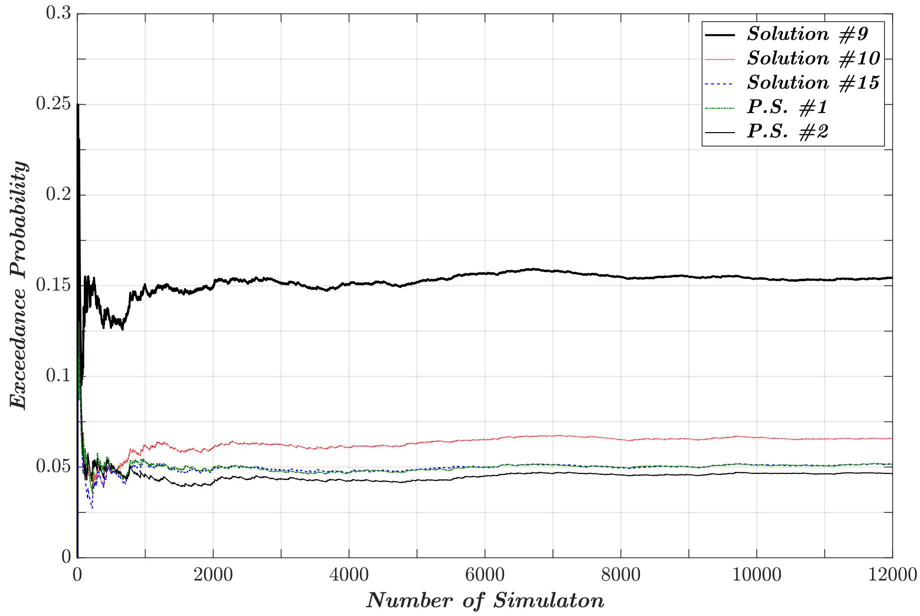

Fig. 17. Convergence assessment of limit state functions for the shortlisted solutions.

shows a higher probability of exceedance, suggesting that it is less reliable in calculating the plastic hinge length compared to the other solutions. The exceedance probability of the shortlistd solutions was estimated using different thresholds for the plastic hinge length L_p^T spans from 0 to H_e and illustrated in Fig. 18.

Fig. 18 presents the exceedance probability of shortlisted solutions as a function of the normalized plastic hinge length (L_p/H_e). As the plastic hinge length increases from 0 to H_e, all solutions demonstrate a significant decrease (beyond $0.1H_e$) in exceedance probability, eventually approaching zero, except for solution #9. Solution #9 consistently shows the highest probability of exceedance across all L_p values, indicating it is the least reliable solution for all considered threshold values. In contrast, proposed solution #2 has the lowest exceedance probability, demonstrating the highest reliability. Proposed solution #1 and solution #15 exhibit similar reliability, both performing better than solution #9 but not as well as proposed solution #2. Solution #10 also shows a higher

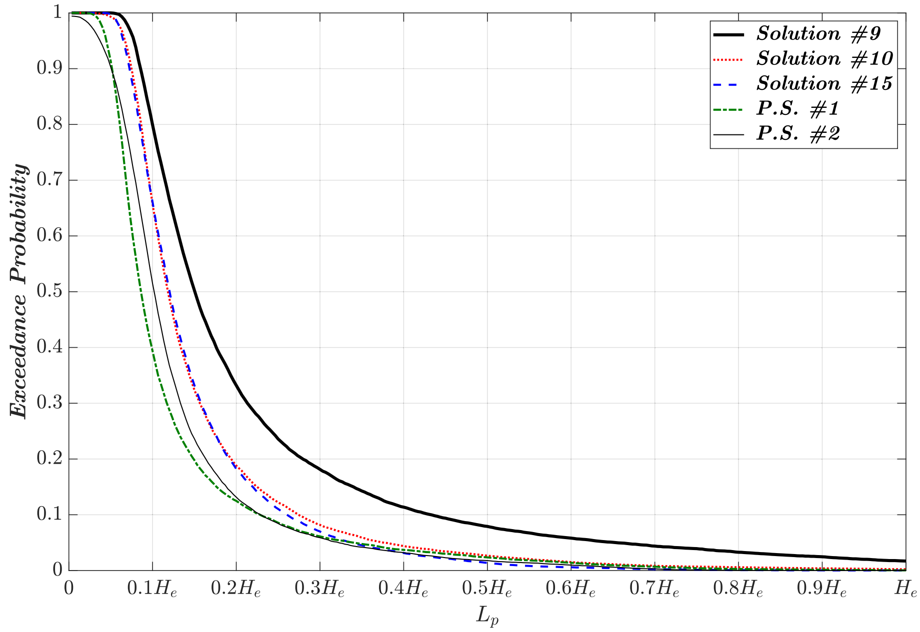

Fig. 18. Exceedance probability variation of the shortlisted solutions for different L_p^T values.

exceedance probability than proposed solutions #1 and #2 but is lower than Solution #9. The results indicate that proposed solution #2 is the most effective in minimizing exceedance probability, while solution #9 is the least effective. The higher reliability of the proposed solutions compared to the empirical ones can be attributed to the inclusion of more features following feature selection and the fine-tuning of developed networks.

To provide a comprehensive overview of the collapse probability and reliability of the existing and proposed solutions at the threshold of the plastic hinge length ($H_e/3$), the reliability indexes and the exceedance probability of all solutions were estimated at the threshold and demonstrated in Fig. 19.

Fig. 19 provides a comparative analysis of the reliability index and exceedance probability for various solutions at the threshold plastic hinge length ($H_e/3$). Solutions #11, #13, and #14 show high reliability; however, they significantly underperform compared to the shortlisted solutions in overall performance (see Figs. 9, 10, 11 and 12). The greater robustness of solutions #11, #13, and #14, despite their lower performance, can be attributed to their limited upper bounds, as they were derived for lightly reinforced shear walls (see Table 1). As illustrated in Fig. 19, proposed solution #2 emerges as the most reliable option, exhibiting the highest reliability index and the lowest exceedance probability, underscoring its superior robustness in predicting plastic hinge length. Proposed solution #1

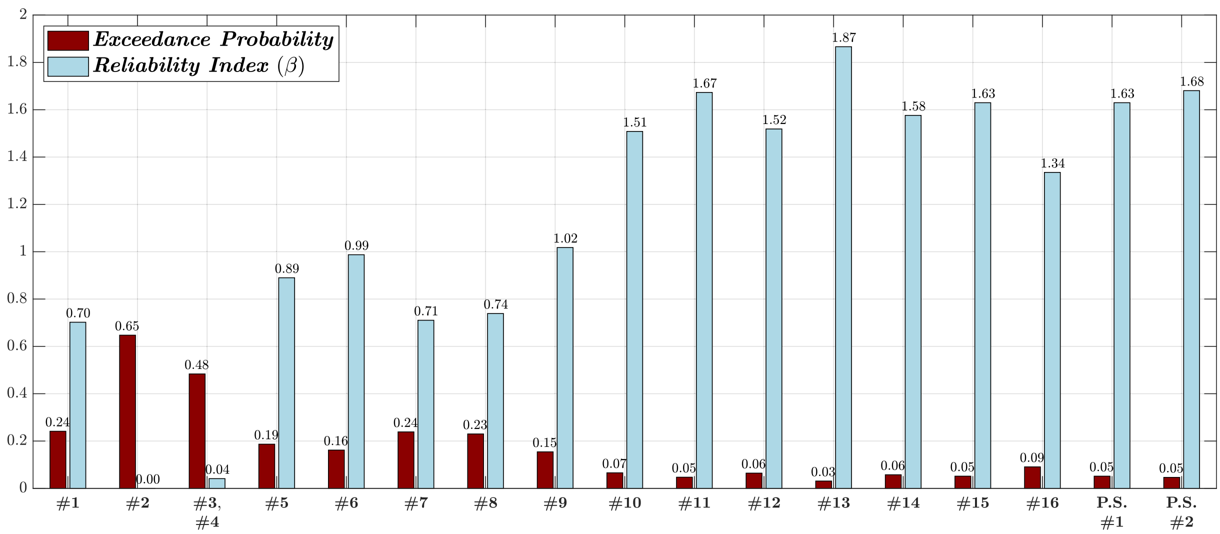

Fig. 19. Exceedance probability and reliability index of the listed solutions at the plastic hinge length threshold ($H_e/3$).

and solution #15 follow closely, showing high reliability indices and low exceedance probabilities, reflecting significant robustness, although they are not as optimal as proposed solution #2. Solution #10 demonstrates moderate performance with a mid-range reliability index and exceedance probability, suggesting it provides reasonable reliability but is less optimal than the proposed solutions. In contrast, solutions #1-#9 display considerably lower reliability indices and higher exceedance probabilities, making them less reliable and indicating a higher likelihood of failure in predicting plastic hinge length. Overall, the results emphasize the effectiveness of the proposed solutions, especially proposed solution #2, in enhancing reliability and minimizing the risk of exceedance in estimating the plastic hinge length.

4. Conclusions

This research presents a comprehensive study on the plastic hinge length of RC shear walls using ML approach and probabilistic assessment. The main objectives were to derive accurate and universal closed-form solutions for plastic hinge length and to evaluate their reliability under stochastic conditions. The study involved collecting an extensive database of experimental and numerical results, conducting feature selection analysis, developing ML models to predict plastic hinge length, and performing a reliability analysis of the derived solutions. The performance and robustness of the derived solutions were thoroughly compared to existing empirical solutions. Based on the detailed investigation of both the proposed and empirical solutions, the following conclusions are identified:

- The wall length, shear span, axial load ratio, longitudinal and transverse reinforcement ratios, concrete compressive strength, yield and ultimate strength of reinforcements, and peak lateral capacity are the most influential features in estimating the plastic hinge length of shear walls.
- Two closed-form solutions for plastic hinge length were proposed using NNs with specialized architectures: one with a single neuron and another with two neurons in the hidden layer, employing the Log-Sigmoid activation function. The first solution, utilizing one neuron, is practical and straightforward, while the second, more complex solution aims to enhance performance.
- The proposed solutions show significant superiority over empirical ones, demonstrated by their lower root mean square error values. Proposed solution #1, despite underestimating plastic hinge length with a minor mean error, proves conservative for assessing RC shear walls' deformation capacity. On the other hand, proposed solution #2 slightly overestimates plastic hinge length; however, its overall superior performance renders these overestimations negligible. Moreover, the proposed solutions exhibit much stronger performance ($R^2 > 0.8$) in explaining and predicting plastic hinge length compared to empirical solutions.
- SHAP analysis revealed that wall length, shear span, concrete compressive strength, and axial load ratio are the top contributing features that increase plastic hinge length. Overall, wall length and shear span are the predominant factors driving model predictions.
- Given the dominant role of shear span on the plastic hinge length, the plastic hinge threshold was proportional to the shear span. Considering the collected experimental data, the plastic hinge length threshold set to $H_e/3$ to define the limit state function for conducting the reliability analysis.
- Both proposed solutions exhibit a low exceedance probability (with the second being lower), indicating higher reliability than empirical solutions. This also demonstrates that proposed solution #2 is the most effective in minimizing exceedance probability, likely due to fine-tuning the model network and the higher number of neurons in its hidden layer. Furthermore, the proposed solutions show a larger reliability index than the empirical ones.

CRediT authorship contribution statement

Abouzar Jafari: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Masoomeh Mirrashid:** Writing – review & editing, Validation, Methodology, Investigation, Formal analysis. **Ryan D. Hoult:** Data curation, Writing – review & editing, Methodology. **Ying Zhou:** Writing – review & editing, Validation, Supervision.

Declaration of competing interest

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

Acknowledgments

This research was supported by the National Natural Science Foundation of China (NSFC) under Grant No. 52350410464.

Data availability

The complete database is publicly available and can be downloaded from the following link: https://zenodo.org/records/5513710.

References

[1] L. Xue, G. He, Q. Sun, Failure analysis of monolithic precast shear wall structure subjected to strong earthquake, Eng. Fail. Anal. 146 (2023) 107078.
[2] A. Koçak, The effects of infill walls and basement shear walls on the seismic performance of existing damaged building, Eng. Fail. Anal. 118 (2020) 104797.
[3] T.A. Ozkul, et al., Effect of shear wall on seismic performance of RC frame buildings, Eng. Fail. Anal. 100 (2019) 60–75.
[4] L. Miao, et al., Effect of shear-span ratio and vertical reinforcement ratio on the failure of geometrical-similar RC shear walls, Eng. Fail. Anal. 139 (2022) 106407.
[5] R. Park, T. Paulay, Reinforced Concrete Structures, John Wiley & Sons, New York, US, 1975.
[6] T. Paulay, The Design of ductile reinforced concrete structural walls for earthquake resistance, Earthq. Spectra 2 (4) (1986) 783–823.
[7] J.P. Moehle, Displacement-based design of rc structures subjected to earthquakes, Earthq. Spectra 8 (3) (1992) 403–428.
[8] J.W. Wallace, J.P. Moehle, Ductility and detailing requirements of bearing wall buildings, J. Struct. Eng. 118 (6) (1992) 1625–1644.
[9] ACI, *Building code requirements for reinforced concrete (ACI 318-89)*. 1989, American Concrete Institute: Farmington Hills, Michigan, USA.
[10] ICBO, *Uniform Building Code (UBC 97)*. 1997, International Conference of Building Officials (ICBO): Whittier,California, USA.
[11] CSA, *Design of concrete structures (CAN CSA A23. 3-04)*. 2004, Canadian Standards Association: Toronto, ON, Canada.
[12] NZS, *Concrete Structures Standard – Part 1: The Design of Concrete Structures (NZS 3101.1:2006)*. 2006, Council of Standards New Zealand: Wellington, New Zealand.
[13] İ. Kazaz, Analytical study on plastic hinge length of structural walls, J. Struct. Eng. 139 (11) (2013) 1938–1950.
[14] Bohl, A.G., *Plastic hinge length in high-rise concrete shear walls*, in *Civil Engineering*. 2006, University of British Columbia: Vancouver, BC, Canada.
[15] R. Hoult, H. Goldsworthy, E. Lumantarna, Plastic hinge length for lightly reinforced rectangular concrete walls, J. Earthq. Eng. 22 (8) (2018) 1447–1478.
[16] Y. Lu, et al., Cyclic testing of reinforced concrete walls with distributed minimum vertical reinforcement, J. Struct. Eng. 143 (5) (2017) 04016225.
[17] R.D. Hoult, H.M. Goldsworthy, E. Lumantarna, Plastic hinge length for lightly reinforced C-Shaped Concrete Walls, J. Earthq. Eng. 24 (7) (2020) 1083–1114.
[18] Constantin, R.-T., *Seismic behaviour and analysis of U-shaped RC walls*. 2016, EPFL: Lausanne, Switzerland. p. 296.
[19] A.G. Bohl, P. Adebar, Plastic hinge lengths in high-rise concrete shear walls, ACI Struct. J. 108 (2) (2011).
[20] Priestley, M., G. Calvi, and M. Kowalsky, *Displacement–Based Seismic Design of Structures*. 2007, Pavia, Italy: IUSS.
[21] J.H. Thomsen, J.W. Wallace, Displacement-based design of slender reinforced concrete structural walls-experimental verification, J. Struct. Eng. 130 (4) (2004) 618–630.
[22] A. Wibowo, et al., Seismic performance of lightly reinforced structural walls for design purposes, Mag. Concr. Res. 65 (13) (2013) 809–828.
[23] Altheeb, A.H., *Seismic drift capacity of lightly reinforced concrete shear walls*, in *Department of Infrastructure Engineering*. 2016, University of Melbourne: Melbourne, Australia.
[24] A. Jafari, et al., Plastic hinge length for coupled and hybrid-coupled shear walls, Steel Compos. Struct. 48 (4) (2023) 367.
[25] D.R. Hoult, Seismic Assessment of Reinforced Concrete Walls in Australia, in: Department of Infrastructure Engineering, University of Melbourne, Melbourne, Australia, 2017.
[26] K. Beyer, A. Dazio, M.J.N. Priestley, Inelastic wide-column models for U-shaped reinforced concrete walls, J. Earthq. Eng. 12 (sup1) (2008) 1–33.
[27] R.D. Hoult, Universal plastic hinge length for reinforced concrete walls, ACI Struct. J. 119 (4) (2022).
[28] M.S. Barkhordari, A. Jawdhari, Machine learning based prediction model for plastic hinge length calculation of reinforced concrete structural walls, Adv. Struct. Eng. 26 (9) (2023) 1714–1734.
[29] T. Paulay, S.M. Uzumeri, A Critical review of the seismic design provisions for ductile shear walls of the canadian code and commentary, Can. J. Civ. Eng. 2 (4) (1975) 592–601.
[30] R.G. Oesterle, et al., Web Crushing of reinforced concrete structural walls, ACI Journal Proceedings 81 (3) (1984).
[31] T. Paulay, M.J.N. Priestley, Seismic design of reinforced concrete and masonry buildings, Wiley, USA, 1992.
[32] T. Paulay, M.J.N. Priestley, Stability of ductile structural walls, ACI Struct. J. 90 (4) (1993).
[33] M. Sasani, A.D. Kiureghian, Seismic Fragility of RC structural walls: displacement approach, J. Struct. Eng. 127 (2) (2001) 219–228.
[34] 1998-3:2005, E., *Eurocode 8: Design of Structures for Earthquake Resistance – Part 3: Assessment and Retrofitting of Buildings*, in *Comité Européen de Normalisation: Brussels, Belgique*. 2005, European Committee for Standardization: Brussels, Belgium. p. 91.
[35] Daelemans, W., *Machine Learning Approaches*, in *Syntactic Wordclass Tagging*, H. van Halteren, Editor. 1999, Springer Netherlands: Dordrecht. p. 285-304.
[36] H. Adeli, Neural Networks in Civil Engineering: 1989–2000, Comput. Aided Civ. Inf. Eng. 16 (2) (2001) 126–142.
[37] R.D. Hoult, H.M. Goldsworthy, E. Lumantarna, Plastic hinge analysis for lightly reinforced and unconfined concrete structural walls, Bull. Earthq. Eng. 16 (10) (2018) 4825–4860.
[38] Lundberg, S.M. and S.-I. Lee. A unified approach to interpreting model predictions. in Advances in neural information processing systems. 2017. Long Beach, CA, USA.
[39] J.B. Mander, M.J. Priestley, R. Park, Theoretical stress-strain model for confined concrete, J. Struct. Eng. 114 (8) (1988) 1804–1826.
[40] Ramberg, W. and W.R. Osgood, *Description of stress-strain curves by three parameters*. 1943, National Bureau of Standards: University of Washington, Seattle, WA, US.