computers

MDPI

Article

New Predictive Models for the Computation of Reinforced Concrete Columns Shear Strength

Anthos I. Ioannou [1,*], David Galbraith [2], Nikolaos Bakas [3], George Markou [4,*] and John Bellos [1]

[1] Department of Civil Engineering, Neapolis University Pafos, 2 Danais Avenue, Pafos 8042, Cyprus
[2] Department of Civil Engineering, University of Pretoria, Private Bag X20, 002 Hatfield, Pretoria 0028, South Africa; u19027436@tuks.co.za
[3] Machine Intelligence Research and Engineering P.C., 1 Tsakasianou, 11141 Athens, Greece; nb@machine-intelligence.ai
[4] Department of Civil Engineering and Geomatics, Cyprus University of Technology, Limassol 3036, Cyprus
* Correspondence: a.ioannou.7@nup.ac.cy (A.I.I.); george.markou@cut.ac.cy (G.M.)

Abstract: The assessment methods for estimating the behavior of the complex mechanics of reinforced concrete (RC) structural elements were primarily based on experimental investigation, followed by the collective evaluation of experimental databases from the available literature. There is still a lot of uncertainty in relation to the strength and deformability criteria that have been derived from tests due to the differences in the experimental test setups of the individual research studies that are being fed into the databases used to derive predictive models. This research work focuses on structural elements that exhibit pronounced strength degradation with plastic deformation and brittle failure characteristics. The study's focus is on evaluating existing models that predict the shear strength of RC columns, which take into account important factors including the structural element's ductility and axial load, as well as the contributions of specific resistance mechanisms like that of concrete, transverse, and longitudinal reinforcement. Significantly improved predictive models are proposed herein through the implementation of machine learning (ML) algorithms on refined datasets. Three ML models, LREGR, POLYREG-HYT, and XGBoost-HYT-CV, were used to develop different predictive models that were able to compute the shear strength of RC columns. According to the numerical findings, POLYREG-HYT- and XGBoost-HYT-CV-derived models outperformed other ML models in predicting the shear strength of rectangular RC columns with the correlation coefficient having a value R greater than 99% and minimal errors. It was also found that the newly proposed predictive model derived a 2-fold improvement in terms of the correlation coefficient compared to the best available equation in international literature.

Keywords: seismic assessment; reinforced concrete columns; shear strength; machine learning; design equations

Academic Editor: Kartik B. Ariyur

Received: 30 September 2024
Revised: 2 November 2024
Accepted: 26 November 2024
Published: 24 December 2024

Citation: Ioannou, A.I.; Galbraith, D.; Bakas, N.; Markou, G.; Bellos, J. New Predictive Models for the Computation of Reinforced Concrete Columns Shear Strength. *Computers* **2025**, *14*, 2. https://doi.org/10.3390/computers14010002

1. Introduction

Following the development of Performance-Based Assessment (PBA) frameworks, between 1995 and 2010, practical evaluation of the seismic behavior of reinforced concrete (RC) structures became a priority when accounting for a large number of existing RC buildings in urban centers. Earthquakes that occurred in the last 30 years affected urban regions such as Loma-Prieta (1989), Northridge (1994, California), Athens (1999, Parnitha), Izmit (1999, Turkey), ChiChi (1999), L'Aquila (2009, Italy), Haiti (2010), Kahramanmaras (2023, Turkey), etc., all highlighting the catastrophic potential and risk to human life imparted by old construction. Damage was more intense in RC buildings with soft storeys.

For the first time in this period, the explicit interest in the literature is focused on the drift capacity of columns at collapse. A large number of studies have been published thus far, attempting to quantify the deformation capacity of columns, with reference to the seismic risk presented by existing construction. In particular, crucial parameters that affect the seismic behavior of this type of element at advanced stages of deformation are of significant importance, especially when elements with inadequate steel reinforcing configurations that represent old-type practices are evaluated.

An important factor responsible for the dispersion of results is the perceived insensitivity of the analytical models to some critical parameters that control the onset of failure. In columns controlled by flexural yielding before failure (flexure–shear elements), the load-carrying capacity against horizontal load is generally controlled by flexure. On the other hand, the deformation capacity is generally much lower than that specified by the analytical models, which superimpose a limiting envelope on the resistance curve. This approach aims to effectively control the interplay between shear and flexure after yielding, based on the ratio $V_n(\mu)/V_{flex}$ by limiting drift capacity, which is suggested in KAN.EPE., 2014 [1]. Flexural strength is hardly the only controlling variable; for example, strength loss in lap splices exacerbated by cyclic deformation reversals may alter the hierarchy expressed by the preceding ratio. For these reasons, the estimation of the shear capacity of RC columns is still an open problem that requires further investigation and the development of more accurate design formulae.

For the purposes of this study, a dataset was obtained from the PEER database of columns by Berry et al. (2004) [2], which included an extensive collection of tests gathered from published experimental studies. The type of failure of the RC column specimens was thoroughly examined and served as the primary criterion for inclusion in the selected dataset. The research motivation centers on examining the accuracy of the most promising proposed models for evaluating the shear strength of RC columns, as outlined herein and in previous publications found in international literature. The collected data were subsequently used to compare these models' shear strength calculations with reported values, enabling an assessment of their predictive accuracy.

Many studies have evaluated existing predictive models and introduced new ones to improve the accuracy of predicting shear strength and failure [1–10]. Key factors that affect shear strength include column dimensions, concrete strength, aspect ratio, axial load, and displacement ductility [7]. Comparisons between various models and experimental data have shown that some models perform better than others, with the Eurocode 8 (2005) [11] and Sezen and Moehle (2004) [7] models demonstrating high accuracy [10]. To validate the shear mechanism of the proposed model, Jiang et al. (2023) [12] thoroughly examined the contribution of each shear component in relation to four existing models: Sezen and Moehle (2004) [7], ACI (2011) [13], Priestley et al. (1996) [4], and MOHURD (2010) [14]. While the Sezen [7] and ACI (2011) [13] models acknowledge the shear contributions of concrete and stirrups, the Priestley et al. (1996) [4] and MOHURD (2010) [14] models also consider the positive impact of the axial compression force component. A comparative analysis highlights the superior predictive ability of the newly proposed model. Shear strength envelope models, which take into account strength degradation under cyclic loading, have been particularly effective in simulating column behavior. However, further research is needed to deepen our understanding of how axial load and shear span affect cyclic behavior.

According to Sasani (2007) [8], experimental results from 89 reinforced concrete (RC) columns, along with the mechanics of shear transfer and column deformation, were used to develop new models for shear strength and drift ratio capacity. The proposed shear strength model demonstrates a coefficient of variation of just 0.14, offering significantly

improved predictions of column shear strength compared to current codes and standard equations. The study further indicates that predictions improve when approximately two-thirds of the strength provided by transverse reinforcement, based on a simple truss model, is considered effective.

Yu et al. (2019) [15] similarly proposed a probabilistic shear strength model for RC columns, also leveraging a variable-angle truss-arch model with MCMC methods. Their approach includes a probabilistic calibration method to assess and validate the accuracy of existing shear strength models. Ma et al. (2024) [16] introduce a probabilistic model for shear capacity in RC columns based on truss-arch theory and Monte Carlo Markov Chain (MCMC) methods. This model captures the influence of middle longitudinal reinforcement, validated by a confidence interval to demonstrate its precision and supplemented by an analysis of each shear component's contributions.

Recent research has investigated the use of machine learning (ML) techniques for the seismic evaluation of reinforced concrete (RC) structures. These methods offer faster and more efficient alternatives to traditional approaches by predicting building performance, seismic responses, and damage states with high accuracy. Models such as Random Forest, Artificial Neural Networks, and Extreme Gradient Boosting have consistently demonstrated excellent predictive capabilities [17–19]. Critical factors affecting seismic performance include the age of the building, concrete strength, and steel properties [17]. ML models have been successfully applied to estimate parameters such as the maximum interstorey drift ratio, seismic performance level and fragility curves [18]. Moreover, Active Machine Learning has proven effective in assessing damage states, even with limited datasets [19]. Additionally, user-friendly graphical interfaces have been developed to make ML-based seismic assessment tools accessible for practical use [18,19].

Thai (2022) [17] highlights machine learning (ML) as a transformative tool in structural engineering, offering the potential to replace conventional empirical models. Thai's extensive review of ML in structural engineering encompasses five distinct topics and seven groups of ML algorithms. Additionally, Thai provides resources, such as open-source ML codes, libraries, and datasets, to facilitate ML model development for structural engineers. Key challenges and potential directions for ML in structural engineering are also discussed.

Phan et al. (2022) [20] demonstrated that Artificial Neural Networks (ANNs) and K-Nearest Neighbors (KNN) models effectively predict shear strength in rectangular RC columns with a high R^2 value (>0.99). The KNN model proved particularly adept at classifying failure modes with nearly 100% accuracy. A graphical user interface (GUI) tool developed in this study allows practical prediction of shear strength and failure mode identification during design and structural evaluation processes.

Design codes like ASCE/SEI 41-13 [21], EN1998-3 [11], 2005, and Pardalopoulos et al. (2013) [22] all specify useful formulae for determining the shear strength of RC columns. Comparing experimental tests and equations between the aforementioned criteria and proposed models, however, revealed a discrepancy [11,21,22]. Additionally, a number of formulae that had been suggested in earlier research had limitations in particular situations, such as short, rectangular RC columns with light transverse reinforcement [23]. Due to their numerous coefficients, only a few models were even challenging in application techniques [24]. The existing models that are used to assess the shear strength of RC columns were supplemented and improved using an optimization methodology in order to enhance the investigation. The proposed model, which was created using machine learning methods, is contrasted with the models proposed by ASCE/SEI 41-13 [21], EN1998-3, 2005 [11], and Pardalopoulos et al. (2013) [22].

2. Crucial Behavior Parameters

A qualifying criterion for the type of failure and the behavior of RC columns is the yielding of the longitudinal reinforcement before the occurrence of shear failure. (Shear failure is identified by transverse reinforcement yielding precluding other inadequacies such as lap-splice failure). If longitudinal bar yielding precedes stirrup yielding, the failure is described as being of flexure–shear type, whereas if the sequence is reversed, the failure is referred to as brittle–shear. For the brittle–shear type of failure, the drift capacity is particularly small and in any case is less than the nominal yielding drift θ_y of the element.

Because flexural yielding is the benchmark reference in assessment procedures, the drift at the onset of longitudinal bar yielding is needed to define all other forms of failure by comparison. Thus, θ_y is calculated either by member analysis (i.e., Response 2000), or according to the Greek Code for Interventions KANEPE, 2017 [1] and the EN 1998-3, 2005 [11]. Alternatively, mechanics-based relationships may be used (i.e., $\theta_y = \frac{1}{3}.\varphi_y.L_s$, where $\varphi_y = min\left\{2.1\frac{\varepsilon_y}{h}, \frac{0.75\varepsilon_{c0}}{\chi}\right\}$, χ is the height of the compression zone that equals to ξd, L_s is the shear span, and h is the cross-sectional height). The values of drift where brittle failures are developed are obtained from $\theta = \theta_y \frac{V_n}{V_{flex}}$, where V_n is the member shear strength and V_{flex} is the flexural strength calculated from the yielding moment of the critical section divided by the shear span [25]. It is noted that in the case of shear failure before flexural yielding, the ratio $\frac{V_n}{V_{flex}} \leq 1$. The shear strength V_n for the needs of seismic assessment is obtained from EN1998-3, 2005 [11], Pardalopoulos et al. (2013) [22], and ASCE/SEI 41-13 [21], and by using mean values for material strengths as shown in Table 1.

Table 1. Typical shear strength models for shear critical RC columns.

References	*Description of Models*
EN1998-3, 2005 [11]	$V_n^E = V_c + V_w + V_N = \eta^E(\mu)\cdot 0.16\sqrt{f_c}(0.8A_c)[\max\{0.5, 100\rho_{tot}\}]\cdot(1 - 0.16\cdot\min\left\{5, \frac{L_s}{h}\right\})] + \eta^E(\mu)\frac{A_{st}f_{ft}(d-d')}{s} + \min\{N, 0.55A_cf_c'\}\cdot\tan\alpha$ Where $\eta^E(\mu) = \left(1 - 0.05\ \min\left(5, \mu_\Delta^{pl}\right)\right)$
Pardalopoulos et al. *(2013)* [22]	$V_n = 0.4\xi bd\sqrt{f_c'} + \lambda\cdot\nu\cdot(bdf_c)\tan\alpha + A_{st}f_{yt}\frac{d(1-\xi)}{s}\cdot\cot\theta_v$ $\theta_v = 45^\circ$ for $\nu \leq 0.10$ $\theta_v = 45^\circ - 15^\circ.\frac{\nu}{0.25} \geq 30^\circ$ for $\nu \geq 0.$
ASCE/SEI 41-13 (2014) [21]	$V_n = \eta(\mu\_\Delta)\cdot(V_s + V_c) = \eta(\mu\_\Delta)\left[0.5\cdot\sqrt{f_c'}\sqrt{1 + \frac{N_u}{0.5\cdot\sqrt{f_c'}\cdot A_g}}\right]\frac{0.8\cdot A_g}{\frac{L_s}{d}} + \eta(\mu\_\Delta)\frac{A_{st}\cdot f_{yt}\cdot d}{s}$ $0.7 \leq \eta(\mu\_\Delta) \leq 1$

Ioannou et al., 2018 [26] considered the EN1998-3 2005 [11] expressions and the model of Pardalopoulos et al. (2013) [22] with some additional modifications that had been proposed earlier. First, the concrete contribution term is taken as per the Tureyen and Frosh (2003) [27] approach where the underlying assumption is that shear transfer in cracked sections under cyclic loading occurs mainly in the compression zone of the member. As for the second, the axial load contribution is accounted for, provided that the axial load ratio exceeds a lower limit of 0.1 (the sign convention used is compression positive). This is performed to ensure that cracks in the compression zone are closed and that the force component transferred may be non-trivial. In the case of the third, the number of stirrups activated in the V_w term is determined from the inclination of the critical crack plane measured with reference to the longitudinal axis, θ_v, which is usually not taken as 45° as it is assumed from previous available models in the literature [$cot\theta_v$ = 1], but rather, it depends on the axial load magnitude acting on the section (Figure 1b).

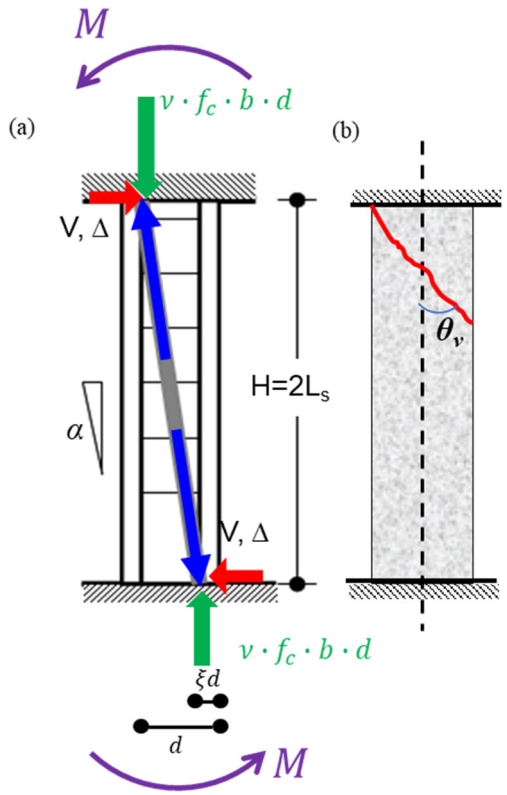

Figure 1. EN model for the contribution of the diagonal strut to shear strength. (**a**) Definition of the strut angle, and (**b**) definition of the critical crack angle, θ_v.

Relevant studies that have applied the abovementioned framework have been conducted by Aschheim and Moehle, 1992 [3]; Lynn and Moehle, 1996 [5]; Sezen and Moehle, 2004 [7]; Elwood and Moehle, 2003 [6]; and Priestley et al., 1996 [4]. Regarding the relationship proposed by Pardalopoulos et al. (2013) [22], for consistency with fundamental principles, a correction was proposed at a later stage where the effective concrete section area contributing to the shear strength at the critical section is equal to the area of compression zone, and therefore the $0.5\sqrt{f_c'} \cdot (0.8A_c)$ is replaced with $0.4\sqrt{f_c'} \cdot (\chi \cdot b)$ [27]—at this point, the compression zone height is taken equal to $\chi = \xi d$ [1].

Consequently, for the model considered, the contribution of the transverse reinforcement (V_w) is taken into consideration as long as it has been confirmed that at least one stirrup meets the critical sliding plane. Finally, all terms are reduced with increasing displacement ductility through a postulated reduction factor $0.7 \leq \eta(\mu_\Delta) \leq 1$.

3. Dataset Collection

The dataset employed in the current investigation was composed of 74 specimens featuring old-type steel reinforcing setups. The reporting of a pure shear or flexure–shear failure, which meant that longitudinal reinforcement yielding occurred before the observed shear failure, was the criterion for evaluating the specimen's selection. This is a consequence of the circumstance that in that case, a reduction in shear strength was more prominent when the resistance envelope curve's maximum shear force value was exceeded during a corresponding increase in drift. More specifically, 44 specimens showed a flexure–shear type of failure, whereas 30 specimens failed in shear. Axial failure and drift during shear might also be more clearly identified. A few of the specimens were taken from databases that have been published and made accessible by Berry et al. (2004) [2] and Kim et al. (2019) [28]. The data collected are shown in Table 2.

Table 2. Basic parameters and experimental shear strengths of 74 sets of shear critical columns.

	Reference	Test	L_s (mm)	b (mm)	h (mm)	c (mm)	d_{bw} (mm)	s (mm)	ρ_w (%)	ρ_{tot} (%)	f_{yl} (MPa)	f_{yw} (MPa)	$f_c^{'}$ (MPa)	P (kN)	$V_{exp.max}$ (kN)
1	Lynn et al. (1996)	3SLH18	1473.0	457.0	2946.0	38.1	9.5	457.0	0.1	3.0	331.0	400.0	25.6	503.0	270.0
2	Lynn et al. (1996)	3CMD12	1473.0	457.0	2946.0	38.1	9.5	305.0	0.2	3.0	331.0	400.0	27.7	1512.0	355.0
3	Lynn et al. (1996)	3CMH18	1473.0	457.0	2946.0	38.1	9.5	457.0	0.1	3.0	331.0	400.0	27.7	1512.0	328.0
4	Henkhaus et al., 2013	B1	736.5	457.0	1473.0	35.0	9.5	457.0	0.1	1.5	455.0	490.0	20.0	1545.5	565.5
5	Henkhaus et al., 2013	B2	736.5	457.0	1473.0	35.0	6.4	203.0	0.1	1.5	455.0	455.0	19.3	1531.7	317.2
6	Henkhaus et al., 2013	B3	736.5	457.0	1473.0	35.0	9.5	457.0	0.1	1.5	455.0	490.0	22.1	969.3	562.3
7	Henkhaus et al., 2013	B4	736.5	457.0	1473.0	35.0	9.5	457.0	0.1	2.5	441.0	490.0	24.1	2164.3	771.2
8	Henkhaus et al., 2013	B5	736.5	457.0	1473.0	35.0	9.5	457.0	0.1	2.5	441.0	490.0	23.4	2248.1	699.4
9	Henkhaus et al., 2013	B6	1473.0	457.0	2946.0	35.0	9.5	305.0	0.2	2.5	490.0	469.0	27.6	634.1	334.2
10	Henkhaus et al., 2013	B7	1473.0	457.0	2946.0	35.0	9.5	305.0	0.2	2.5	490.0	469.0	28.3	650.1	331.7
11	Henkhaus et al., 2013	B8	1473.0	457.0	2946.0	35.0	9.5	305.0	0.1	2.5	490.0	469.0	29.0	666.2	336.7
12	Zhou et al., 1987	104-08	160.0	160.0	320.0	12.5	5.0	40.0	0.7	2.2	341.0	559.0	19.8	406.0	82.7
13	Zhou et al., 1987	114-08	160.0	160.0	320.0	12.5	5.0	40.0	0.7	2.2	341.0	559.0	28.8	406.0	91.3
14	Kim et al., 2018	SBd2	1200.0	400.0	1200.0	40.0	12.7	165.0	0.4	2.5	571.0	500.0	32.0	870.4	340.2
15	Kim et al., 2018	SBd4	1200.0	400.0	1200.0	40.0	12.7	82.0	0.8	2.5	571.0	500.0	32.0	870.4	326.9
16	Kim et al., 2018	SCd2	1200.0	400.0	1200.0	40.0	12.7	165.0	0.4	2.5	571.0	500.0	32.0	870.4	325.4
17	Kim et al., 2018	SDd2	1200.0	400.0	1200.0	40.0	12.7	165.0	0.4	2.5	571.0	500.0	32.0	512.0	319.0
18	Kim et al., 2018	SDd4	1200.0	400.0	1200.0	40.0	12.7	82.0	0.8	2.5	571.0	500.0	32.0	870.4	326.9
19	Kim et al., 2018	RFd2	1200.0	250.0	1200.0	40.0	9.5	105.0	0.8	2.4	566.0	530.0	32.0	870.4	220.2
20	Nagasaka 1982	HPRC10-63	300.0	200.0	600.0	12.0	5.5	35.0	0.7	1.3	371.0	344.0	21.0	146.9	86.9
21	Arakawa et al., 1989	OA2	225.0	180.0	450.0	10.0	4.0	64.3	0.2	3.1	340.0	249.0	31.8	189.6	130.6
22	Arakawa et al., 1989	OA5	225.0	180.0	450.0	10.0	4.0	64.3	0.2	3.1	340.0	249.0	33.0	475.8	134.0
23	Umehara and Jirsa 1982	CUS	455.0	410.0	910.0	25.0	6.0	89.0	0.3	3.0	441.0	414.0	34.9	533.2	322.7
24	Umehara and Jirsa 1982	CUW	455.0	230.0	910.0	25.0	6.0	56.0	0.3	3.0	441.0	414.0	34.9	533.2	255.3
25	Umehara and Jirsa 1982	2CUS	410.0	410.0	230.0	25.0	6.0	89.0	0.3	3.0	441.0	414.0	42.0	1069.4	409.4
26	Bet et al., 1985	1_1	305.0	305.0	305.0	25.0	6.0	210.0	0.2	2.4	462.0	414.0	29.9	289.3	213.6
27	Aboutaha et al., 1999	SC3	457.2	457.2	914.4	38.0	9.5	406.4	0.1	1.9	434.0	400.0	21.9	0.0	407.0

Table 2. *Cont.*

	Reference	Test	L_s (mm)	b (mm)	h (mm)	c (mm)	d_{bw} (mm)	s (mm)	ρ_w (%)	ρ_{tot} (%)	f_{yl} (MPa)	f_{yw} (MPa)	$f_c^{'}$ (MPa)	P (kN)	$V_{exp.max}$ (kN)
28	Aboutaha et al., 1999	SC9	914.4	914.4	457.2	38.0	9.5	406.4	0.1	1.9	434.0	400.0	16.0	0.0	604.5
29	Sokoli and Ghannoum (2016)	CS-60	457.2	457.2	457.2	38.1	16.0	140.0	1.5	4.7	464.0	472.0	26.4	290.0	779.6
30	Sokoli and Ghannoum (2016)	CS-100	457.2	457.2	457.2	38.1	10.0	114.0	0.7	2.9	700.0	820.0	32.0	1806.0	734.6
31	Lynn et al., (1996)	2CLH18	457.0	457.0	457.0	38.1	9.5	457.0	0.1	2.0	331.0	400.0	33.1	503.0	240.8
32	Lynn et al., (1996)	3CLH18	457.0	457.0	457.0	38.1	9.5	457.0	0.1	3.0	331.0	400.0	25.6	503.0	277.0
33	Lynn et al., (1996)	2SLH18	457.0	457.0	457.0	38.1	9.5	457.0	0.1	2.0	331.0	400.0	33.1	503.0	229.0
34	Lynn et al., (1996) (lap splice)	2CMH18	457.0	457.0	457.0	38.1	9.5	457.0	0.1	3.0	331.0	400.0	25.7	1512.0	306.0
35	Lynn et al., (1996) (lap splice)	3SMD12	457.0	457.0	457.0	38.1	9.5	305.0	0.2	3.0	331.0	400.0	25.7	1512.0	367.0
36	Matchulat et al., 2005	Sp.1	457.0	457.0	457.0	39.7	9.5	460.0	0.3	2.5	441.3	372.3	20.7	2159.5	414.0
37	Matchulat et al., 2005	Sp.2	457.0	457.0	457.0	39.7	9.5	460.0	0.3	2.5	441.3	372.3	23.4	1663.0	363.2
38	Sezen and Moehle 2002	Specimen 1	457.0	457.0	457.0	65.1	9.5	304.8	0.2	2.5	434.4	476.0	21.1	665.4	302.5
39	Sezen and Moehle 2002	Specimen 2	457.0	457.0	457.0	65.1	9.5	304.8	0.2	2.5	434.4	476.0	21.1	2666.1	301.0
40	Sezen and Moehle 2002	Specimen 4	457.0	457.0	457.0	65.1	9.5	304.8	0.2	2.5	434.4	47.0	21.8	664.7	294.6
41	Wight and Sozen 1973	40.033a(East)	152.0	152.0	305.0	32.0	6.3	127.0	0.3	2.5	496.0	345.0	34.7	188.2	98.8
42	Wight and Sozen 1973	40.033a(west)	152.0	152.0	305.0	32.0	6.3	127.0	0.3	2.5	496.0	345.0	34.7	188.2	101.3
43	Wight and Sozen 1973	40.048.(East)	152.0	152.0	305.0	32.0	6.3	89.0	0.5	2.5	496.0	345.0	26.1	177.9	104.6
44	Wight and Sozen 1973	40.048(West)	152.0	152.0	305.0	32.0	6.3	89.0	0.5	2.5	496.0	345.0	26.1	177.9	98.5
45	Wight and Sozen 1973	40.033(East)	152.0	152.0	305.0	32.0	6.3	127.0	0.3	2.5	496.0	345.0	33.6	177.6	94.2
46	Wight and Sozen 1973	40.033(West)	152.0	152.0	305.0	32.0	6.3	127.0	0.3	2.5	496.0	345.0	33.6	177.6	104.9
47	Wight and Sozen 1973	25.033(East)	152.0	152.0	305.0	32.0	6.3	127.0	0.3	2.5	496.0	345.0	33.6	110.6	87.9
48	Wight and Sozen 1973	25.033(West)	152.0	152.0	305.0	32.0	6.3	127.0	0.3	2.5	496.0	345.0	33.6	110.6	93.3
49	Wight and Sozen 1973	40.067(East)	876.0	152.0	305.0	32.0	19.0	64.0	0.7	2.5	496.0	345.0	33.4	178.1	93.1
50	Wight and Sozen 1973	40.067(West)	876.0	152.0	305.0	32.0	19.0	64.0	0.7	2.5	496.0	345.0	33.4	178.1	99.4
51	Wight and Sozen 1973	40.147(East)	875.0	152.0	305.0	32.0	19.0	64.0	1.5	2.5	496.0	317.0	33.5	178.6	119.8
52	Wight and Sozen 1973	40.147(West)	875.0	152.0	305.0	32.0	19.0	64.0	1.5	2.5	496.0	317.0	33.5	178.6	114.7
53	Wight and Sozen 1973	40.092(East)	875.0	152.0	305.0	32.0	19.0	102.0	0.9	2.5	496.0	317.0	33.5	178.6	115.9
54	Wight and Sozen 1973	40.092(West)	875.0	152.0	305.0	32.0	19.0	102.0	0.9	2.5	496.0	317.0	33.5	178.6	121.0
55	Ohue et al., 1985	2D16RS	400.0	200.0	200.0	11.0	16.0	50.0	0.6	2.0	369.0	316.0	32.0	183.0	110.6
56	Ohue et al., 1985	4D13RS	400.0	200.0	200.0	12.0	13.0	50.0	0.6	2.7	370.0	316.0	29.9	183.0	101.4
57	Zhou et al., 1987	124-08	160.0	160.0	160.0	12.5	9.5	40.0	1.8	2.2	341.0	559.0	19.8	406.0	115.0
58	Zhou et al., 1987	204-08	320.0	160.0	160.0	12.5	9.5	40.0	0.7	2.2	341.0	559.0	21.1	432.7	66.5

Table 2. *Cont.*

	Reference	Test	L_s (mm)	b (mm)	h (mm)	c (mm)	d_{bw} (mm)	s (mm)	ρ_w (%)	ρ_{tot} (%)	f_{yl} (MPa)	f_{yw} (MPa)	$f_c^{'}$ (MPa)	P (kN)	$V_{exp.max}$ (kN)
59	Zhou et al., 1987	223-09	320.0	160.0	160.0	12.5	9.5	40.0	1.8	2.2	341.0	559.0	21.1	486.1	67.4
60	Zhou et al., 1987	302-07	480.0	160.0	160.0	12.5	9.5	40.0	0.7	2.2	341.0	559.0	28.8	516.8	51.2
61	Zhou et al., 1987	312-07	480.0	160.0	160.0	12.5	9.5	40.0	0.6	2.2	341.0	559.0	28.8	516.8	54.9
62	Zhou et al., 1987	322-07	480.0	160.0	160.0	12.5	9.5	40.0	1.0	2.2	341.0	559.0	28.8	516.8	51.8
63	Amitsu et al., 1991	CB060C	323.0	278.0	278.0	28.0	28.0	52.0	0.9	4.1	441.0	414.0	46.3	2633.6	505.6
64	Xiao and Martirossyan 1998	HC4-8L16-T6-0.1P	508.0	254.0	254.0	13.0	15.9	51.0	1.6	2.5	510.0	449.0	86.0	532.6	267.6
65	Xiao and Martirossyan 1998	HC4-8L16-T6-0.2P	508.0	254.0	254.0	13.0	15.9	51.0	1.6	2.5	510.0	510.0	86.0	1065.3	324.1
66	Kim et al., 2018	Sad2	1200.0	400.0	400.0	40.0	25.4	165.0	0.4	2.5	571.0	500.0	32.0	870.4	310.3
67	Kim et al., 2018	RGd2	1200.0	250.0	640.0	40.0	22.2	105.0	0.8	2.4	566.0	530.0	32.0	870.4	240.8
68	Nagasaka 1982	HPRC19-32	300.0	200.0	200.0	12.0	12.7	20.0	1.2	1.3	371.0	344.0	21.0	294.0	110.7
69	Zhou et al., 1985	No.806	80.0	80.0	80.0	10.0	6.0	80.0	0.4	1.8	336.0	341.0	32.3	124.0	31.6
70	Zhou et al., 1985	No. 1007	80.0	80.0	80.0	10.0	6.0	80.0	0.4	1.8	336.0	341.0	34.0	152.3	36.7
71	Zhou et al., 1985	No.1309	80.0	80.0	80.0	10.0	6.0	80.0	0.4	1.8	336.0	341.0	32.8	188.9	29.6
72	Imai and Yamamoto 1986	No 1	825.0	500.0	400.0	37.0	22.0	100.0	0.3	2.7	318.0	336.0	27.1	390.2	471.3
73	Ono et al., 1989	CA025C	300.0	200.0	200.0	19.0	9.5	70.0	0.8	2.1	361.0	426.0	25.8	265.2	130.0
74	Ono et al., 1989	CA060C	300.0	200.0	200.0	19.0	9.5	70.0	0.8	2.1	361.0	426.0	25.8	635.7	133.8

Notation: L_s = element shear span; b = section width; h = section height; c = concrete cover; d_{bw} = transverse reinforcement diameter; s = transverse reinforcement spacing; ρ_w = geometrical transverse reinf. ratio in the direction of seismic action; ρ_{tot} = longitudinal reinforcement ratio; f_{yl} = longitudinal reinforcement yielding stress; f_{yw} = transverse reinforcement yielding stress; f'_c = concrete compressive strength; P = axial load; $V_{exp.max}$ = experimental shear force.

The cross-section's height h, width b, and element shear span L_s are geometric dimensions that are usually used for the development of predictive models. It is important to note that the column's aspect ratio $\frac{L_S}{d}$ varies from 1.0 to 4.8, covering short and slender RC columns. The concrete cover c, reinforcement details such as the transverse reinforcement diameter d_{bw}, the spacing of the transversal reinforcements, the longitudinal reinforcement ratio ρ_{tot}, and the transversal reinforcing bar ratio ρ_w, are additional parameters that are included in the dataset. The uniaxial compressive strength of the concrete f_c', the yield strength of the longitudinal reinforcement f_{yl}, and transverse reinforcing bars f_{yw} make up the material properties. The values of f_c' vary from 16 to 86 MPa, covering the case of high-strength concrete materials. Moreover, crucial parameters such as the axial load P (superstructural loads) and the maximum experimental shear strength $V_{exp,max}(kN)$ are considered in the dataset.

4. Machine Learning Algorithms

In a range of scientific areas, including simulation-based engineering research, the usage of ML algorithms has grown significantly in recent years [29]. ML algorithms are a very promising tool for tackling computationally challenging issues given their predictive abilities that have been demonstrated in several applications. Different techniques are used to produce answers that cannot be reached by standard numerical procedures. Some of the often-used ML numerical techniques are Random Forest, gradient boosting [30], and ANNs [31]. These models, however, are frequently too sophisticated to be effectively deployed and tested in practical situations. Improved ML approaches have recently been created in order to solve these shortcomings [32].

The present research investigation has employed the following techniques:

- Linear Regression (LRegr): The simplest regression model in ML is called linear regression, and it assumes that the relationship between the input and output variables is best described by a straight line (linear function). This method is used as a base numerical technique, and it is compared to the more advanced ML algorithms.
- Nonlinear Regression (NLRegr): In nonlinear regression, Ruckstuhl (2016) [33] states that functions with nonlinear parameters are utilized. Such a function was frequently drawn from theory. Markou et al. (2024) [29] proposed POLYRG-HYT with advanced predictive features; thus, it is adopted herein for the needs of the numerical analysis.
- Extreme Gradient Boosting (XGBoost): The XGBoost algorithm was proposed by Chen and Guestrin (2016) [34] in order to accelerate the GBM algorithm, which is currently highly difficult to implement owing to its sequential model training. XGBoost uses numerous strategies for surpassing GBM in speed. For example, the randomization technique is used to reduce overfitting and increase training speed. A compressed column-based structure is also used to store data to reduce the cost of sorting which is the most time-consuming part of tree learning. For the needs of this research work, the XGBoost-HYT-CV is adopted as presented by Markou et al. (2024) [29].

According to Markou et al. (2024) [29], the proposed tuning for the case of the XGBoost-HYT-CV is performed for the following significant training parameters:

1. Maximum number of XGBoost rounds ∈ [10, 20, . . ., 1000];
2. Maximum tree depth ∈ [1,7,15];
3. Learning rate eta ∈ [0.05, 0.2, 0.5];
4. Colsample_bytree ∈ [0.5, 1];
5. Subsample ∈ [0.5, 1].

5. Investigation of the Proposed and Existent Formulae

Figure 2 demonstrates a comparison between experimental results from the dataset presented in Table 2 and the proposed models from EN1998-3 (2005) [11], Pardalopoulos et al. (2013) [22], and ASCE/SEI 41-13 (2014) [21] for the shear strength evaluation of the understudy RC columns. The objective of the study was to propose an improved model utilizing a numerical technique as described in the next section by confirming the consistency of the under-study models in terms of experimental RC column shear or flexure–shear failures.

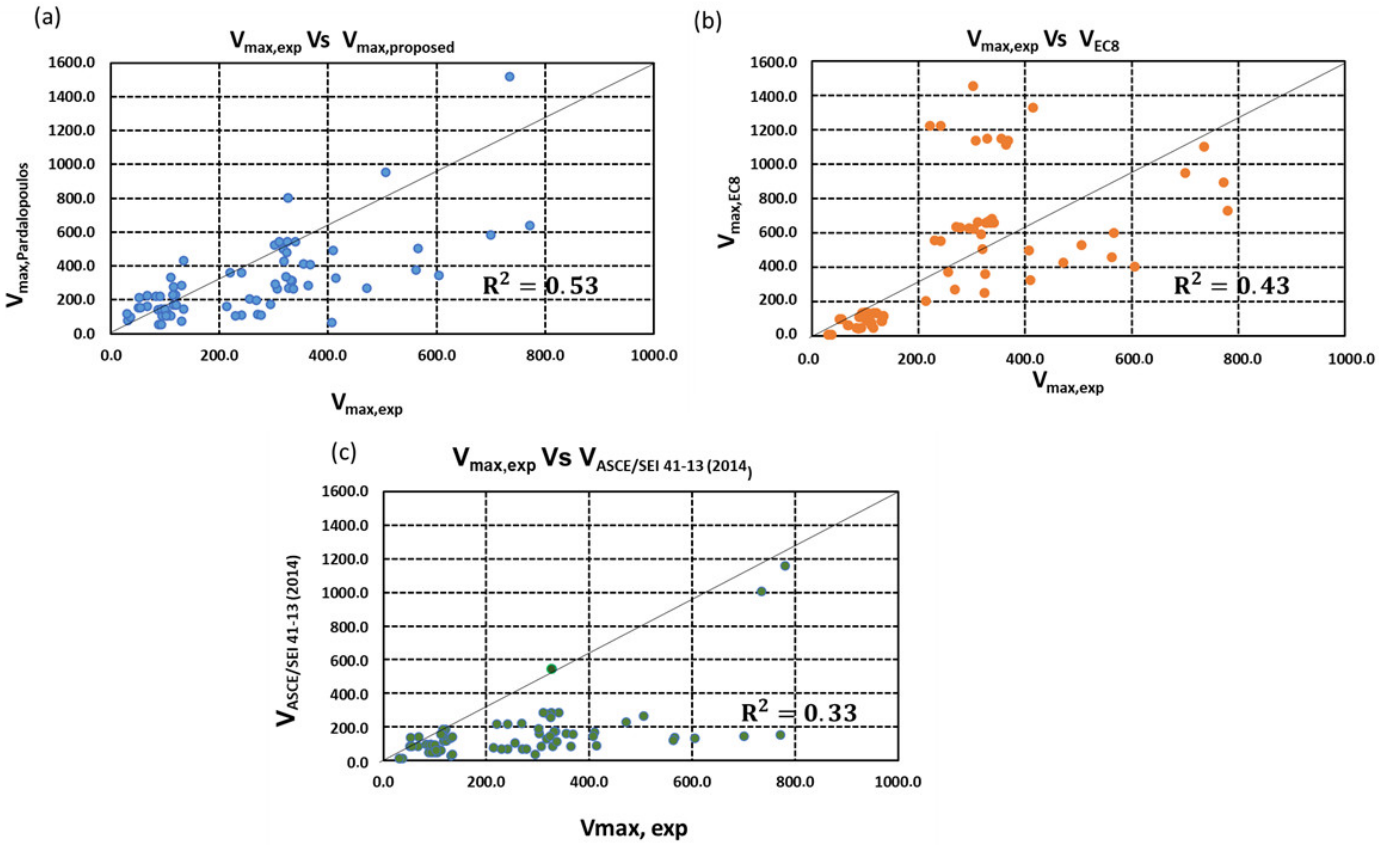

Figure 2. Correlation of experimental corrected maximum shear strength data vs. model predictions: (**a**) model proposed by Pardalopoulos et al. (2013) [22]; (**b**) model proposed by EN1998-3, 2005 [11], and (**c**) model proposed by ASCE/SEI 41-13 (2014) [21].

The 45° line in Figure 2 represents the equal value correlation, while the points above and below indicate instances where the analytical estimate differs from the corresponding experimental value. Deviations from the equal value line suggest poor predictive accuracy, whereas consistent overestimation highlights the need for a safety factor. It is evident that the Pardalopoulos et al. (2013) [22] model has a correlation coefficient of R = 0.54, which is higher than that of the current EN1998-3 (2005) [11] model (R = 0.43) and the corresponding ASCE/SEI 41-13 (2014) [21] model which derived a correlation coefficient R = 0.33. This underscores the necessity for improving maximum shear strength estimates.

5.1. Proposed Predictive Models

The basic principles of the POLYREG-HYT ML algorithm are based on the polynomial regression approach with hyperparameter tuning and are targeted at the construction of nonlinear terms composed of different combinations of independent variables up to the third degree. The approach may automatically derive the optimal predictive model while

avoiding overfitting by selecting nonlinear characteristics that correspond to the lowest prediction error.

In total, 95, 90, 85, and 80 percent of the data were used to train the POLYREG-HYT algorithm [29] by identifying relative relationships. The remaining 5, 10, 15, and 20 percent, respectively, were used to evaluate the performance of the numerically obtained shear strength equations (testing dataset). In a cross-validation context, this process was repeated five times using randomly permuted subsets of the training set. It must be noted herein that all analyses were performed through the use of the nbml software [29], which is written in Python. Given that the predictive models operate as black boxes, researchers can readily utilize the link provided (https://github.com/nbakas/nbml (accessed on 15 November 2024)) to access the nbml software and develop predictive models themselves through the open-source code.

Three datasets were utilized in the parametric investigation performed for the needs of this research work. The first matched the independent variables of Pardalopoulos et al. (2013) [22], another matched the independent variables of EN1998-3 (2005) [11], and the last matched the independent variables of ASCE/SEI 41-13 (2014) [21]. Five iterations of the POLYREG-HYT algorithm were run for every train–test split ratio. The performance of POLYREG-HYT and linear regression (LRegr) algorithms was assessed by means of a comparative study of the obtained data. The correlation coefficient (R) and the mean absolute percentage error (MAPE) were used to assess the predictions made by the predictive models.

After running the ML algorithm, three predictive models were obtained as can be seen in Equations (1)–(3), which correspond to the POLYREG-HYT. The three equations correspond to the three different datasets described above [11,21,22].

According to the numerical results, using the POLYREG-HYT method and analyzing dataset 1 (based on EN1998-3 (2005) [11] variables), we determined the optimal split ratio to be 90/10%. The highest R values achieved were 98.44% for training and 99.4% for testing, with a MAPE value of 17.7% on the test dataset. For dataset 2 (based on the Pardalopoulos et al. (2013) [22] model variables), the best split ratio was again found to be 90/10%. In this case, the training and testing R were 95.26% and 99.65%, respectively, with a MAPE of 13.9%. Finally, for dataset 3 (based on ASCE/SEI 41-13 [21], 2014 variables), the optimal split ratio was the same as before (90/10%), where the R values obtained were 99.14% for training and 99.0% for testing, with a MAPE of 15.5% same on the testing dataset.

$$
\begin{aligned}
V_{exp,1} = -4.51614 \quad & \times 10^{-1} \times h + 1.83925 \times 10^{-2} \times \rho_{tot} \times s + 1.77130 \times 10^{-1} \times d \\
& - 3.67223 \times 10^{-1} \times a_d \times s + 2.69541 \times 10 \times \rho_{tot} + 1.72100 \times 10 \times d_{bl} \\
& + 4.41065 \times 10^{-1} \times b \times \rho_w - 2.63221 \times 10^{-4} \times V_r^2 + 1.22815E \times 10^{-1} \times d_{bw} \times s \\
& + 2.38053 \times 10^{-2} \times h \times f_c' + 3.25673 \times 10^{-1} \times V_r \\
& + 6.82692 \times 10^{-4} \times A_{tr} \times f_{yw} \times -9.40552 \times 10^{-1} \times f_c' \times d_{bw} \\
& + 1.99045 \times 10 \times c \times \rho_w - 3.70053 \times 10 \times \rho_w^2 + 8.01932 \times 10 \times v \\
& - 8.63480 \times 10 \times c \times v + 4.64570 \times 10^{-4} \times f_{yl} \times s + 9.46359 \times 10^{-5} \times P \times V_r \\
& + 2.91722 \times 10^{-3} \times f_{yl} \times f_c' + 1.09822 \times 10^{-4} \times d \times P \\
& - 8.39975 \times 10^{-1} \times a_d \times f_c' \; where \; P \; in \; KN
\end{aligned}
\tag{1}
$$

$$
\begin{aligned}
V_{exp,2} = 4.50889 \quad & \times 10^{-3} \times f_{yl} \times s + 1.23497 \times 10 \times h + 3.49000 \times 10 \times a_d - 1.11066 \times 10 \times d \\
& + 2.97440 \times 10 \times \rho_{tot} - 9.80018 \times 10^{-5} \times A_{tr}^2 \\
& + 4.30251 \times 10^{-4} \times P \times 0.4 \times \xi \times b \times d \times f_c' + 1.59616 \times 10^{-3} \times s \times 0.4 \times \xi \times b \times d \times f_c' \\
& - 3.91027 \times 10^{-1} \times A_{tr} - 2.49454 \times 10 \times s + 3.61139 \times 10 \times 0.4 \times \xi \times b \times d \times f_c' \\
& + 3.58780 \times 10^{-1} \times V_r - 1.34864 \times 10 \times a_d \times 0.4 \times \xi \times b \times d \times f_c' \\
& + 1.78231 \times 10^{-2} \times f_c' \times A_{tr} + 1.46317 \times 10^{-4} \times f_{yl} \times V_r - 2.12422 \times 10^{-3} \times h \times f_{yl} \\
& + 1.77066 \times 10^{-2} \times d \times d_{bl} - 2.05816 \times 10^{-1} \times v * V_r + 1.1749510^{-1} \times d_{bw} \times s \\
& - 2.04027 \times 10^{-1} \times a_d \times f_c' - 4.38284 \times 10 \times a_d \times v + 6.14056 \times 10 \times d_{bw} \\
& - 3.24594 \times 10^{-2} \times v \times P - 1.11574 \times 10 \times d_{bw}^2 + 1.18462 \times 10^{-1} \times f_{yw} \times v \\
& - 9.82636 \times 10 \times d_{bw} \\
& \times v \; where \; P \; in \; KN
\end{aligned} \tag{2}
$$

$$
\begin{aligned}
V_{exp,3} = 1.17316 \quad & \times 10^2 a_d - 6.90891 \times 10^{-1} \times d - 1.26582 \times 10 \times c + 3.39341 \times 10^{-1} \times f_{yl} \\
& - 4.92902 \times 10 \times a_d \times \rho_{tot} - 1.96488 \times 10 \times d_{bw} + 1.48838 \times 10 \times V_c \\
& + 5.59181 \times 10 \times c \times \rho_{tot} + 1.60358 \times 10^{-1} \times d \times d_{bw} - 2.39573 \times 10^{-1} \times c \times f_c^i \\
& - 4.45350 \times 10 \times c \times v + 1.39423 \times 10^{-2} \times h \times f_c^i + 9.45579 \times 10^{-3} \times h \times V_c \\
& - 9.21490 \times 10^{-2} \times b \times a_d + 2.81197 \times 10^{-1} \times f_{yw} + 4.43107 \times 10^{-2} \times c \times c \\
& - 2.95561 \times 10^{-2} \times d_{bw} \times f_{yw}
\end{aligned} \tag{3}
$$

Figure 3 shows the results of the numerically derived models during the testing and training of POLYREG-HYT for the three datasets. It is easy to observe a strong correlation between the relationships of the experimental values versus those derived from the formulae for each scenario. It is also worth noting that, despite the fact that the newly developed predictive models differ significantly from the original models proposed in the international literature (especially the number of terms), the proposed formulae were not only able to predict the shear strengths accurately but also to provide a 2-fold improvement in terms of correlation when compared with the three previously proposed predictive equations (i. EN1998-3, 2005 [11]; ii. Pardalopoulos et al., 2013 [22]; and iii. ASCE/SEI 41-13, 2014 [21]). This highlights the numerical superiority of the adopted ML algorithm of POLYREG-HYT, which has the ability to provide us with a closed-form solution [29].

To enhance the accuracy of the predictive models, XGBoost-HYT-CV [29] was employed in addition to POYREG-HYT and LRegr. The results for the test and training datasets using XGBoost-HYT-CV are presented in Figure 4 and Tables 3–5. These findings demonstrate that the advanced ML method achieved the highest correlation coefficient values between experimental and predicted results. Specifically, Figure 4 and Tables 3–5 showcase the outcomes of applying XGBoost-HYT-CV to the three datasets with a 90/10% test–train ratio.

Based on the correlation coefficient R values and the values of the MAPE metric presented in Tables 3–5, the dataset developed based on Pardalopoulos et al. (2013) [22] derived the highest correlation coefficient value (R = 99.65%) when using the ML algorithm of XGBoost-HYT-CV. The ASCE/SEI 41-13, 2014 [21] dataset follows with a correlation coefficient of 99.58% and an error of 8.95% (on the test dataset). Lastly, the EN 1998-3 (2005) [11] prepared dataset derived an optimum correlation coefficient value of 98.45% and an error of 17.7% for the case of the POLYREG-HYT ML algorithm followed by a 97% correlation coefficient and a corresponding 15.3% MAPE value for the case of XGBoost-HYT-CV. It is easy to observe that even though the POLYREG-HYT managed to derive a higher correlation coefficient (1.45% higher) compared to the one obtained by XGBoost-HYT-CV, the latter resulted in a lower MAPE value when used on the test dataset.

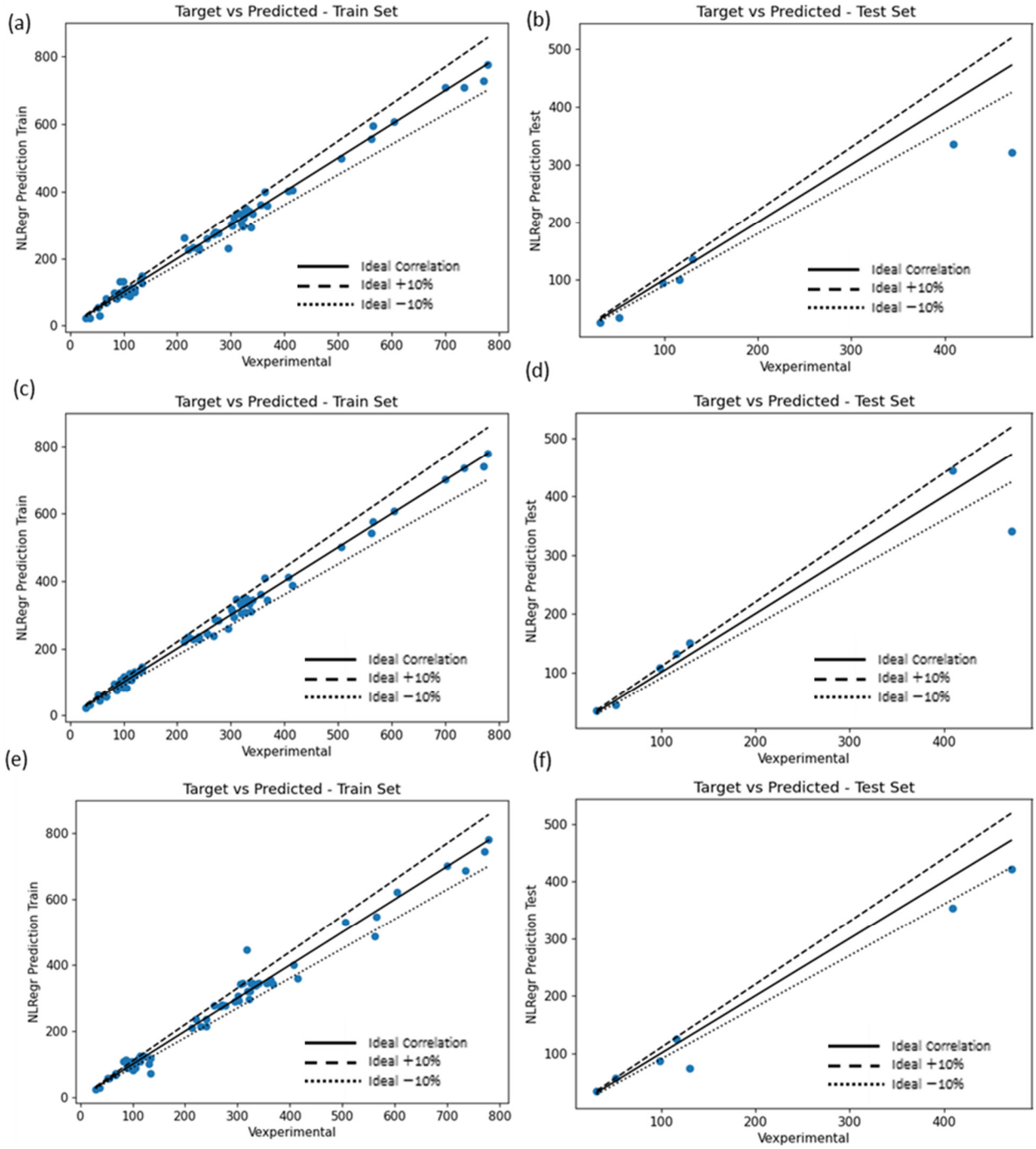

Figure 3. Correlation between experimental shear strengths and numerically predicted values using the POLYREG-HYT method based on the following models: (**a**,**b**) EN1998-3 (2005) [11]: training and testing datasets; (**c**,**d**) Pardalopoulos et al. (2013) [22]: training and testing datasets; (**e**,**f**) ASCE/SEI 41-13 (2014) [21]: training and testing datasets.

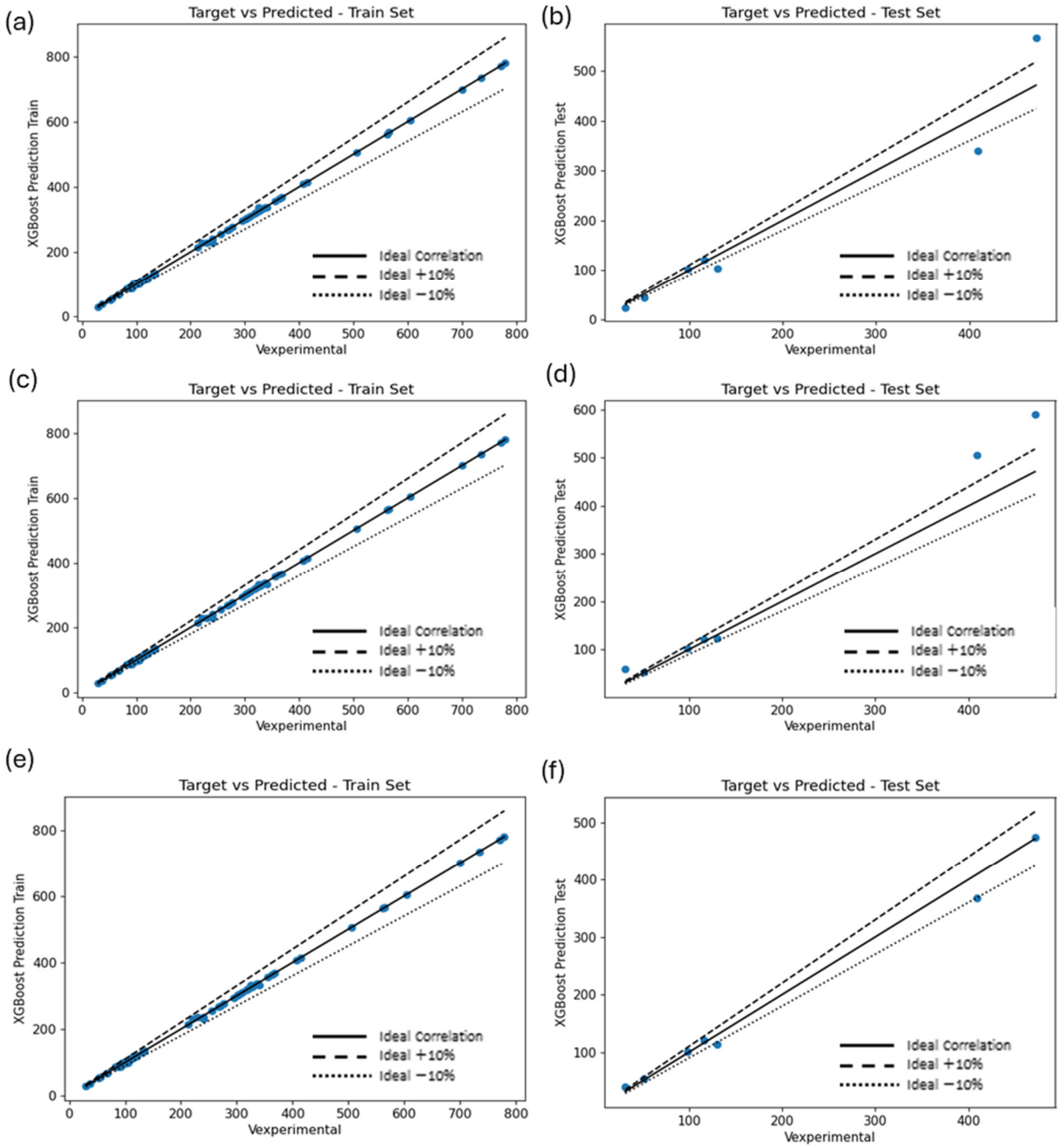

Figure 4. Comparison of numerically predicted shear strengths using XGBoost-HYT-CV to experimental values and proposed model predictions: (**a**,**b**) EN1998-3 (2005) [11] dataset; (**c**,**d**) Pardalopoulos et al. (2013) [22] dataset; and (**e**,**f**) ASCE/SEI 41-13 (2014) [21] dataset.

Table 3. Error metric evaluation for various machine learning algorithms on the EN1998-3 (2005) dataset, using a 90/10 train–test split.

MODEL	DATASET	R	MAPE	MAMPE	MAE	RMSE
POLYREG-HYT	Train	0.99490	0.07903	0.05084	13.224	18.4979
LREGR	Train	0.96821	0.19220	0.13054	33.9533	45.8741
XGBOOST-HYT-CV	Train	0.99987	0.01225	0.00644	1.67587	2.92632
POLYREG-HYT	Test	0.98448	0.17764	0.20863	38.9874	63.9059
LREGR	Test	0.84845	0.42230	0.33075	61.8105	92.0577
XGBOOST-HYT-CV	Test	0.97007	0.15299	0.16567	30.9598	46.2470

Table 4. Error metric evaluation for various machine learning algorithms on the Pardalopoulos et al. (2013) dataset using a 90/10 train–test split.

MODEL	DATASET	R	MAPE	MAMPE	MAE	RMSE
POLYREG-HYT	Train	0.99647	0.06761	0.04623	12.0254	15.3862
LREGR	Train	0.97552	0.16353	0.11609	30.1960	40.3127
XGBoost-HYT-CV	Train	0.99989	0.00913	0.00495	1.28678	2.72715
POLYREG-HYT	Test	0.95266	0.13873	0.16834	31.4586	52.3988
LREGR	Test	0.87672	0.49074	0.30313	56.6489	85.0794
XGBoost-HYT-CV	Test	0.99652	0.21500	0.19804	37.0085	59.0950

Table 5. Error metric evaluation for various machine learning algorithms on the ASCE/SEI 41-13 (2014) dataset using a 90/10 train–test split.

MODEL	DATASET	R	MAPE	MAMPE	MAE	RMSE
POLYREG-HYT	Train	0.99001	0.08229	0.06268	16.3040	25.8498
LREGR	Train	0.97806	0.14642	0.10738	27.9311	38.2091
XGBoost-HYT-CV	Train	0.99988	0.00958	0.00532	1.38284	2.86823
POLYREG-HYT	Test	0.99140	0.15463	0.14728	27.5236	36.1777
LREGR	Test	0.92665	0.25136	0.23429	43.7839	62.0201
XGBoost-HYT-CV	Test	0.99575	0.08957	0.06142	11.4780	17.5487

5.2. Sensitivity Analysis

Sensitivity analysis was performed to evaluate the effect of a specific input on the target variable. According to Hariri-Ardebili (2018) [35], ML is a particular kind of artificial intelligence that "learns" as it finds new patterns in data and develops techniques to enhance decision-making based on knowledge concealed in large datasets. Sensitivity analysis may be applied to both the ML tasks of classification and regression.

To allow for the permutation of only one feature around its given values, all features at certain values (i.e., 25%, median, and 75% quantiles) are kept fixed as soon as the trained models are formed. The matching sensitivity curves for every feature and every trained model are obtained after these values are established in training mode. Therefore, it is possible to determine the equivalent influence of each input feature on the target variable, which is shown in a summarized manner in Figure 5.

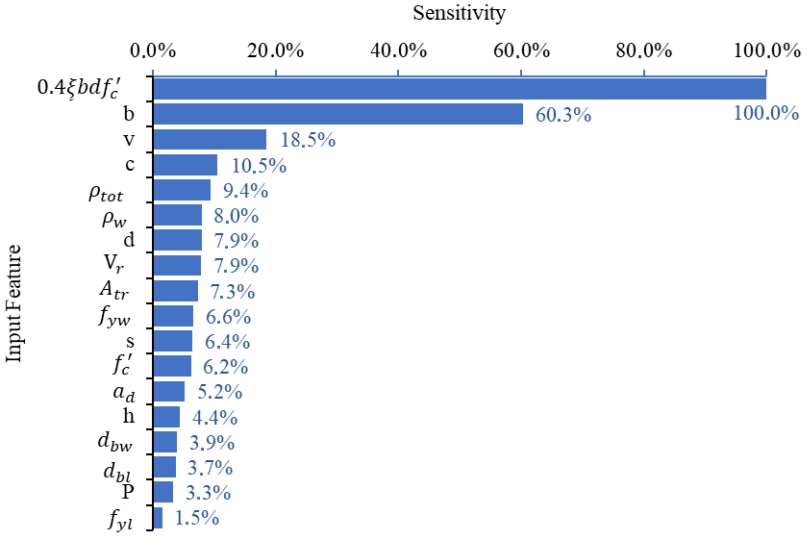

Figure 5. A bar chart-based sensitivity analysis to assess the impact of different variables in the Pardalopoulos et al. (2013) [22] dataset on model output.

It is important to note at this point that given its superior predictive performance, XGBoost-HYT-CV (used on the Pardalopoulos et al. (2013) [22] dataset) was selected to perform the sensitivity analysis shown in Figure 5. This numerically developed predictive model exhibited the highest correlation coefficient, lowest MAPE, and the most accurate predictions on the test dataset. Additionally, the sensitivity analysis across all three datasets revealed a consistent hierarchy of parameter importance, reinforcing the credibility of the Pardalopoulos et al. (2013) [22] dataset.

The sensitivity analysis, as illustrated in Figure 5, reveals key parameters that are consistent with previous research findings, thus strengthening the conclusions derived from the numerical findings of this research work. According to Figure 5, the primary factors influencing the shear strength of RC columns include the compressive zone height, specimen geometry, axial load ratio, concrete cover, longitudinal and transverse reinforcement percentages, effective depth, stirrup yield stress, transverse reinforcement area within diagonal cracks, stirrup spacing, and the aspect ratio.

Furthermore, these findings align with prior studies focused solely on the experimental evaluation of shear strength in older-style RC structural elements. Sezen and Mohele (2004) [7] demonstrated that the shear strength of lightly reinforced rectangular columns is notably influenced by variables such as aspect ratio, axial load, amount of transverse reinforcement, and deformation ductility requirements.

6. Conclusions

The present study investigated the strength and deformation parameters defining the mechanical behavior of RC columns under lateral sway such as what occurs during seismic loading, focusing on the details of the failure mechanisms of columns. Carefully chosen specimens from the experimental database that highlight the behavior of columns experiencing shear failure after flexural yielding were assembled to study the parametric sensitivity of the examined data, while at the same time, an evaluation of existing relationships for the limit state failure in strength terms was carried out (shear mechanism failure and bearing capacity at failure).

To develop a predictive model for RC columns, a dataset of 74 samples was employed to train a polynomial regression model optimized through hyperparameter tuning. The resulting predictive formula, when using POLYREG-HYT, outperformed existing models in the literature, achieving a correlation coefficient R of 99.1% compared to 53.1%, 43.3%, and 32.6% for the cases of the EN1998-3 [11], Pardalopoulos et al. (2013) [22], and ASCE/SEI 41-13 [21] predictive models, respectively.

Furthermore, the XGBoost-HYT-CV ML algorithm was found to derive the most accurate predictive model, reaching a correlation coefficient R of 99.65% when used on the test dataset. This is currently the most accurate model found in the international literature, which corresponds to a 2.3 times improved correlation coefficient compared to the EN1998-3 [11] model and an almost 2 times improvement compared to the Pardalopoulos et al. (2013) [22] predictive model. This highlights the numerical superiority of the proposed predictive model when used to compute the shear strength of RC columns, leading to safer and more sustainable designs. Finally, it is advisable to use the proposed predictive models for RC column geometries that fall within the geometrical spectrum defined by the minimum and maximum RC column geometry included in the dataset presented in this manuscript.

It must be noted that the preparation of the dataset foresaw the use of specific experimental results that followed a specific type of failure. Given that ML algorithms have the ability to provide complicated connections between the input and output features, future work will foresee the use of larger datasets for the development of enhanced predictive

models that will not be limited by the type of shear failure of the RC column or any geometrical restraints. This research work will allow the development of universal predictive models that will allow civil engineers to be able to compute the ultimate shear strength of RC columns without the need to make any adjustments or assumptions when using the predictive models.

Future advancements aim to extend these models with broader datasets, addressing limitations such as sample size constraints due to strict inclusion criteria and enhancing degradation curve predictions of RC column resistance. Broader datasets and refined models will ultimately support universal predictive capabilities, enabling civil engineers to accurately compute RC column shear strength without requiring model adjustments.

Author Contributions: Conceptualization: A.I.I.; Methodology: A.I.I., G.M., N.B. and D.G.; Software: N.B. and G.M.; Validation: A.I.I., G.M. and D.G.; Formal analysis: A.I.I. and D.G.; Investigation: A.I.I. and D.G.; Resources: A.I.I., G.M. and N.B.; Data curation: A.I.I. and D.G.; Writing—original draft preparation: A.I.I.; Writing—review and editing: A.I.I., G.M. and J.B.; Visualization: A.I.I.; Supervision: A.I.I. and G.M.; Project administration: A.I.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy restrictions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Organization of Seismic Planning and Protection (OASP). *Regulation of Interventions (KAN.EPE.)*, 2nd ed.; OASP: Athens, Greece, 2014.
2. Berry, M.; Parrish, M.; Eberhard, M. *PEER Structural Performance Database User's Manual (Version 1.0)*; PEER Center, University of California: Berkeley, CA, USA, 2004.
3. Aschheim, M.; Moehle, J.P. *Shear Strength and Deformability of RC Bridge Columns Subjected to Inelastic Cyclic Displacements*; Report No. PEER 1992/02; California Department of Transportation: Sacramento, CA, USA, 1992.
4. Priestley, M.J.N.; Verma, R.; Xiao, Y. Seismic shear strength of reinforced concrete columns. *J. Struct. Eng.* **1994**, *120*, 2310–2319. [CrossRef]
5. Lynn, A.C.; Moehle, J.P. Seismic Evaluation of Existing Reinforced Concrete Building Columns. Master's Thesis, University of California, Berkeley, CA, USA, 2001.
6. Elwood, K.J.; Moehle, J.P. Drift capacity of reinforced concrete columns with light transverse reinforcement. *Earthq. Spectra* **2005**, *21*, 71–89. [CrossRef]
7. Sezen, H.; Moehle, J.P. Shear strength model for lightly reinforced concrete columns. *J. Struct. Eng.* **2004**, *130*, 1692–1703. [CrossRef]
8. Sasani, M. Shear strength and deformation capacity models for RC columns. In Proceedings of the 13th World Conference on Earthquake Engineering, Vancouver, BC, Canada, 1–6 August 2004.
9. Markou, G.; Bakas, N.P. Prediction of the shear capacity of reinforced concrete slender beams without stirrups by applying artificial intelligence algorithms in a big database of beams generated by 3D nonlinear finite element analysis. *Comput. Concr.* **2021**, *28*, 433–447.
10. Nguyen, T.A.; Pham, N.M.; Vo, T.C.; Nguyen, D.D. Assessment of shear strength models of reinforced concrete columns. *Eng. Technol. Appl. Sci. Res.* **2022**, *12*, 9440–9444. [CrossRef]
11. *EN 1998-3:2005*; Design of Structures for Earthquake Resistance—Part 3: Assessment and Retrofitting of Buildings. European Committee for Standardization: Brussels, Belgium, 2005.
12. Jiang, J.; Cui, Y.; Han, Z. Experimental study and analysis-oriented model of PET FRP confined ECC cylinders under monotonic axial compression. *Eng. Struct.* **2023**, *280*, 115610. [CrossRef]
13. American Concrete Institute (ACI). *Building Code Requirements for Structural Concrete (ACI Committee 318)*; ACI: Farmington Hills, MI, USA, 2011.
14. *GB 50010–2010*; Code for Design of Concrete Structures. Ministry of Housing and Urban-Rural Development of the People's Republic of China (MOHURD). Architecture and Building Press of China: Beijing, China, 2010.

15. Yu, B.; Liu, S.; Li, B. Probabilistic calibration for shear strength models of reinforced concrete columns. *J. Struct. Eng.* **2019**, *145*, 04019026. [CrossRef]
16. Ma, C.; Deng, Z.; Chang, X.; Zhou, L.; Pan, K.-M.; Zeng, J.-J. Shear capacity model of rectangular RC columns based on Bayesian update. *Structures* **2024**, *61*, 106096. [CrossRef]
17. Coşgun, C. Machine learning for the prediction of evaluation of existing reinforced concrete structures performance against earthquakes. *Structures* **2023**, *48*, 1994–2003. [CrossRef]
18. Kazemi, F.; Asgarkhani, N.; Jankowski, R. Machine learning-based seismic response and performance assessment of reinforced concrete buildings. *Arch. Civ. Mech. Eng.* **2023**, *23*, 94. [CrossRef]
19. Zhang, H.; Cheng, X.; Li, Y.; He, D.; Du, X. Rapid seismic damage state assessment of RC frames using machine learning methods. *J. Build. Eng.* **2023**, *65*, 105797. [CrossRef]
20. Phan, V.-T.; Tran, V.-L.; Nguyen, V.-Q.; Nguyen, D.-D. Machine learning models for predicting shear strength and identifying failure modes of rectangular RC columns. *Buildings* **2022**, *12*, 1493. [CrossRef]
21. *ASCE/SEI 41-13*; Seismic Evaluation and Retrofit of Existing Buildings. American Society of Civil Engineers (ASCE): Reston, VA, USA, 2014.
22. Pardalopoulos, S.J.; Thermou, G.E.; Pantazopoulou, S.J. Preliminary seismic assessment method for identifying RC structural failures. *Comput. Methods Earthq Eng.* **2013**, *2*, 111–128.
23. Tran, C.T.N.; Li, B. Shear strength model for reinforced concrete columns with low transverse reinforcement ratios. *Adv. Struct. Eng.* **2014**, *17*, 1373–1385. [CrossRef]
24. Aval, S.B.; Ketabdari, H.; Gharebaghi, S.A. Estimating shear strength of short rectangular reinforced concrete columns using nonlinear regression and gene expression programming. *Structures* **2017**, *12*, 13–23. [CrossRef]
25. Pardalopoulos, S.J.; Thermou, G.E.; Pantazopoulou, S.J. Rapid preliminary seismic assessment methodology for non-conforming reinforced concrete buildings. In Proceedings of the ECCOMAS Thematic Conference—COMPDYN 2011: 3rd International Conference on Computational Methods in Structural Dynamics and Earthquake Engineering: An IACM Special Interest Conference Programme, Corfu, Greece, 25–28 May 2011.
26. Ioannou, A.; Pantazopoulou, S.J.; Georgiou, A.; Illampas, R. Limit states behaviour (or failure) of reinforced concrete columns. In Proceedings of the Hellenic Concrete Conference, Athens, Greece, 29–31 March 2018.
27. Tureyen, A.K.; Robert, J.F. Concrete shear strength: Another perspective. *ACI Struct. J.* **2003**, *100*, 609–615. [CrossRef]
28. Kim, C.-G.; Park, H.-G.; Eom, T.-S. Cyclic load test and shear strength degradation model for columns with limited ductility tie details. *J. Struct. Eng.* **2019**, *145*, 04018249. [CrossRef]
29. Markou, G.; Bakas, N.P.; Chatzichristofis, S.A.; Papadrakakis, M. General framework of high-performance machine learning algorithms: Application in structural mechanics. *Comput. Mech.* **2024**, *73*, 705–729. [CrossRef]
30. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
31. Bakas, N.P.; Langousis, A.; Nicolaou, M.; Chatzichristofis, S.A. A gradient-free neural network framework based on universal approximation theorem. *arXiv* **2019**, arXiv:1909.13563.
32. Gravett, Z.D.; Mourlas, C.; Taljaard, V.L.; Bakas, P.N.; Markou, G.; Papadrakakis, M. New fundamental period formulae for soil-reinforced concrete structures interaction using machine learning algorithms and ANNs. *Soil Dyn. Earthq. Eng.* **2021**, *144*, 106656. [CrossRef]
33. Ruckstuhl, A. *Introduction to Nonlinear Regression*; IDP Institut für Datenanalyse und Prozessdesign, ZHAW: Zurich, Switzerland, 2016.
34. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
35. Hariri-Ardebili, M. Risk, reliability, resilience (R3) and beyond in dam engineering: A state-of-the-art review. *Int. J. Disaster Risk Reduct.* **2018**, *31*, 806–831. [CrossRef]