COMPDYN 2025
10th ECCOMAS Thematic Conference on
Computational Methods in Structural Dynamics and Earthquake Engineering
M. Papadrakakis, M. Fragiadakis (eds.)
Rhodes Island, Greece, 15-18 June 2025

RANDOM FORESTS SUPERVISED MACHINE LEARNING FOR CLASSIFYING FAILURE MODES IN RC COLUMNS USING BASIC STRUCTURAL INFORMATION

Konstantinos G. Megalooikonomou[1] **and Grigorios N. Beligiannis**[1,2]

[1] Hellenic Open University, School of Science and Technology,
Parodos Aristotelous 18, 26335 Patras, Greece
std153412@ac.eap.gr

[2] Department of Food Science and Technology, University of Patras,
Agrinio Campus, G. Seferi 2, 30100 Agrinio, Greece
gbeligia@upatras.gr

Abstract

Columns are crucial to structural performance, and this paper addresses the critical need for predicting failure modes in reinforced concrete (RC) columns by evaluating the potential of a random forest machine learning (ML) model. This model is built using data from the well-known PEER structural performance database, which compiles results from over 400 cyclic, lateral-load tests on reinforced concrete columns. The database includes tests on spiral or circular hoop-confined columns, rectangular tied columns, and columns with or without lap splices of longitudinal reinforcement at critical sections. Here, the effectiveness of supervised ML techniques is examined, specifically random forests, using a randomly selected test set from the Pacific Earthquake Engineering Research Center (PEER) database. The model achieved an overall accuracy of 94% for rectangular RC columns and 86% for circular RC columns. Additionally, the model's predicted failure modes matched or even outperformed those calculated using code-defined equations (the traditional method) in some cases. This study demonstrates that random forest models are highly effective for postdicting RC column failure modes, highlighting the transformative potential of machine learning in earthquake engineering.

Keywords: Reinforced Concrete Columns; PEER Structural Performance Database, Machine Learning, Random Forests, Failure Mode.

1 INTRODUCTION

The failure mode of structural elements, such as reinforced concrete columns, is influenced by various factors, including their geometric properties, longitudinal reinforcement, the effectiveness of confinement provided by transverse reinforcement, and the loading history. The response of these columns across the loading range is governed by competing resistance mechanisms, such as flexure, shear, buckling of longitudinal bars under compressive loads, and, in cases involving lap splices, the behavior of the lap splice mechanism for reinforcing bar development. Often, a combination of these mechanisms defines the overall behavior of the column, particularly under cyclic load reversals. Numerous predictive models have been proposed to estimate both the strength and deformation capacity of columns. However, as evidenced by test comparisons, the uncertainty in predicting deformation capacity is significantly greater—by at least an order of magnitude—than that for strength [1].

System identification and damage detection is a dual-focused field that leverages machine learning (ML) to replicate structural systems and predict their deterministic seismic responses. Laboratory testing of reinforced concrete (RC) structures has provided valuable data, enabling ML techniques to identify failure modes, strength, capacities, and constitutive behaviors. Recently, ML approaches, which rely on algorithms to learn from data, have been applied to risk assessment and predictive modeling in civil engineering. Some studies have specifically explored failure mode prediction and shear strength estimation for beam–column joints. For example, Mitra et al. (2011) [2] classified non-ductile joint shear failure and ductile beam yielding failure in interior beam–column joints. Similarly, Tang et al. (2022) [3] conducted low-cycle reciprocating loading tests on 23 recycled aggregate concrete-filled steel tube columns and 3 ordinary concrete-filled steel tube columns. Their study employed artificial intelligence, specifically random forests with hyperparameters optimized using the firefly algorithm, to assess the effects of parameter variations on the seismic performance of concrete columns. Related studies, including multi-objective optimization analyses, are discussed in Tang et al (2023) [4].

This study evaluates the performance of a supervised learning algorithm, the random forest, as a predictive model for the first time in postdicting the failure mode of reinforced concrete (RC) columns. The evaluation is conducted using a widely referenced experimental dataset originally compiled by Berry and Eberhard (2004) [5]. Known as the PEER (Pacific Earthquake Engineering Research Center) Structural Performance Database, this resource aggregates results from over 400 cyclic lateral-load tests of RC columns. It includes data on spiral or circular hoop-confined columns, rectangular tied columns, and columns with or without lap splices in the longitudinal reinforcement at critical sections. For each test, where available, the database provides details on column geometry, material properties, reinforcement details, test configuration (including P-Delta effects), axial load, digital lateral force-displacement histories at the column top, and top displacement associated with various damage observations.

In Berry and Eberhard's experimental database (2004) [5], column failure modes were categorized as:

(a) Flexure critical,

(b) Flexure–shear critical, or

(c) Shear critical

These classifications were based on the following criteria:

- If no shear damage was reported, the column was categorized as flexure critical.
- If shear damage (diagonal cracks) was noted, the absolute maximum effective force (F_{eff})-the highest measured force in the experimental response—was compared to the calculated "ideal" force ($F_{0.004}$), corresponding to a maximum axial compressive strain of 0.004 (the strain at which unconfined concrete spalls). The failure

displacement ductility (μ_{fail}) was defined as the displacement ductility at 80% of the maximum effective force (F_{eff}). If $F_{eff} < 0.95 \cdot F_{0.004}$ or if $\mu_{fail} \leq 2$ the column was classified as shear critical. Otherwise, the column was categorized as flexure–shear critical.

All columns in the database were further grouped by cross-sectional shape (rectangular or circular).

2 TRADITIONAL METHOD PREDICTING COLUMN FAILURE MODE BASED ON ENGINEERING MECHANICS

RC columns are crucial to the overall performance of a structure, as their failure can lead to disproportionate consequences for the entire system. The behavior of RC columns under combined axial load, shear, and flexure has been extensively studied for decades. When it comes to flexural behavior, sectional analysis or a fiber model in a one-dimensional stress field can provide reasonable estimates of ultimate strength and yielding deformation along with the flexural failure mode prediction. However, the performance of RC columns dominated by shear or shear-flexure cannot be accurately predicted through sectional analysis alone, as shear force transfer involves stress fields that extend through the member to its supports. Figure 1 illustrates the shear strength degradation models used by EN 1998-3 (2005) [6] and ASCE-SEI 41 (2007) [7] to represent the envelope of resistance curves for reinforced concrete columns as a function of displacement ductility.

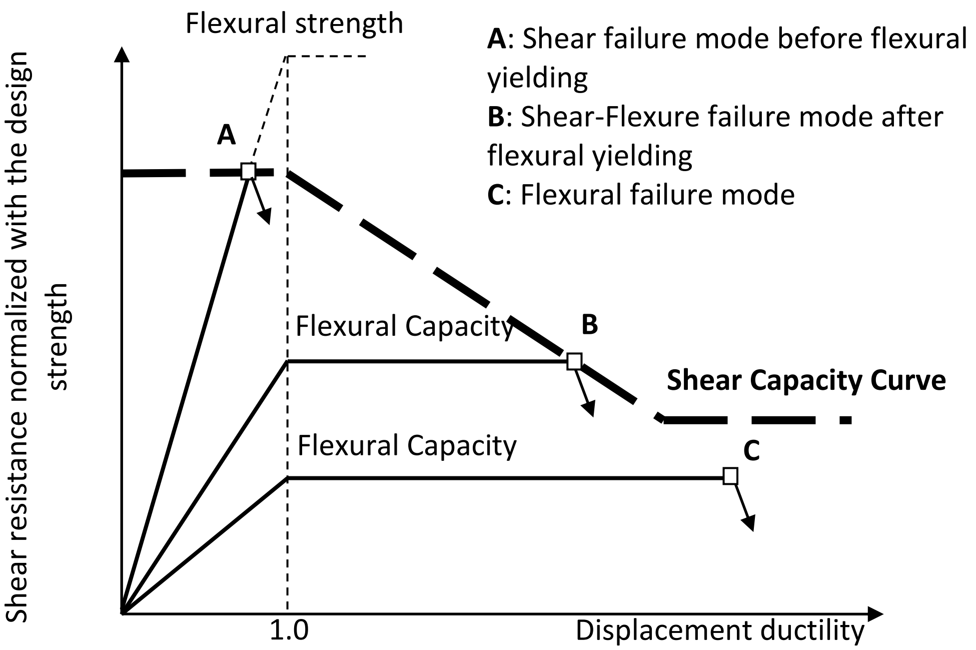

Figure 1: Shear strength degradation model for RC column failure mode prediction adopted by current codes of assessment.

These models serve as the primary criterion for identifying shear failure occurring before or after flexural yielding (the point where the shear curve intersects the flexural capacity curve). To define the shear failure strength and deformation of the reinforced concrete column, it is essential to first establish the flexural capacity curve using classic flexural analysis and then integrate it with the shear strength reduction curve proposed by the codes. This approach is applied in the current section to evaluate the accuracy of the code provisions in identifying the exact failure mode of RC columns before testing the same performance with ML methods in the next Section.

2.1 Shear Demand Curve

Shear demand curves (flexural capacity, Figure 1) for RC columns of the PEER database are obtained through nonlinear static pushover or cyclic analysis using a well-known MATLAB toolbox called FEDEAS Lab [8-10], assuming flexural behavior without the occurrence of shear failure.

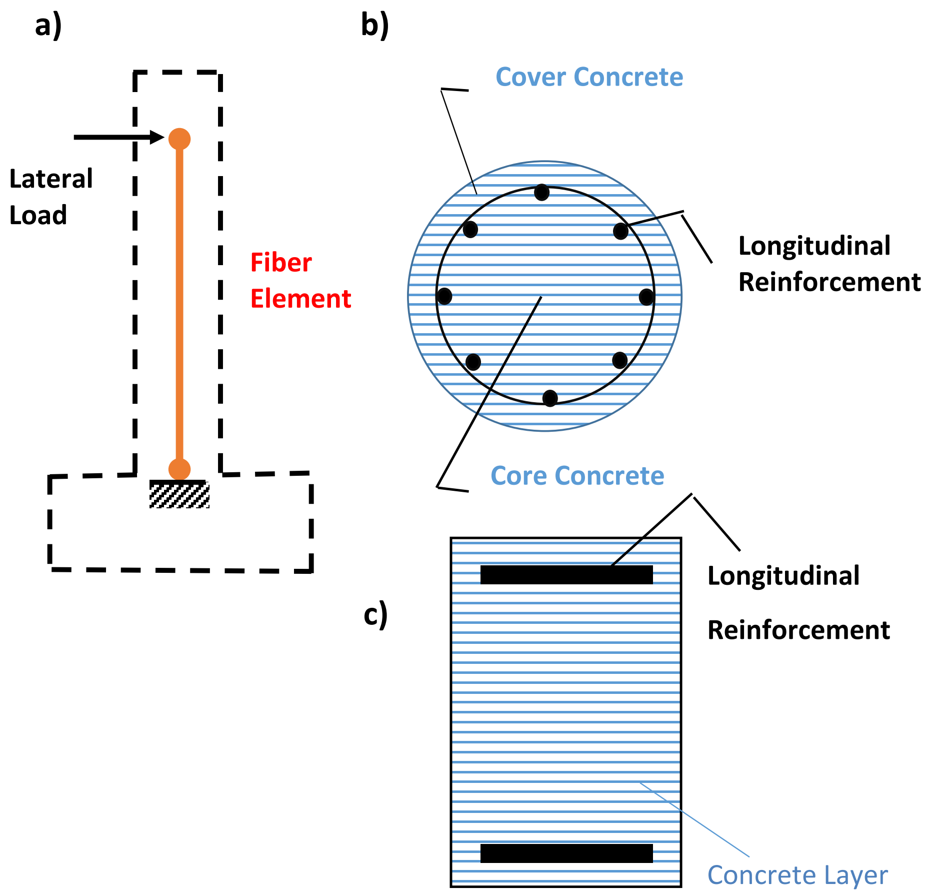

Figure 2: a) Numerical model for Circular and Rectangular RC Columns failed in flexure in FEDEAS MATLAB toolbox [8-10] b) Circular Section discretization in fibers/layers c) Rectangular Section discretization in fibers/layers.

Numerical simulations were performed using a nonlinear fiber beam-column element that accounts for the spread of plasticity. In this analysis, the longitudinal beam element employs a force-based formulation with a linear moment distribution to construct a flexibility matrix that evolves step by step with increasing nonlinearity. Strain-displacement relationships are implicitly defined by inverting the flexibility matrix to obtain stiffness. Assuming strain compatibility among the materials in the member, the formulation evaluates sectional response at selected integration points along its length. At the sectional level, the Bernoulli hypothesis (plane sections remain plane and perpendicular to the member axis) is applied to relate fiber strains to sectional curvature and axial strain. Nonlinear uniaxial material models are used to define the relationship between normal stress and strain in the fibers, neglecting the impact of shear on altering the principal stress orientations through the cross-sectional height. A typical column section discretization is depicted in Figure 2. Sectional stress resultants, such as moment and axial load, are derived by balancing the contributions of fiber stress resultants.

2.2 Shear Capacity Curve

Shear capacity curves of the RC columns of PEER database are obtained based on the shear strength degradation model of Figure 1.

According to EN 1998-3 (2005)[6], the part of the cyclic shear resistance that depends on concrete and transverse steel contribution (excluding the part owing to axial load contribution), V_R, decreases with the plastic part of ductility demand, expressed in terms of ductility ratio of the transverse deflection of the shear span or of the chord rotation at member end: $\mu_{\Delta}{}^{pl} = \mu_{\Delta} - 1$. For this purpose $\mu_{\Delta}{}^{pl}$ may be calculated as the ratio of the plastic part of the chord rotation, θ_p, normalized to the chord rotation at yielding, θ_y.

Thus, EN 1998-3 (2005) [6] defines shear strength accounting for the above reduction as follows:

$$V_R = [(h-x)/2L_s] \cdot \min(N; 0.55A_c f_c) + [1 - 0.05min(5; \mu_{\Delta}{}^{pl})] \cdot \{0.16max(0.5; 100\rho_{tot})[1 - 0.16min(5; L_V/h)]\sqrt{f_c}A_c + V_w\} \quad (1)$$

where $\boldsymbol{h}$: is the depth of the cross-section (equal to the diameter D for circular sections); $\boldsymbol{x}$: is the compressive zone depth; $\boldsymbol{N}$: is the compressive axial force (positive, taken as being zero for tension); $\boldsymbol{L_s} = M/V$ is the shear span of the member; $\boldsymbol{A_c}$: is the cross-sectional area, taken as being equal to $\boldsymbol{b_w d}$ for a cross-section with a rectangular web of width (thickness) $\boldsymbol{b_w}$ and structural depth $\boldsymbol{d}$ or to $\boldsymbol{\pi D_c{}^2/4}$ (where $\boldsymbol{D_c}$ is the diameter of the concrete core to the inside of the hoops) for circular sections; $\boldsymbol{f_c}$: is the concrete compressive strength, and $\boldsymbol{\rho_{tot}}$: is the total longitudinal reinforcement ratio. Term $\boldsymbol{V_w}$ is the contribution of transverse reinforcement to shear resistance, taken as equal to

$$V_w = \frac{\pi}{2}\frac{A_{sw}}{S} f_{yw}(D - 2c) \quad (2)$$

where, $\boldsymbol{f_{yw}}$ is the yield stress of the transverse reinforcement, $\boldsymbol{A_{sw}}$ the area of the spiral wire, $\boldsymbol{c}$ the concrete cover, and $\boldsymbol{S}$ is the spiral step (spacing between successive turns of a spiral). Similarly, for rectangular cross-sections with a web having width $\boldsymbol{b_w}$:

$$V_w = \rho_w b_w z f_{yw} \quad (3)$$

where $\boldsymbol{\rho_w}$ is the transverse reinforcement ratio, $\boldsymbol{z}$ is height of the equivalent truss (internal lever arm between longitudinal tension and compression resultants, i.e., *d-d'* in beams and columns).

In concrete columns with shear span ratio of L_s/h less or equal to 2, the shear strength, V_R cannot exceed the value corresponding to failure by web crushing along the diagonal of the column after flexural yielding, $V_{R,max}$, which under cyclic loading may be calculated from the expression:

$$V_{R,max} = (4/7)[1 - 0.02min(5; \mu_{\Delta}{}^{pl})][1 + 1.35(N/A_c f_c)][1 + 0.45(100\rho_{tot})] \cdot \sqrt{min(40; f_c)}b_w z \cdot sin2\delta \quad (4)$$

where $\boldsymbol{\delta}$ is the angle between the diagonal strut that is defined by the centroids of the compression zones at the column ends, and the axis of the column ($\tan\delta = h/2L_s$).

ASCE/SEI 41 [7] is the latest in a series of documents developed after the FEMA [11] initiatives in the 1990s and 2000s towards the development of a consistent assessment framework for existing structures. The FEMA/ATC documents form the first integrated reference for performance-based engineering, whereby deformation and force demands for different seismic hazards are compared against the capacities at various performance limits (i.e. states of damage). At the outset of this momentous project by FEMA, available data on the performance of existing components were rather limited and therefore reliability concepts were not applied evenly towards the establishment of performance criteria. The issue of dependably estimating the shear strength of a RC element appears to be rather complicated as it presumes the full understanding of the several interacting behavior mechanisms under reversed cyclic loading, whereas it is strongly affected by the imposed loading history, the dimensions of the element (e.g. the aspect ratio), the concrete strength, the longitudinal reinforcement ratio but mostly the ratio and the detailing of the transverse reinforcement. So far it has not been possible to theoretically describe the strength of the shear mechanism from first principles of mechanics without the use of calibrated empirical constants. Therefore, the shear strength estimates obtained from calibrated design expressions necessarily rely on the pool of experimental data used for correlation of the empirical expressions, as well as on the preconceived notions of the individual researchers as to the role each variable has in the mechanics of shear.

The following expression for estimation of the shear strength of reinforced concrete columns is proposed by the Code for seismic rehabilitation of existing buildings of the American Society of Civil Engineers ASCE/SEI 41 (2007) [7]:

$$V_R = V_c + V_w = k(\mu_\Delta)\left[\left(0.5\sqrt{f_c}/(L_s/d)\right)\sqrt{1 + N/\left(0.5A_g\sqrt{f_c}\right)}\right]0.8A_g + k(\mu_\Delta)\cdot \left[A_{sw}f_{yw}d/S\right] \quad (5)$$

where $\boldsymbol{V_c}$ is the concrete contribution in shear resistance; $\boldsymbol{V_w}$ is the contribution of transverse reinforcement; $\boldsymbol{d}$ is the effective depth; $\boldsymbol{L_s}$ is the shear span of the column; $\boldsymbol{N}$ is the axial force (compression positive, taken zero for tension); $\boldsymbol{A_g}$ is the gross cross-sectional area of the column; $\boldsymbol{A_{sw}}$ is the cross-sectional area of one layer of stirrup reinforcement parallel to the shear action; and $\boldsymbol{S}$ is the centerline spacing of stirrups along the length of the member. If $\boldsymbol{S}$ is equal to or greater than half of the effective depth of the column then the contribution of steel reinforcement $\boldsymbol{V_w}$ in shear strength is reduced to 50% of its estimated value from the above equation. If $\boldsymbol{S}$ is equal to or greater than the effective depth of the column then zero shear strength contribution from steel reinforcement $\boldsymbol{V_w}$ is considered; $\boldsymbol{f_c}$ is the concrete compressive strength; $\boldsymbol{k(\mu_\Delta)}$ is the shear strength reduction coefficient that depends on ductility demand. If ductility demand is less than or equal to 2 then the factor is set to equal to 1 (i.e. no strength reduction). If the ductility is greater than 6, then the reduction factor is equal to 0.6. For ductility between 2 and 6 the reduction factor is linearly interpolated between the proposed values.

2.3 Failure Mode Prediction

In the next Section the efficiency of ML methods and especially random forests in predicting the failure mode of RC columns testcases taken from PEER structural performance database will be tested. In this Section however the same performance inquiry will be examined for the traditional method based on engineering mechanics using the same test set of columns as the one that will be employed in the next Section. Moreover, the confusion matrix as performance metric of the traditional method will be reported here too both for circular and rectangular RC columns of the PEER database.

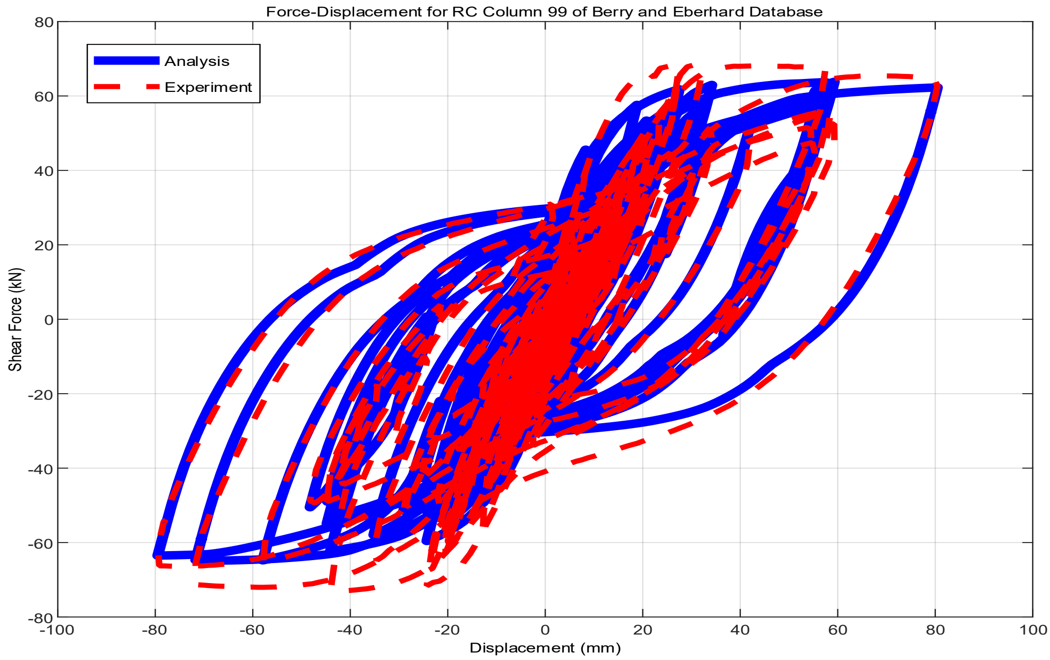

Figure 3: Comparison between numerical and experimental response of circular column failed in flexure (ID#99) (specimen case obtained from the Berry and Eberhard Database 2004 and was analyzed herein).

For the case of a circular column failed in flexure for example, a single beam-column element was used to represent the entire length of the cantilever column (Figure 2), with five Gauss-Lobatto integration points defined along the element [FEDEAS Lab [8-10]]. The effect of confinement on the concrete core was modeled by appropriately modifying the properties of the uniaxial stress-strain law for concrete in compression [12,13]. The P-Δ effect was not included in this simulation. The computed lateral force-lateral displacement response of the column is plotted in Figure 3 for comparison with experimental results.

For a rectangular RC column from the test set that failed in flexure, a single fiber element was assigned to represent the entire height of the cantilever column here too (Figure 2). Five Gauss-Lobatto integration points were defined along the element. The uniaxial stress-strain response of concrete was modeled using the relationship proposed by Mander et al. (1988). The differing confinement effects between the unconfined concrete cover and the confined concrete core were not incorporated into the section discretization (Figure 2). The longitudinal reinforcement's stress-strain behavior was simulated using the model by Menegotto and Pinto (1973) [14]. Once again, the P-Delta effect was excluded from the analysis. The lateral force-lateral displacement response from the numerical simulation of the column was compared with experimental results, as shown in Figure 4. Similar to the case of circular section columns, good agreement was observed between numerical and experimental results. Thus, in the performance metrics it will be considered that the flexural mode of failure both for circular and rectangular columns can be 100% predicted by traditional methods based on engineering mechanics and code provisions.

Figure 5 presents a comparison between the analytical and experimental responses of a rectangular column failed in shear (ID#140). The correlation is quite poor, even in terms of the initial stiffness predicted by flexural analysis. This discrepancy arises from the omission of deformation contributions due to reinforcement pullout and shear deformation. It is evident that only the degrading shear strength model from ASCE-SEI 41 (2007) [7] intersects the flexural capacity curve, indicating the occurrence of shear strength failure as a result of shear strength degradation. However, the displacement at which the latter failure is predicted occurs earlier than the actual onset of strength degradation observed in the experimental results. Regarding performance in this case since not all the code provisions predicted correctly the failure mode it will be considered that the prediction was not successful.

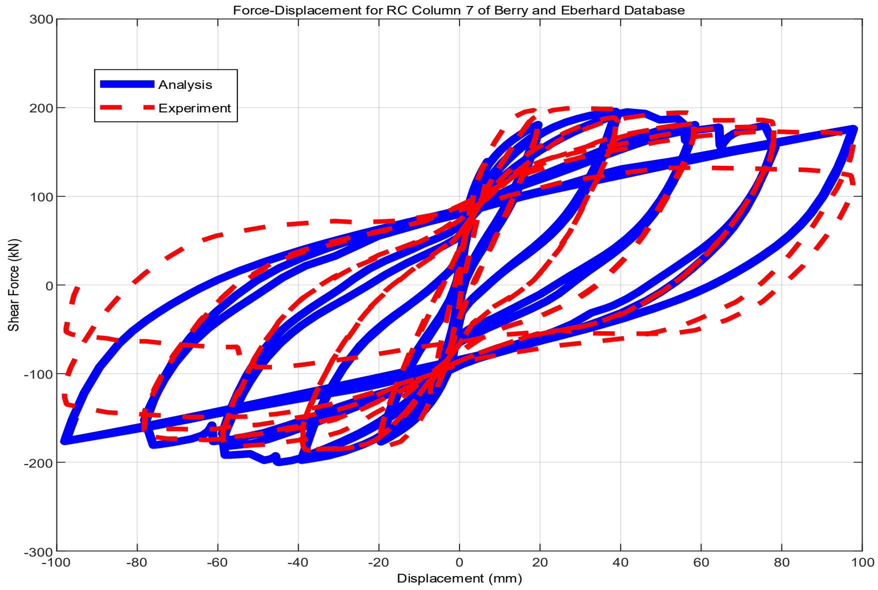

Figure 4: Comparison between numerical and experimental response of rectangular column failed in flexure (ID#7) (specimen case obtained from the Berry and Eberhard Database 2004 and was analyzed herein).

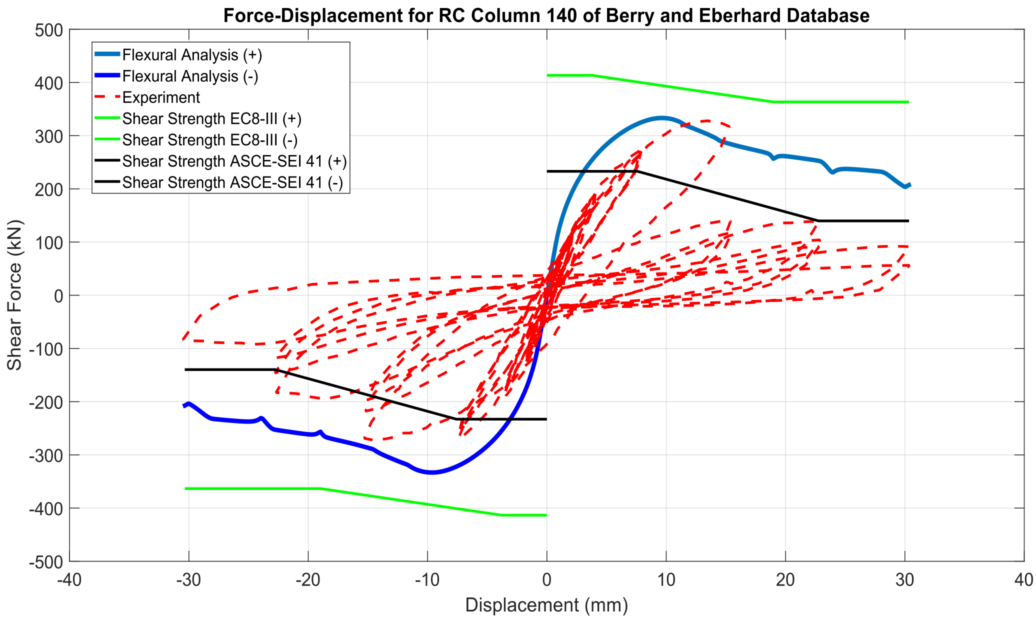

Figure 5: Comparison between numerical and experimental response of rectangular column failed in shear (ID#140) (specimen case obtained from the Berry and Eberhard Database 2004 and was analyzed herein).

Both shear strength degradation models depicted in Figure 6 identified shear failure after yielding, but at a displacement significantly lower than observed in the corresponding experimental results. The EN 1998-3 (2005) [6] model provided a more accurate estimation of strength at shear failure compared to the ASCE-SEI 41 (2007) [7] model. Here the traditional method was successful in predicting the failure mode. More comparison cases are included in [15].

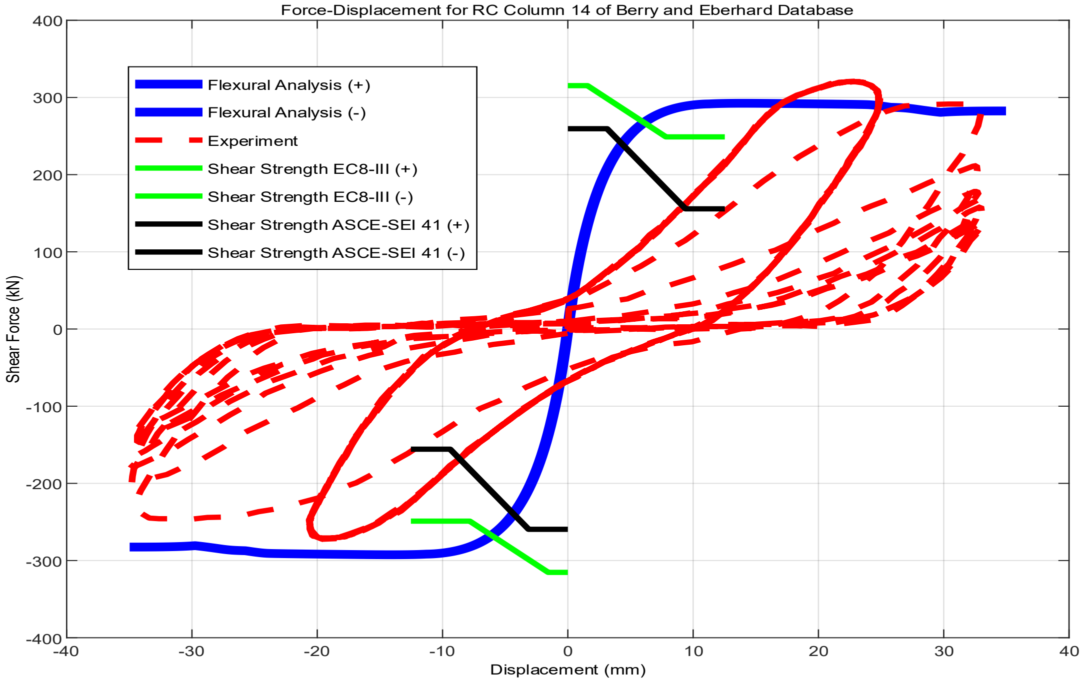

Figure 6: Comparison between numerical and experimental response of circular column failed in shear (ID#14) (specimen case obtained from the Berry and Eberhard Database 2004 and was analyzed herein).

The performance metrics for circular RC columns through failure mode prediction with the traditional method are presented below. The traditional method based on engineering mechanics and code provisions (see also [15]) achieves an accuracy of 81% in predicting the actual failure mode of the tested columns. This accuracy can be calculated using Table 1 by dividing the sum of the diagonal elements by the total sum of all elements in the table.

Table 1: Confusion matrix in numbers for prediction through engineering mechanics and code provisions of the failure mode of circular RC columns of PEER structural performance database.

		Confusion Matrix in Numbers		
True Values	Flexure	12	0	0
	Flexure–Shear	2	5	0
	Shear	0	2	0
		Flexure	Flexure–Shear	Shear
		Predicted Values		

Finally, the performance metrics for rectangular RC columns through failure mode prediction with the traditional method are presented below too. The traditional method based on engineering mechanics and code provisions (see also [15]) achieves an accuracy of 97 % in predicting the actual failure mode of the tested columns. This accuracy can be calculated using Table 2 by dividing the sum of the diagonal elements by the total sum of all elements in the Table.

Table 2: Confusion matrix in numbers for prediction through engineering mechanics and code provisions of the failure mode of rectangular RC columns of PEER structural performance database.

		Confusion Matrix in Numbers		
True Values	Flexure	57	0	0
	Flexure–Shear	1	3	0
	Shear	1	0	0
		Flexure	Flexure–Shear	Shear
		Predicted Values		

3 SUPERVISED MACHINE LEARNING PREDICTION OF COLUMN FAILURE MODE WITH RANDOM FORESTS

3.1 Random Forests with Python

After providing the failure modes along with the performance metrics of traditional failure mode prediction based on nonlinear structural analysis, the methodology for predicting the failure modes of reinforced concrete columns by exploring the potential of machine learning (ML) methods is introduced. The process for achieving this goal is detailed here, focusing on supervised ML techniques, such as random forests, applied to a randomly assigned test set derived from the PEER database [16].

In any machine learning problem, the process typically involves the following steps:

1. Define the problem and identify the required data.
2. Collect the data in a usable format.
3. Address any data gaps or uncertainties and resolve them as needed.
4. Prepare the data for use in the machine learning model.
5. Establish a baseline model to serve as a benchmark for improvement.
6. Train the model using the training dataset.
7. Use the model to make predictions on the test dataset.
8. Compare the predictions to the known test targets and calculate performance metrics.
9. If the performance is inadequate, refine the model, gather additional data, or explore alternative modeling techniques.

3.2 Results

The performance metrics of applying the above-described methodology for rectangular RC columns are presented below. Random forests achieve a 94% accuracy rate in predicting the actual failure modes of the tested data. This accuracy can be calculated from Table 3 by dividing the sum of the diagonal matrix terms by the total sum of all table terms. Additional performance metrics are provided in Table 4.

Table 3: Confusion matrix in numbers for ML prediction of the failure mode of rectangular RC columns of PEER structural performance database with random forest method.

		Confusion Matrix in Numbers *		
True Values	Flexure	55	2	0
	Flexure–Shear	2	2	0
	Shear	0	0	1
		Flexure	Flexure–Shear	Shear
		Predicted Values		

* See also Figure 7

Table 4 : Performance metrics

	Performance Metrics *						
	True Positive	True Negative	False Positive	False Negative	Accuracy	Precision	Recall
Flex-ure	55	2 + 1 + 0 + 0 = 3	2 + 0 = 2	2 + 0 = 2	(55 + 3)/(55 + 3 +2 +2) = 58/62 = 94%	(55)/(55 +2) = 55/57 = 97%	(55)/(55 +2) = 55/57 = 97%
Flex-ure–Shear	2	55 + 0 + 0 + 1= 56	2 + 0 =2	2 + 0 = 2	(2 + 56)/(2 + 56 +2 +2) = 58/62 = 94%	(2)/(2 +2) = 2/4 = 50%	(2)/(2 +2) = 2/4 = 50%
Shear	1	55 +2 +2 +2 = 61	0 + 0 = 0	0 + 0 = 0	(1 + 61)/(1 + 61 +0 +0) = 62/62 = 100%	(1)/(1 +0) = 1/1 = 100%	(1)/(1 +0) = 1/1 = 100%

* See also Table 3

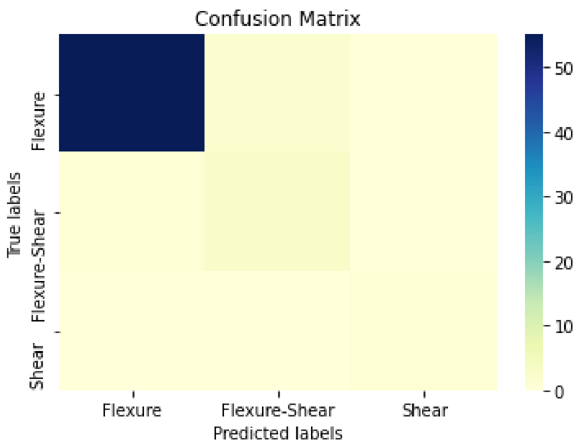

Figure 7: Confusion matrix as performance metric for ML prediction of the failure mode of rectangular RC columns of PEER structural performance database with random forest method.

Analyzing the influence of individual parameters within the feature set reveals that the transverse reinforcement ratio is the most critical factor for the model's success. This finding confirms that the model has correctly identified the relationships between the input features and the target failure modes, underscoring the development of an ML prediction model grounded in physical principles.

Furthermore, Table 4 indicates that the model is particularly effective at predicting flexural and shear failure modes compared to flexure-shear failure modes. This aligns with practical seismic assessment challenges, as flexure-shear is inherently more difficult to identify in real-world engineering contexts. Lastly, it is important to note that Table 3 clarifies the insights illustrated in Figure 7, while Table 4 provides additional context for Table 3.

The performance metrics for circular RC columns are presented below too. Random forests achieve an 86% accuracy rate in predicting the actual failure modes of the tested data. This accuracy is calculated from Table 5 by dividing the sum of the diagonal terms by the total sum of all terms in the table. Additional performance metrics are provided in Table 6.

A detailed analysis of the individual influencing parameters in the feature set reveals that the transverse reinforcement ratio is the most critical factor for the model's success. This finding confirms that the model accurately established the relationship between the features and the target failure mode, reinforcing the development of an ML prediction model grounded in physical principles.

Table 5 further highlights that, for circular RC columns, the model performs better at predicting flexural and flexure-shear failure modes compared to other failure modes. This is expected, as brittle failures typically require nonlinear structural analyses for deterministic detection. Lastly, it should be noted that Table 5 provides clarification for the insights depicted in Figure 8, while Table 6 elaborates on the information in Table 5.

Table 5: Confusion matrix in numbers for ML prediction of the failure mode of circular RC columns of PEER structural performance database with random forest method.

		Confusion Matrix in Numbers *		
True Values	Flexure	12	0	0
	Flexure–Shear	1	5	1
	Shear	0	1	1
		Flexure	Flexure–Shear	Shear
		Predicted Values		

* See also Figure 8

Table 6: Performance metrics

	Performance Metrics *						
	True Positive	True Negative	False Positive	False Negative	Accuracy	Precision	Recall
Flexure	12	5 + 1 + 1 + 1 = 8	1 + 0 = 1	0 + 0 = 0	(12 + 8)/(12 + 8 +1 +0) = 20/21 = 95%	(12)/(12 +1) = 12/13 = 92%	(12)/(12 + 0) = 12/12 = 100%
Flexure–Shear	5	12 + 0 + 0 + 1= 13	0 + 1 =1	1 + 1 = 2	(5 + 13)/(5 + 13 +1 +2) = 18/21 = 95 %	(5)/(5 +1) = 5/6 = 83%	(5)/(5 +2) = 5/7 = 71%
Shear	1	12 +0 +1 +5 = 18	0 + 1 = 1	0 + 1 = 1	(1 + 18)/(1 + 18 +1 +1) = 19/21 = 90%	(1)/(1 +1) = 1/2 = 50%	(1)/(1 +1) = 1/2 = 50%

* See also Table 5.

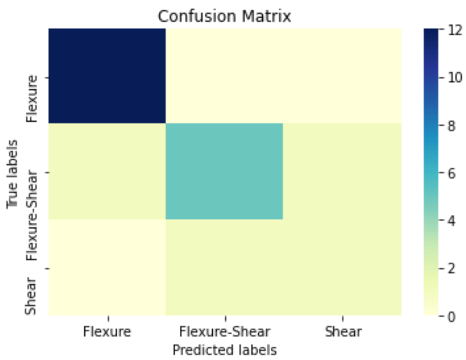

Figure 8: Confusion matrix as performance metric for ML prediction of the failure mode of circular RC columns of PEER structural performance database with random forest method.

Finally, in order to understand better how random forests builds its decision trees one of the decision trees of the process will be plotted for the cases of circular and rectangular RC columns displaying the decision-making process of a single tree within the ensemble in order to predict the failure mode of RC columns (Figures 9-12).

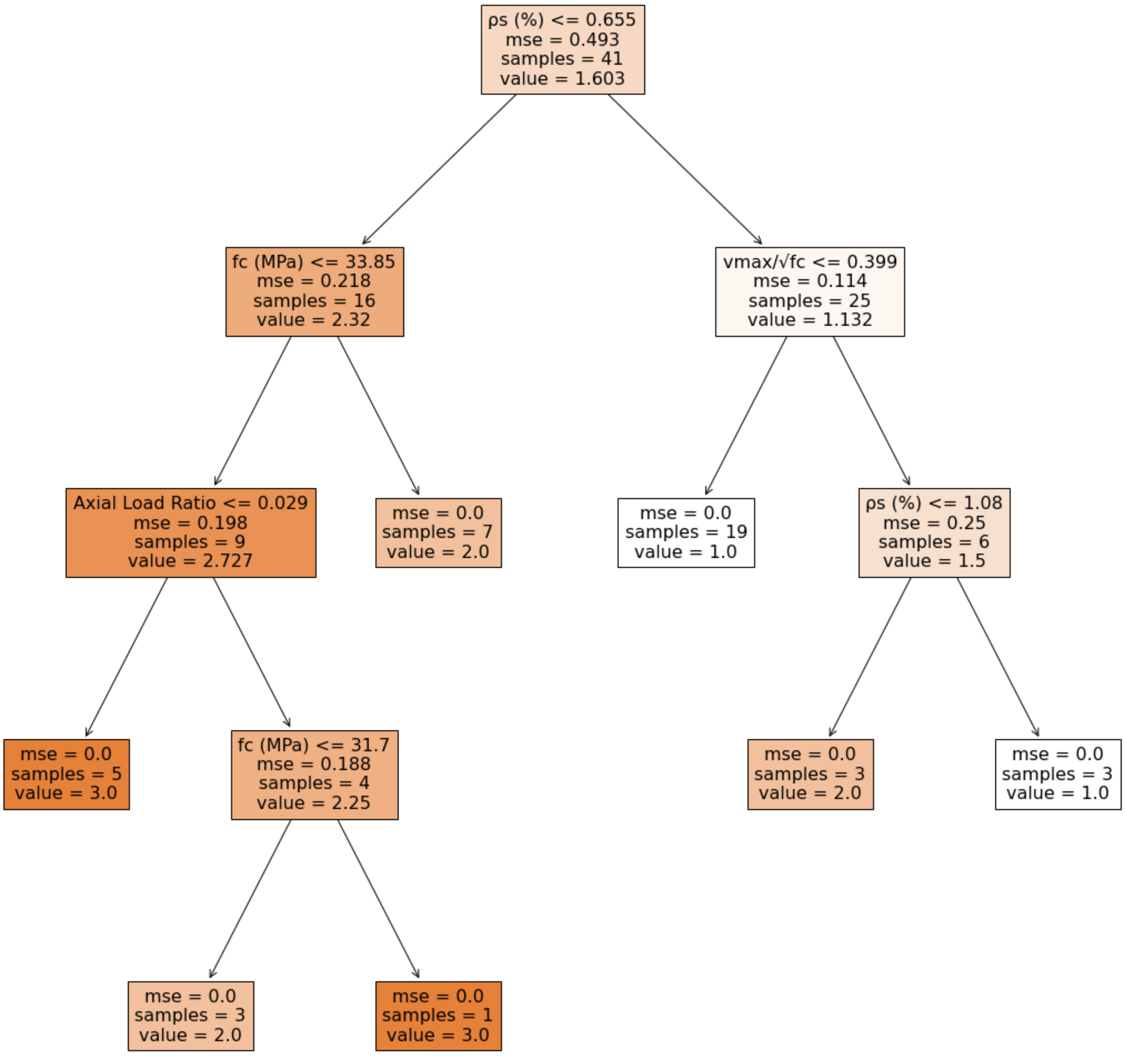

Figure 9: Decision-making process of a single tree within the ensemble of random forests for circular columns.

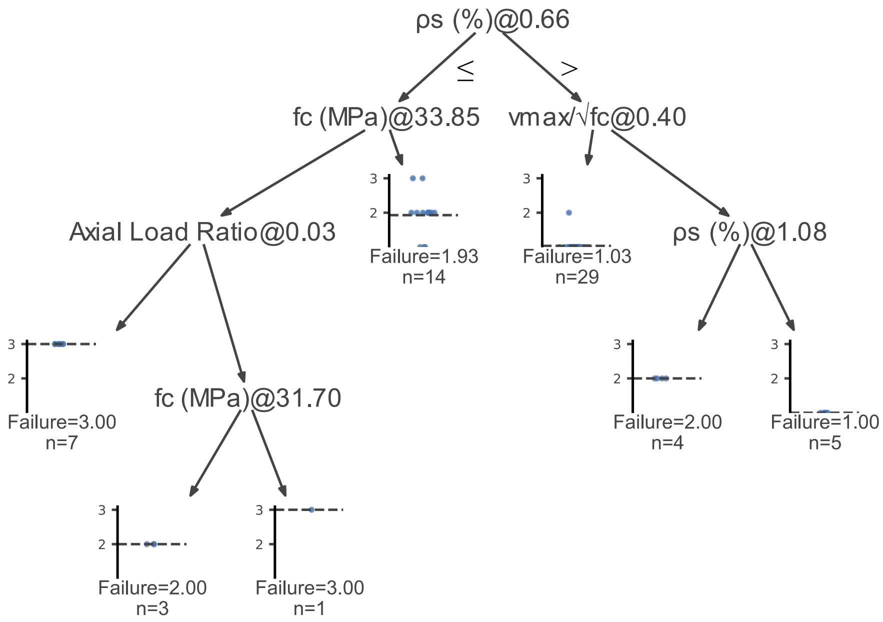

Figure 10: Decision-making process of a single tree within the ensemble of random forests for circular columns.

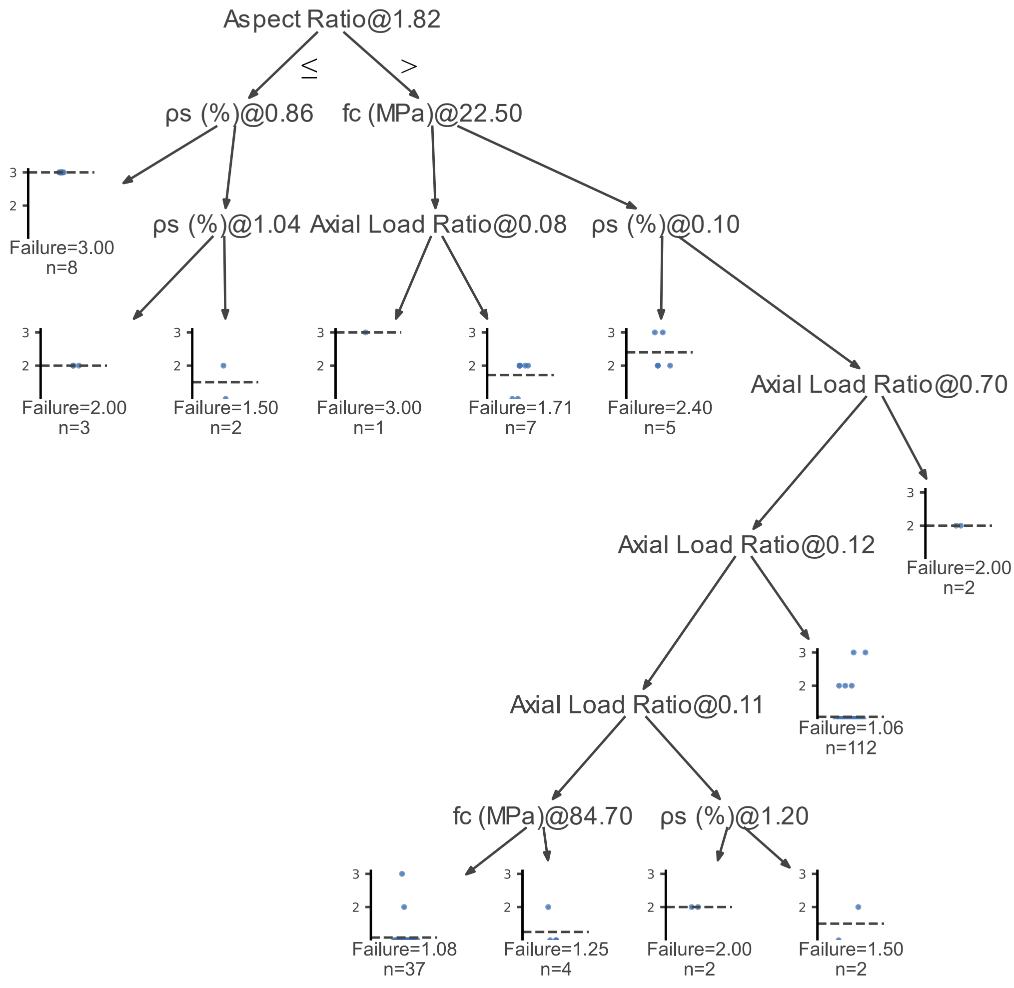

Figure 11: Decision-making process of a single tree within the ensemble of random forests for rectangular columns.

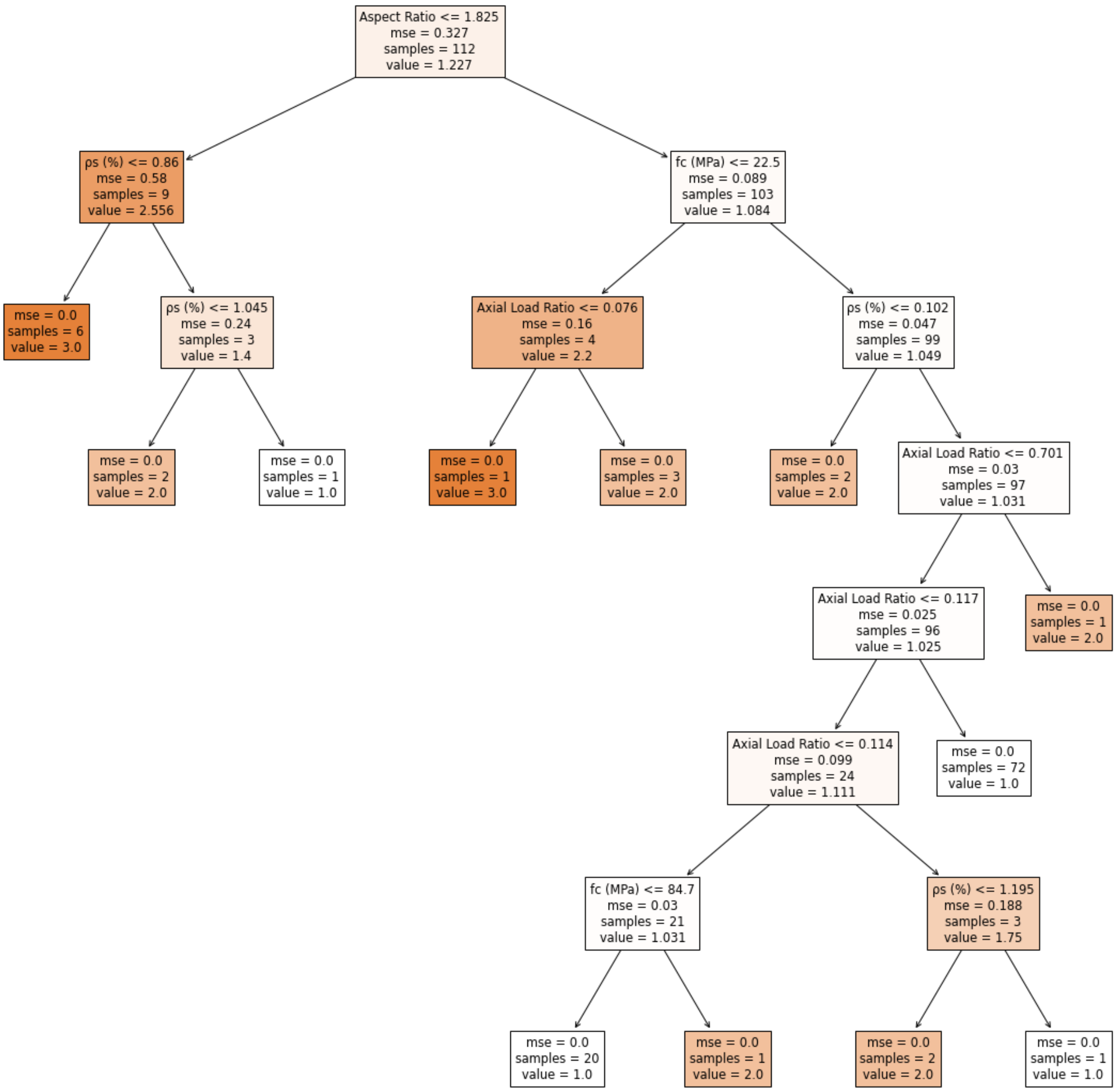

Figure 12: Decision-making process of a single tree within the ensemble of random forests for rectangular columns.

4 CONCLUSIONS

This paper makes the following contributions to the application of ML methods in earthquake engineering research:

- To the best of the authors knowledge, the PEER structural performance database is utilized for the first time to predict the failure modes of RC columns.
- Rectangular RC columns are analyzed for the first time in failure mode detection using the random forest ML method.
- A comprehensive investigation is conducted into the influence of key design variables on column ductility and failure modes.
- Lastly, all essential performance metrics for evaluating the ML methodology's effectiveness in predicting RC column failure modes are presented.

Predicting the failure modes of RC columns is essential for designing effective retrofitting solutions for modern buildings and bridges. Current approaches often rely on nonlinear structural analysis methods, which are time-intensive and require significant effort to achieve accuracy. This study investigates the potential of integrating physical knowledge with machine learning (ML) techniques to predict the failure modes of RC columns. Utilizing the PEER structural performance database, the study examines the impact of key design variables on column ductility and failure modes. The results demonstrate that supervised ML methods, such as random forests, can accurately classify failure modes when applied to a test set randomly drawn from the PEER database, particularly when enhanced with physical insights. The overall accuracy achieved is 94% for rectangular columns and 86% for circular columns while code-based prediction based also on nonlinear structural analysis achieves as it was demonstrated previously 97% and 81% respectively. This signifies that random forests outperformed traditional method based on engineering mechanics for the case of circular columns. These findings highlight the transformative potential of ML in advancing earthquake engineering. Notably, this study is the first to use the PEER structural performance database to identify RC column failure modes through supervised ML approaches. Additionally, the influence of column section geometry, often overlooked in previous research focused mainly on circular columns, is explored. The findings lay a foundation for future studies on other supervised ML techniques, including Decision Trees, k-Nearest Neighbors, Neural Networks, and Deep Learning, to further enhance failure mode detection for RC columns.

REFERENCES

[1] Megalooikonomou, K.G. Modeling the Behavior of Shear-Critical Reinforced Concrete Columns under Lateral Loads. Ph.D. Thesis, Department of Civil and Environmental Engineering, Faculty of Engineering, University of Cyprus, Nicosia, Cyprus, December, 2019. https://doi.org/10.12681/eadd/47504.

[2] Mitra, N.; Mitra, S.; Lowes, L.N. Probabilistic model for failure initiation of reinforced concrete interior beam-column connections subjected to seismic loading. Eng. Struct., 33, 154–162, 2011.

[3] Tang, Y.; Wang, Y.; Wu, D.; Liu, Z.; Zhang, H.; Zhu, M.; Chen, Z.; Sun, J.; Wang, X. An experimental investigation and machine learning-based prediction for seismic

performance of steel tubular column filled with recycled aggregate concrete. Rev. Adv. Mater. Sci., 61, 849–872, 2022.

[4] Tang, Y.; Wang, Y.; Wu, D.; Chen, M.; Pang, L.; Sun, J.; Feng, W.; Wang, X. Exploring temperature-resilient recycled aggregate concrete with waste rubber: An experimental and multi-objective optimization analysis. Rev. Adv. Mater. Sci., 62, 20230347, 2023.

[5] Berry, M.P.; Parrish, M.; Eberhard, M.O. PEER Structural Performance Database User's Manual; Version 1.0, Pacific Earthquake Engineering Research Center (PEER) Report; University of California: Berkeley, CA, USA, 2004. Available online: http://nisee.berkeley.edu/spd/ .

[6] EN 1998-3, Eurocode 8: Design of structures for earthquake resistance -Part 3: Assessment and retrofitting of buildings. European Committee for Standardization (CEN), Brussels, 2005.

[7] ASCE/SEI 41. Seismic Rehabilitation of Existing Buildings. American Society of Civil Engineers, Reston, VA, 2007.

[8] Filippou, F. C., Constantinides, M. FEDEAS Lab – Getting Started Guide and Simulation Examples, NEESgrid Report 2004-22 and SEMM Report 2004-05, 2004. Available online: https://fedeas.net/ .

[9] Filippou, F.C.; Fenves, G.L. Methods of analysis for earthquake-resistant structures. In Earthquake Engineering: From Engineering Seismology to Performance-Based Engineering; Bozorgnia, Y., Bertero, V.V., Eds.; CRC Press: Boca Raton, FL, USA, 2004.

[10] MATLAB: User's Guide, Version R2024b; Mathworks Inc.: 2024, Natick, Massachusetts, U.S.A.

[11] The FEMA 440 Report. Improvement of Nonlinear Static Seismic Analysis Procedures, Federal Emergency Management Agency (FEMA), Washington, DC, USA, p. 392, 2005.

[12] Scott, B.D., Park, R., and Priestley, M.J.N. Stress-strain behavior of concrete confined by overlapping hoops at low and high strain rates. American Concrete Institute Journal, 79, 13-27, 1982.

[13] Mander, J. B., Priestley, M. J. N., and Park, R. Theoretical stress-strain model for confined concrete. Journal of Structural Engineering ASCE, 114(8), 1804–1826, 1988.

[14] Menengotto, M. and Pinto E. Method of Analysis for Cyclically Loaded Reinforced Concrete Plane Frames Including Changes in Geometry and Nonelastic Behavior of Elements under Combined Normal Force and Bending. In IABSE Symposium on Resistance and Ultimate Deformability of Structures Acted on by Well-Defined Repeated Loads, Final Report, Lisbon, Portugal 1973.

[15] Megalooikonomou KG. Random Forests Machine Learning Applied to PEER Structural Performance Experimental Columns Database, Master Thesis in Information Systems, School of Science and Technology, Hellenic Open University, Patras, Greece, 2025, in press.

[16] Megalooikonomou KG, Beligiannis GN Random Forests Machine Learning Applied to PEER Structural Performance Experimental Columns Database, Applied Sciences, MDPI, 13(23):12821, 2023. https://doi.org/10.3390/app132312821