

COMPARISON BETWEEN MACHINE LEARNING PREDICTIONS AND EUROCODE 8 EMPIRICAL EXPRESSIONS FOR THE PLASTIC ROTATION CAPACITY OF RECTANGULAR RC COLUMNS

Andrei-Odey KADHIM - PhD Candidate, Technical University of Civil Engineering Bucharest, Faculty of Civil, Industrial and Agricultural Buildings, e-mail: andrei.kadhim@utcb.ro

Iolanda-Gabriela CRAIFALEANU - Associate Professor, Dr. Habil., Technical University of Civil Engineering Bucharest, Faculty of Civil, Industrial and Agricultural Buildings, e-mail: iolanda.craifaleanu@utcb.ro

Abstract: This paper presents a comparative assessment of the plastic rotation capacity of rectangular reinforced concrete columns, using machine learning predictions and the empirical expressions provided by Eurocode 8, Part 3. The study relies on a database of 258 rectangular RC column specimens subjected to quasi-static cyclic lateral loading, covering a wide range of reinforcement configurations, axial load levels, shear span-to-depth ratios and failure modes. A previously developed and validated Stacking Ensemble model, combining Support Vector Regression, Random Forest and XGBoost through a RidgeCV meta-regressor, is used as the data-driven reference model. Its predictions are compared with the values obtained using expressions A.3 and A.4 from EN 1998-3:2005, evaluated on a subset of 52 rectangular specimens for which all parameters required by the code expressions were available. The comparison is performed in terms of the coefficient of determination, mean absolute error and root mean squared error, together with a residual error analysis. The results show that the Stacking Ensemble model provides a closer agreement with the experimental values, while the code-based expressions exhibit larger dispersion and a more conservative response, particularly when the partial safety factor is considered. The findings indicate that data-driven models can complement empirical code provisions by providing an additional quantitative reference for the seismic assessment of existing rectangular RC columns.

Keywords: reinforced concrete columns, plastic rotation capacity, Eurocode 8, machine learning, empirical expressions, seismic assessment

1. Introduction

In the seismic assessment of existing reinforced concrete buildings, the deformation capacity of structural members becomes a governing parameter once the response exceeds the elastic range. For columns subjected to cyclic lateral loading, this capacity is commonly expressed through the plastic rotation capacity, which describes the ability of the member to sustain inelastic deformations before significant strength degradation or brittle failure occurs. This parameter is directly related to local ductility, to the energy dissipation capacity of the structural system and to the reliability of performance-based assessment procedures [1].

For members subjected to bending, plastic deformations concentrate in the regions of maximum bending moment, where plastic hinges form. The behaviour of these regions is controlled by the simultaneous action of several parameters, including the shear span, the longitudinal and transverse reinforcement ratios, the reinforcement yield strengths, the concrete compressive strength, the axial load ratio and the governing failure mode. Experimental investigations have shown that these parameters do not act independently: high axial load ratios increase the depth of the compression zone and can reduce the rotation capacity, while effective transverse reinforcement improves confinement, delays the buckling of longitudinal bars and stabilizes the post-yield response. The extent of the plastic hinge itself depends on the loading history, cyclic

loading leading to longer effective plastic hinges than monotonic loading [2], as well as on the interaction between member proportions, reinforcement ratio and failure mechanism [3].

Analytical and semi-empirical models remain essential in engineering practice because they are transparent and directly applicable in design and assessment calculations. The plastic rotation can be derived by integrating the plastic curvature over an equivalent plastic hinge length, over which the plastic curvature is assumed constant [1], [3]. For the assessment of existing buildings, EN 1998-3:2005 [4] provides, in Annex A, empirical and semi-empirical expressions for the total and plastic chord rotation capacity of RC members, calibrated on large databases of cyclic tests [5]. These expressions are widely used in practice, as they condense the influence of the main governing parameters into closed-form relationships suitable for code-based verifications.

However, several studies have shown that the application of such expressions may be affected by considerable scatter when their estimates are compared with measured deformation capacities. Mpampatsikos et al. [6] critically reviewed the assessment procedures of Eurocode 8 and of the Italian seismic code for existing RC frame buildings, highlighting the sensitivity of the code-based verifications to the adopted deformation capacity models. Popa and Coțofană [7] compared the experimental drift capacity of seven 300×300 mm RC columns, tested under cyclic lateral loading with axial load ratios between 0.2 and 0.4 and transverse reinforcement ratios between 0.1% and 0.6%, with the estimates of three empirical approaches, including a formulation similar to that adopted in Eurocode 8, and reported considerable differences between the empirical values and the experimental observations. Melo et al. [8] reached similar conclusions for columns reinforced with plain bars and proposed an improvement of the Eurocode 8 formulation for this category of members. This scatter is largely a consequence of the strong nonlinear relationship between the input parameters and the deformation capacity, which regression-based formulations with predefined functional forms can capture only partially.

In this context, machine learning (ML) techniques have been increasingly investigated as complementary, data-driven tools for the prediction of structural response, as they can identify nonlinear dependencies directly from experimental data, without imposing an explicit functional form [9]. Luo and Paal [10] proposed a locally weighted Support Vector Machine model for estimating the drift capacity of RC columns for different failure modes, reporting improved predictions compared with both a conventional SVM model and an empirical formula used in seismic vulnerability assessment. The same authors subsequently applied machine learning to the seismic response prediction of RC frames, obtaining results comparable to fibre-based modelling at a fraction of the computational cost [11]. Aladsani et al. [12] developed an explainable ML model for the drift capacity of RC walls, using SHAP values to quantify the influence of the input parameters, while Yahyapour et al. [13] combined unsupervised and supervised learning for the multi-criteria seismic damage evaluation of RC columns. These studies indicate that ML models can serve not only as prediction tools but also as references for evaluating the dispersion associated with simplified empirical expressions.

The present paper continues a research programme dedicated to the prediction of the deformation capacity of RC members using machine learning. In an initial study, six algorithms were compared on datasets of RC columns and walls, with encouraging results for ductile failure modes but reduced accuracy for brittle failures, largely attributed to unbalanced datasets [14]. In a subsequent study, the rectangular column database was extended with additional specimens, the models were optimized through grid search within a nested cross-validation framework, model interpretability was

addressed through SHAP analysis, and a Stacking Ensemble model was developed, combining Support Vector Regression (SVR), Random Forest (RF) and XGBoost through a RidgeCV meta-regressor [15].

The objective of the present paper is to compare the plastic rotation capacity predicted by this previously validated data-driven model with the values obtained using the empirical expressions of EN 1998-3:2005, on a common subset of rectangular RC column specimens. The study does not introduce a new machine learning algorithm. Instead, the Stacking Ensemble model previously developed [15] is used as a reference prediction model, and its results are compared with expressions A.3 and A.4 from EN 1998-3, evaluated with and without the partial factor γ_{el} . The comparison focuses on global performance indicators and on residual error distributions, in order to quantify the accuracy and the degree of conservatism of the analysed approaches.

2. Rectangular RC column database and selected parameters

The database considered in this paper consists exclusively of rectangular reinforced concrete columns and comprises 258 specimens subjected to quasi-static cyclic lateral loading. The core of the database, extracted from the PEER Structural Performance Database [16] was expanded to improve the balance between flexural, mixed and shear-dominated failure mechanisms, which was identified as the main limitation of the initial dataset, 86 additional rectangular specimens were collected from recent experimental studies and added to the original set, as described in detail in [15].

The final rectangular column dataset includes 147 specimens with predominant flexural failure, 67 specimens with combined flexure-shear failure, and 44 specimens with brittle failure governed mainly by shear. For the comparison with EN 1998-3:2005, a reduced subset of 52 specimens was used, for which all additional reinforcement and section data required by the code expressions could be determined. This subset was selected according to the availability of the parameters needed for expressions A.3 and A.4 and is distinct from the random 20% test subset used for evaluating the machine learning models. The composition of the two datasets is summarized in Table 1.

Table 1 - Composition of the rectangular RC column database and of the Eurocode comparison subset

Dataset	Total specimens	Flexure	Flexure-shear	Shear	Use in the present study
Full rectangular column database	258	147	67	44	Machine learning model development and reference predictions
Eurocode comparison subset	52	45	5	2	Direct comparison between Stacking Ensemble and EN 1998-3 expressions

The input parameters were selected based on their direct mechanical influence on the plastic rotation capacity of RC columns and on their use in existing empirical models. These parameters include the concrete compressive strength f_c , the yield strengths of the longitudinal and transverse reinforcement f_y and f_{yw} , the longitudinal and transverse reinforcement ratios ρ_l and ρ_w , the shear span-to-effective depth ratio L/d , the axial load ratio $P/A_g f_c$, and the observed failure mode. The failure mode was introduced as a categorical variable and converted into binary indicators corresponding to flexural, shear and flexure-shear failure.

The target variable was the plastic rotation capacity, expressed as a percentage. When this value was not explicitly reported in the original source, it was estimated from the measured displacement at the point of failure divided by the shear span or by the effective column height. For ductile specimens without significant strength degradation, the maximum displacement on the hysteretic envelope was used. For brittle or combined failure modes, the displacement corresponding to a drop of the lateral resistance to 80% of the maximum attained value was adopted, following the usual interpretation of strength degradation in cyclic column tests [10], [17]. The statistical ranges of the selected parameters for the full rectangular column database are presented in Table 2.

The statistical ranges indicate a relatively high variability of the database, especially for the concrete compressive strength, transverse reinforcement ratio, and axial load ratio. The large coefficient of variation of ρ_w reflects the diversity of transverse reinforcement configurations among the experimental programs. This is important because confinement effectiveness is not governed only by the amount of transverse reinforcement, but also by spacing, arrangement, bar diameters, longitudinal reinforcement layout, axial load level, and shear span.

Table 2 - Statistical ranges of the selected parameters for the rectangular RC column database

Property	Minimum	Maximum	Average	Standard deviation	Coefficient of variation
f_c [MPa]	13.5	116	41.65	25.3	60.7%
f_{yw} [MPa]	249	1126	427.6	117.7	27.5%
f_y [MPa]	295	609	428.59	65.1	15.2%
ρ_l	0.007	0.060	0.020	0.010	50%
ρ_w	0.002	0.210	0.010	0.030	300%
L/d	0.50	6.04	3.03	1.39	45.8%
$P/A_g f_c$	0.00	0.90	0.28	0.19	67.8%
Plastic rotation capacity [%]	0.61	9.48	3.43	1.93	56.3%

This variability also illustrates why the estimation of plastic rotation capacity cannot be reduced to the influence of a single governing parameter. In rectangular RC columns, similar values of one variable may lead to different deformation capacities depending on the combined effect of axial load, shear span, confinement detailing, and failure mode. Therefore, the selected input variables should be interpreted as descriptors of interacting mechanisms that jointly control the plastic deformation capacity. This aspect supports the use of nonlinear data-driven models, which can account for combined parameter effects that are difficult to represent through simple empirical trends.

The comparison with EN 1998-3:2005 expressions was not performed on the entire rectangular column database, but on the subset of specimens for which all additional parameters required by expressions A.3 and A.4 were available. This distinction is important because the empirical expressions require more detailed information than the machine learning model, especially regarding the distribution of longitudinal reinforcement, transverse reinforcement arrangement, confinement effectiveness and

section-level quantities used for the calculation of curvature-related terms. Consequently, the Eurocode comparison subset is smaller than the full database and has a different failure-mode composition.

As indicated in Table 1 and further illustrated in Fig. 1, the filtering required for the EN 1998-3:2005 comparison reduces both the total number of specimens and the number of specimens available for each failure mode. The figure reports the absolute number of specimens, while the percentages indicate the relative contribution of each failure mode within the corresponding dataset. In the full rectangular database, flexure-controlled specimens represent the largest category, but mixed and shear-dominated failures are still present in relevant proportions. In contrast, the Eurocode comparison subset is dominated by flexural specimens, while flexure-shear and shear-dominated failures are retained only in limited numbers.

This difference affects the interpretation of the subsequent results. The comparison between the Stacking Ensemble model and the EN 1998-3:2005 expressions is mainly representative of columns with flexure-controlled behaviour and, to a lesser extent, of mixed flexure-shear response. It is less representative of brittle shear failures, which remain underrepresented in the comparison subset. This limitation is not caused by the machine learning model itself, but by the additional data requirements of the empirical expressions and by the incomplete availability of detailed reinforcement and sectional information for some experimental specimens.

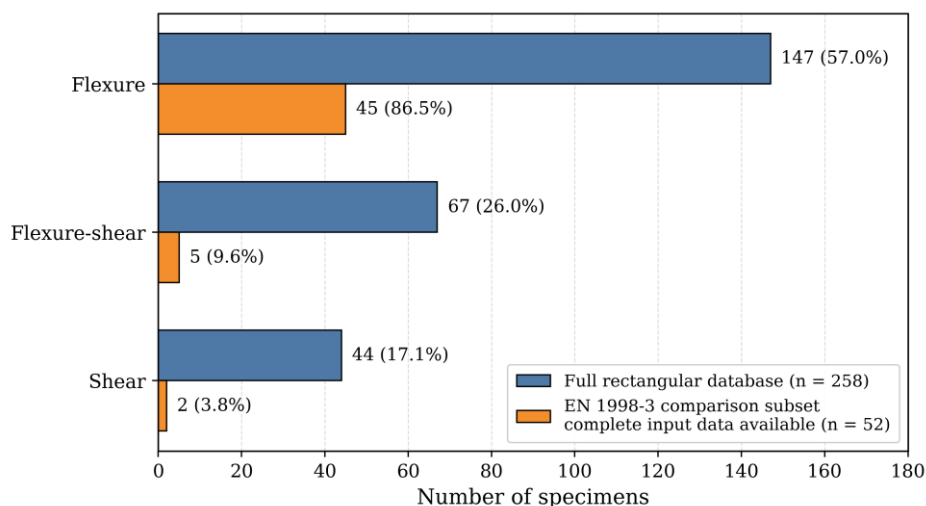


Fig. 1 - Failure-mode composition of the full rectangular column database and of the EN 1998-3 comparison subset, expressed as absolute counts with the corresponding percentages

The reduced number of shear-dominated specimens is relevant because this failure mode is associated with lower deformation capacity and more abrupt strength degradation. In previous stages of the research, the largest residual errors of the machine learning models were also concentrated mainly in specimens with brittle shear failure, which remained the least represented category even after the extension of the rectangular column database. This confirms that the accurate prediction of non-ductile response is strongly dependent on the availability of sufficiently representative experimental data.

For this reason, the results obtained in the following sections should be interpreted as a comparison between data-driven and code-based estimates for a subset of rectangular RC columns in which flexural behaviour is predominant. The conclusions regarding the conservatism and dispersion of the EN 1998-3 expressions are therefore

directly applicable to this type of response, while additional validation would be necessary for a more robust assessment of shear-dominated columns.

3. Machine learning models and optimized configurations

The machine learning models used as the basis for the present comparison were implemented, optimized, and validated by the authors within the previously developed data-driven framework for predicting the plastic rotation capacity of RC columns [15]. The present paper does not propose a new machine learning architecture; instead, it uses these implemented and validated models as data-driven reference predictors for comparison with the EN 1998-3:2005 empirical expressions. For this reason, the models retained for the analysis were Support Vector Regression, Random Forest, XGBoost, and the Stacking Ensemble model.

The selected algorithms represent different regression strategies. Support Vector Regression is based on the support vector learning framework and, when used with a nonlinear kernel, can approximate smooth nonlinear relationships while controlling model complexity through regularization [18]. Random Forest relies on the aggregation of multiple decision trees, which reduces the sensitivity of individual trees and allows local interactions between variables to be captured through an ensemble averaging process [19]. XGBoost is a gradient boosting algorithm in which successive regression trees are added to correct the residual errors of the previous ensemble, with regularization and subsampling used to limit overfitting [20].

The implementation was carried out using Python-based machine learning libraries. Support Vector Regression, Random Forest, and the RidgeCV meta-regressor were implemented using scikit-learn [22], while XGBoost was implemented using the corresponding gradient boosting library [20]. The workflow included preprocessing of the input variables, encoding of the categorical failure mode variable, random train-test splitting, hyperparameter tuning through grid search within a nested cross-validation framework, retraining of the optimized models on the complete training set, and final evaluation on the independent test set.

The workflow of the comparative assessment is summarized in Fig. 2. The data-driven branch is based on the random train-test split used for machine learning model evaluation, whereas the code-based branch is based on a distinct subset selected according to the availability of all parameters required by the EN 1998-3:2005 expressions.

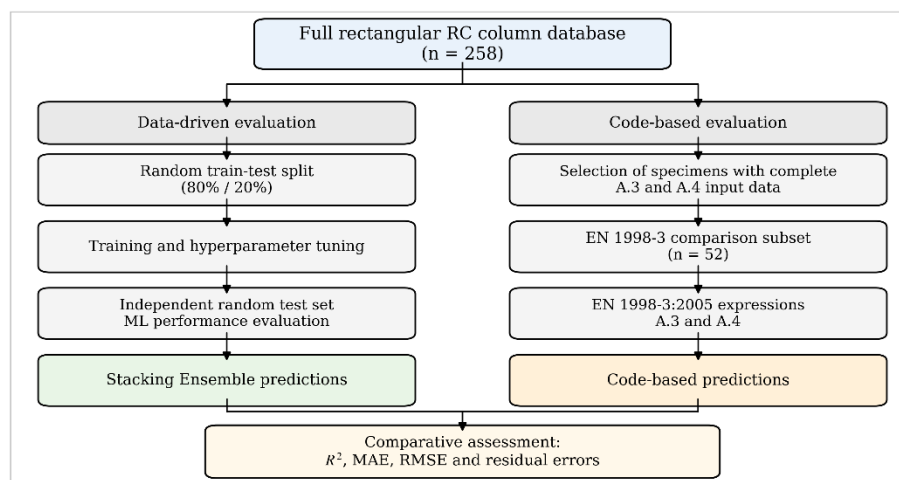


Fig. 2 - Workflow of the comparative assessment, distinguishing between the random train-test split used for machine learning evaluation and the distinct subset with complete input data used for the EN 1998-3 comparison

The optimized configurations used for the rectangular column models are summarized in Table 3. The hyperparameters were obtained through grid search integrated into the nested cross-validation procedure [15]. For SVR, the selected regularization and loss parameters reflect the need to balance tolerance to errors and model complexity in a database characterized by significant dispersion of the plastic rotation capacity. For Random Forest, the use of 200 estimators compensates for the unrestricted depth of the individual trees through the averaging effect of bagging. For XGBoost, the low learning rate, limited tree depth, and subsampling define a stable configuration, appropriate for a dataset with moderate size and high variability.

Table 3 - Optimized hyperparameters of the machine learning models used for rectangular RC columns

Model	Optimized hyperparameters
SVR	$C = 10$; kernel = RBF; $\gamma = \text{auto}$; $\epsilon = 0.2$
Random Forest	$n_estimators = 200$; $max_depth = \text{None}$ (unrestricted); $min_samples_split = 2$
XGBoost	$n_estimators = 200$; $max_depth = 3$; $learning_rate = 0.05$; $subsample = 0.8$
Stacking Ensemble	Base models: SVR + Random Forest + XGBoost in the optimized configurations above; meta-model: RidgeCV

The Stacking Ensemble model combines the predictions of SVR, Random Forest and XGBoost as input features for a RidgeCV meta-regressor. This architecture follows the principle of stacked generalization, in which a second-level model learns how to combine the outputs of several first-level predictors [21]. In the present application, the ensemble exploits the complementary behaviour of the base regressors: SVR captures smooth nonlinear trends, Random Forest provides robustness to local interactions and noise, while XGBoost improves the representation of complex multivariable relationships.

RidgeCV was selected as the meta-regressor because it provides a regularized linear combination of the base-model predictions. Since the outputs of the three regressors are trained to approximate the same target variable, their predictions may be correlated. The L2 regularization introduced by RidgeCV reduces the effect of this multicollinearity and improves the stability of the final ensemble prediction. In addition, the regularization parameter is selected internally through cross-validation, avoiding an additional tuning step at the meta-model level [22]. This configuration makes the Stacking Ensemble suitable as a stable data-driven reference model for the subsequent comparison with the EN 1998-3:2005 expressions.

4. Data-driven reference model

Following the model configurations described in the previous section, the Stacking Ensemble developed and validated in [15] was used as the data-driven reference for the comparison with EN 1998-3:2005. The model combines the predictions of SVR, Random Forest and XGBoost through a RidgeCV meta-regressor, following the principle of stacked generalization [21]. In this section, the focus is placed on the validation procedure and on the predictive performance obtained for the rectangular column dataset. All models were implemented using the scikit-learn library [22].

The base regressors were optimized through grid search integrated in a nested cross-validation framework, in which the hyperparameters were tuned within each training split and the generalization performance was estimated as the average coefficient of determination over five outer folds. After hyperparameter tuning, the final models were retrained on the complete training set and evaluated on an independent test set representing 20% of the database. The performance of the individual and ensemble models is presented in Table 4.

Table 4 - Performance of the optimized machine learning models on the rectangular column dataset

Model	Nested CV R^2	Test R^2	MAE [%]	RMSE [%]
SVR	0.59	0.83	0.64	0.75
Random Forest	0.65	0.77	0.69	0.80
XGBoost	0.71	0.80	0.68	0.84
Stacking Ensemble	-	0.82	0.51	0.82

For rectangular columns, SVR obtained the best individual test-set performance in terms of R^2 and RMSE, while XGBoost obtained the highest average R^2 during nested cross-validation, indicating the most consistent generalization across the outer folds. The Stacking Ensemble was selected as the reference model for the present comparison because it provided the most balanced global performance, obtaining a coefficient of determination very close to that of SVR, together with the lowest mean absolute error among all analysed models. The corresponding observed-versus-predicted representation is shown in Fig. 3, illustrating the level of precision obtained by the selected reference model on the rectangular column test set.

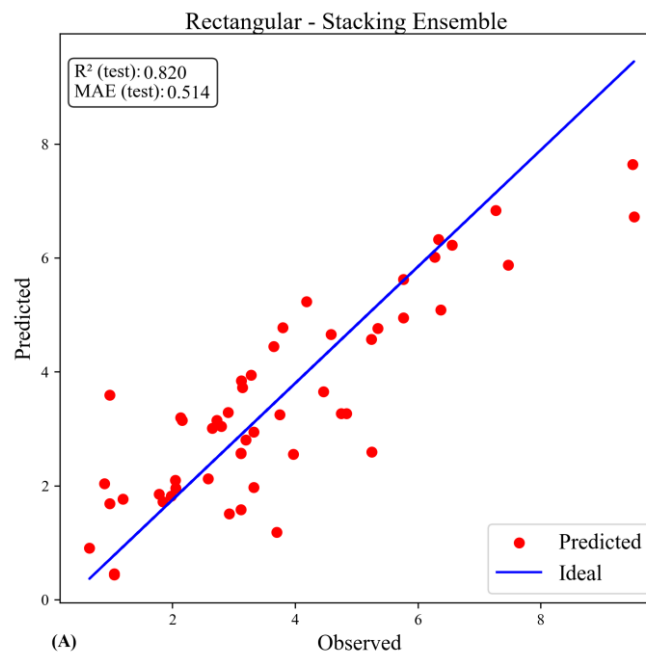


Fig. 3 - Predicted versus observed plastic rotation capacity for rectangular columns using the Stacking Ensemble model

5. Eurocode 8 empirical expressions considered for comparison

The empirical expressions considered in this paper are those provided in Annex A of EN 1998-3 for the estimation of chord rotation and plastic rotation capacity [4]. Expression A.3 directly provides the plastic part of the ultimate chord rotation and was implemented in the following form:

$$\theta_{um}^{pl} = \theta_{um} - \theta_y = \frac{1}{\gamma_{el}} 0.0145 (0.25^v) \left[\frac{\max(0.01; \omega')}{\max(0.01; \omega)} \right]^{0.3} f_c^{0.2} \left(\frac{L_v}{h} \right)^{0.35} 25^{\alpha \rho_{sx} \frac{f_{yw}}{f_c}} (1.275^{100 \rho_d}) \quad (1)$$

where:

γ_{el} is the partial factor, equal to 1.8 for primary seismic elements and to 1.0 for secondary seismic elements;

$v = N/(b \cdot h \cdot f_c)$ is the normalized axial force;

ω and ω' are the mechanical reinforcement ratios of the tension and compression longitudinal reinforcement, respectively;

f_c is the concrete compressive strength, in MPa;

L_v is the shear span and h is the depth of the cross-section;

$\rho_{sx} = A_{sx}/(b \cdot s_h)$ is the ratio of transverse reinforcement parallel to the loading direction;

f_{yw} is the yield strength of the transverse reinforcement;

α is the confinement effectiveness factor, computed according to equation (A.2) of EN 1998-3;

ρ_d is the ratio of diagonal reinforcement, equal to zero in its absence.

Expression A.4 estimates the total ultimate chord rotation as the sum of the yield rotation and of the plastic contribution associated with the difference between the ultimate and yield curvatures. Since the target variable of the machine learning model is the plastic rotation capacity, the comparison was performed using the equivalent plastic component of A.4, obtained by excluding the yield rotation term θ_y :

$$\theta_{um}^{pl} = \frac{1}{\gamma_{el}} (\phi_u - \phi_y) L_{pl} \left(1 - \frac{0.5 L_p}{L_v} \right) \quad (2)$$

where ϕ_u is the ultimate curvature and ϕ_y is the yield curvature at the end section of the member. The equivalent plastic hinge length was determined using the corresponding expression provided in Annex A of EN 1998-3 for use with equation A.4:

$$L_p = 0.1 L_v + 0.17 h + 0.24 \frac{d_{bl} f_y}{\sqrt{f_c}} \quad (3)$$

where d_{bl} is the diameter of the tension reinforcement bars and the material strengths are expressed in MPa.

The expressions were evaluated for two values of the partial factor γ_{el} . First, $\gamma_{el} = 1.0$ was considered, in order to obtain a direct comparison with the experimental values and with the machine learning predictions, which do not intrinsically include a safety reduction. Second, $\gamma_{el} = 1.8$ was used to evaluate the conservative form associated with primary seismic elements in the SR EN 1998-3 assessment framework adopted in this study. The same EN 1998-3 comparison subset of 52 rectangular specimens

was used for all code-based calculations and for the direct comparison with the corresponding Stacking Ensemble predictions.

6. Results and discussion

The comparison between the Stacking Ensemble predictions and the EN 1998-3:2005 empirical expressions is summarized in Table 5. The Stacking Ensemble model obtained $R^2 = 0.82$, MAE = 0.51%, and RMSE = 0.82% on the 52-specimen comparison subset. These values indicate a good capacity to reproduce the variability of the experimental plastic rotation capacity, despite the heterogeneity of the database.

The best result among the empirical formulations was obtained with expression A.3 using $\gamma_{el} = 1.0$. In this case, the coefficient of determination was $R^2 = 0.103$, while MAE and RMSE were 1.255% and 1.710%, respectively. Although this unreduced form was closer to the experimental values than the version with $\gamma_{el} = 1.8$, its predictive performance remained considerably lower than that of the Stacking Ensemble model. When $\gamma_{el} = 1.8$ was introduced, the estimates were reduced, and the deviation from the experimental values increased, leading to $R^2 = -0.825$, MAE = 1.828%, and RMSE = 2.44%. The negative R^2 value indicates that, for this subset, the reduced expression performs worse than using the mean of the experimental values as a global predictor.

The equivalent plastic component derived from expression A.4 did not improve the comparison with the experimental values. For $\gamma_{el} = 1.0$, the resulting metrics were $R^2 = -0.012$, MAE = 1.464% and RMSE = 1.822%, which is slightly weaker than the corresponding A.3 result. For $\gamma_{el} = 1.8$, the performance further decreased, with $R^2 = -0.930$, MAE = 2.066%, and RMSE = 2.516%.

Table 5 - Comparison between the Stacking Ensemble model and EN 1998-3 expressions on the 52-specimen rectangular column subset

Model / Expression	γ_{el}	R^2	MAE [%]	RMSE [%]
Stacking Ensemble	-	0.820	0.510	0.820
EN 1998-3 A.3	1.0	0.103	1.255	1.710
EN 1998-3 A.3	1.8	-0.825	1.828	2.440
EN 1998-3 A.4, plastic component	1.0	-0.012	1.464	1.822
EN 1998-3 A.4, plastic component	1.8	-0.930	2.066	2.516

To complement the numerical comparison presented in Table 5, Figs. 4 and 5 summarize the global performance of the analysed approaches in terms of MAE, RMSE, and R^2 . These graphical representations make the relative differences between the Stacking Ensemble model and the EN 1998-3:2005 expressions more explicit and provide a clearer overview of the increase in error and reduction in predictive capability associated with the empirical formulations.

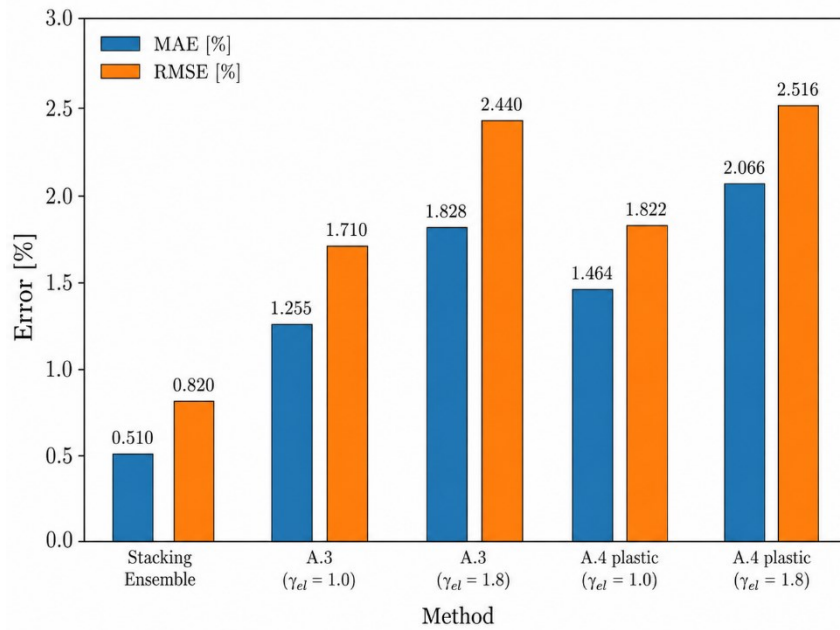


Fig. 4 - Comparison of MAE and RMSE values obtained for the Stacking Ensemble model and the EN 1998-3:2005 empirical expressions

As shown in Fig. 4, the Stacking Ensemble model clearly provides the lowest error values, while all EN 1998-3-based estimates lead to larger MAE and RMSE values. The increase in error is especially pronounced for the reduced forms with $\gamma_{el} = 1.8$.

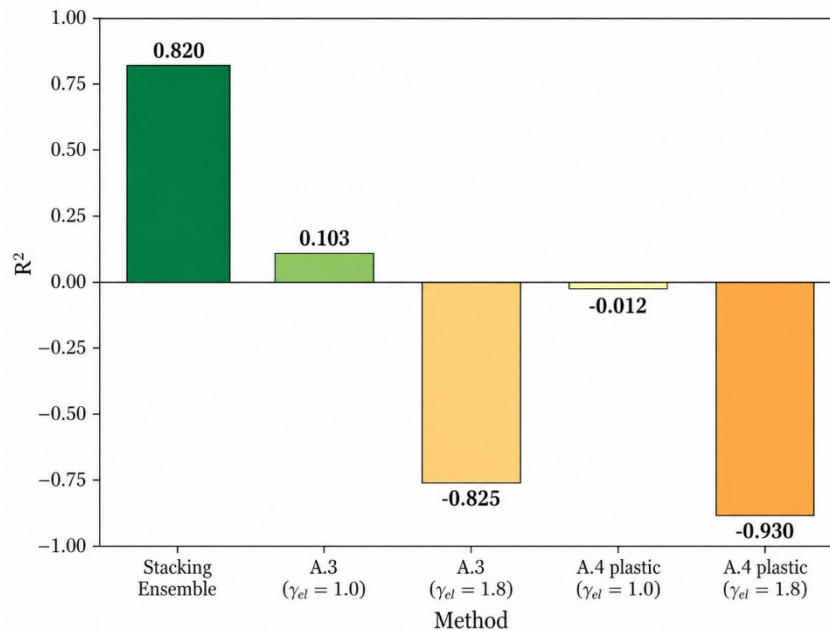


Fig. 5 - Comparison of R^2 values obtained for the Stacking Ensemble model and the EN 1998-3:2005 empirical expressions.

Fig. 5 confirms the same tendency in terms of the coefficient of determination. While the Stacking Ensemble model retains a high positive R^2 value, the empirical expressions show limited or negative correlation with the experimental values, particularly when the reduction associated with $\gamma_{el} = 1.8$ is considered.

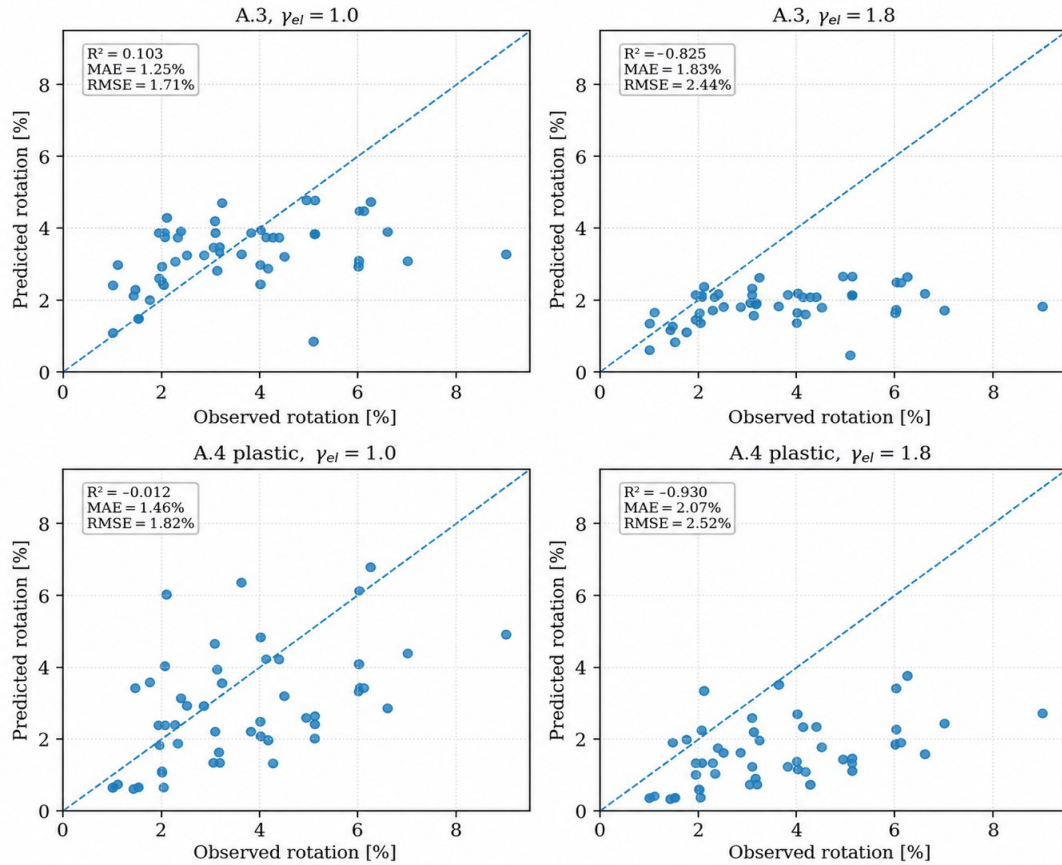


Fig. 6 - Observed versus predicted plastic rotation capacity obtained using expressions A.3 and A.4 from EN 1998-3, evaluated for $\gamma_{el} = 1.0$ and $\gamma_{el} = 1.8$

Fig. 6 illustrates the dispersion obtained when the code expressions are compared with the experimental values. For expression A.3 with $\gamma_{el} = 1.0$, the predictions partially follow the general trend in the range of moderate rotations; however, specimens with observed plastic rotations larger than approximately 5% are frequently underestimated. The use of $\gamma_{el} = 1.8$ shifts the predictions towards lower values, which confirms the conservative character of the reduced expression and increases the distance from the experimental values. A similar trend is observed for the plastic component derived from A.4. The unreduced variant exhibits broad scatter without improving the global correlation, while the reduced variant accentuates the underestimation of high-capacity specimens.

The residual error distributions shown in Fig. 7 complement the numerical comparison. The residuals of the Stacking Ensemble model are concentrated around zero, with reduced spread and without a pronounced positive tail. For expression A.3 with $\gamma_{el} = 1.0$, 17 out of the 52 specimens exhibit residual errors larger than 1.5 percentage points in absolute value, and the positive tail of the distribution shows that the expression tends to underestimate some of the largest observed rotations. For the plastic component derived from A.4 with $\gamma_{el} = 1.0$, the number of specimens with residuals exceeding 1.5 percentage points increases to 25. Compared with A.3, the A.4 distribution extends towards both negative and positive values, confirming that the curvature-based plastic component does not reduce the dispersion for the analysed subset. A possible explanation is the sensitivity of the curvature terms and of the plastic hinge length estimate to the sectional assumptions adopted in the calculation, which introduces additional sources of uncertainty compared with the fully empirical formulation of A.3.

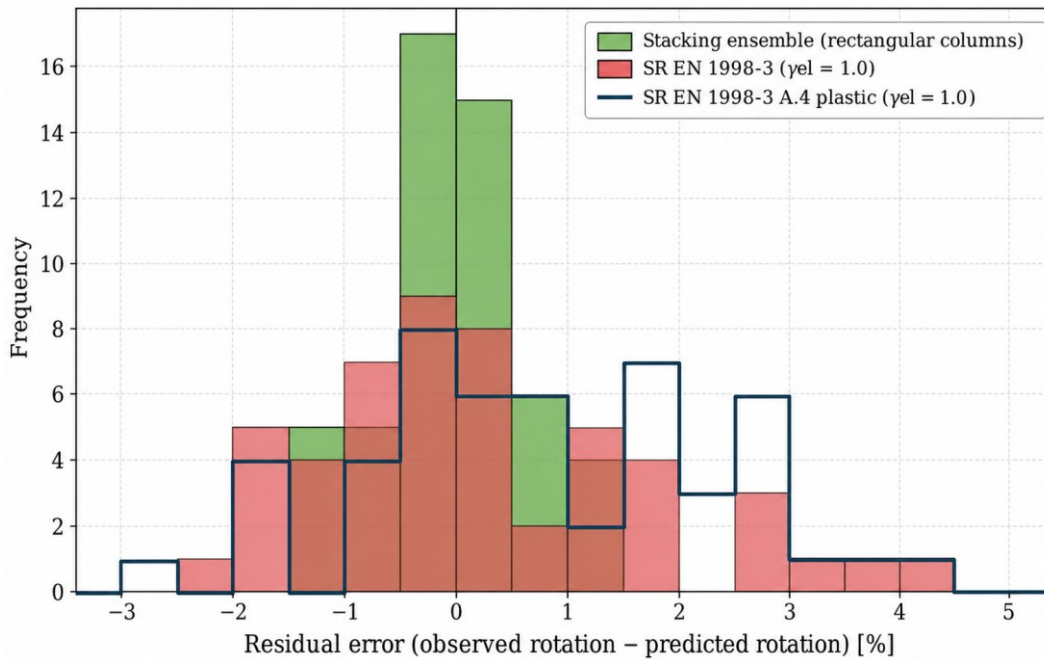


Fig. 7 - Residual error distribution, defined as observed minus predicted plastic rotation capacity, for the Stacking Ensemble model and the EN 1998-3 A.3 and A.4 expressions with $\gamma_{el} = 1.0$

These results should not be interpreted as invalidating the code expressions. Empirical formulations such as those in EN 1998-3:2005 remain useful because they are transparent, simple to apply, and suitable for design-oriented assessment. Their purpose is not necessarily to reproduce each experimental result with minimum error, especially when partial factors are included to ensure an adequate safety margin. However, the comparison shows that their dispersion can be significant when they are evaluated directly against experimental plastic rotation capacities, in agreement with previous observations reported in the literature [6], [7]. In this context, machine learning predictions can provide an additional quantitative reference for identifying the level of conservatism of the code provisions and for supporting the future calibration of simplified empirical expressions.

A limitation of the present comparison is the reduced number of shear-dominated specimens in the 52-specimen subset. Although the full rectangular database includes 44 shear failures, only 2 shear-dominated specimens could be retained in the Eurocode comparison subset because of the additional parameter requirements of the empirical expressions. Consequently, the results are more representative of flexure-controlled and moderately mixed behaviour than of brittle shear failure. Furthermore, the machine learning model remains dependent on the parameter ranges covered by the training database, so its use outside these ranges requires caution. Further validation on larger datasets with complete reinforcement and section details remains necessary before data-driven predictions can be considered as direct assessment tools.

7. Conclusions

This study presented a comparative assessment of the plastic rotation capacity of rectangular reinforced concrete columns, using both a data-driven prediction model and the empirical expressions provided in EN 1998-3. The analysis was focused on the comparison between the Stacking Ensemble model and the A.3 and A.4 formulations, considering the influence of the safety factor γ_{el} .

The results confirm that the estimation of plastic rotation capacity remains strongly affected by the variability of experimental data and by the interaction between axial load, reinforcement detailing, shear span, and failure mode. These effects are difficult to represent through simplified empirical expressions, especially when a single formulation is applied to specimens with different deformation capacities and degradation mechanisms. In this context, the Stacking Ensemble model provided a closer agreement with the experimental values, with more concentrated residual errors and a better ability to reproduce the observed dispersion of the data.

The comparison with EN 1998-3 showed that both A.3 and A.4 expressions lead to larger prediction errors than the machine learning model. For $\gamma_{el} = 1.0$, the empirical estimates remained closer to the experimental values, although the scatter was still significant. When $\gamma_{el} = 1.8$ was introduced, the predicted rotation capacities decreased considerably, leading to a more conservative response and to a systematic underestimation of specimens with larger observed deformation capacity. This behaviour was reflected not only in the statistical indicators but also in the distribution of residual errors and in the observed-versus-predicted comparisons.

The results should not be interpreted as a direct replacement of code-based formulations by machine learning models. The empirical expressions in EN 1998-3 remain practical and transparent tools for seismic assessment, especially because they are simple to apply and include safety-oriented assumptions. However, the comparison indicates that data-driven models can provide a useful complementary reference, particularly for evaluating the level of conservatism of empirical predictions and for identifying cases where simplified formulations may not capture the actual deformation capacity of rectangular RC columns.

Overall, the study highlights the potential of combining experimental databases, machine learning predictions, and code-based expressions in a common assessment framework. Such an approach can support a more informed interpretation of plastic rotation capacity, while preserving the practical role of empirical formulations in engineering applications. Further developments should focus on extending the experimental basis for shear-dominated and mixed failure mechanisms, since these cases remain critical for both empirical calibration and data-driven prediction.

Declaration on the use of AI-assisted tools:

During the preparation of the initial and revised versions of this manuscript, AI-assisted tools were used for language editing, text refinement, improvement of clarity, and grammatical checking. The tools used were ChatGPT, developed by OpenAI, and Grammarly. These tools were not used to generate scientific results, perform data processing, conduct numerical analyses, produce the reported metrics, or draw the scientific conclusions. All scientific content, data processing, numerical results, figures, interpretations, and conclusions were checked and validated by the authors, who take full responsibility for the final content of the manuscript.

References

- [1] Paulay, T., Priestley, M. J. N. (1992). *Seismic Design of Reinforced Concrete and Masonry Buildings*. New York: Wiley.
- [2] Yuan, F., Wu, Y. (2017). Effect of load cycling on plastic hinge length in RC columns. *Engineering Structures*, 147, 90–102. <https://doi.org/10.1016/j.engstruct.2017.05.046>
- [3] Pokhrel, M., Bandelt, M. J. (2019). Plastic hinge behavior and rotation capacity in reinforced ductile concrete flexural members. *Engineering Structures*, 200, 109699. <https://doi.org/10.1016/j.engstruct.2019.109699>.

- [4] European Committee for Standardization. (2005). *Eurocode 8: Design of Structures for Earthquake Resistance – Part 3: Assessment and Retrofitting of Buildings*. EN 1998-3, Brussels.
- [5] Panagiotakos, T. B., Fardis, M. N. (2001). Deformations of reinforced concrete members at yielding and ultimate. *ACI Structural Journal*, 98(2), 135–148.
- [6] Mpampatsikos, V., Nascimbene, R., Petrini, L. (2008). A Critical Review of the R.C. Frame Existing Building Assessment Procedure According to Eurocode 8 and Italian Seismic Code. *Journal of Earthquake Engineering*, 12, 52–82, <https://doi.org/10.1080/13632460801925020>.
- [7] Popa, V., Coțofană, D. (2006). Displacement capacity estimation for RC columns. Comparison between analytical and experimental results. *1st European Conference on Earthquake Engineering and Seismology*, Geneva, Switzerland.
- [8] Melo, J., Varum, H., Rossetto, T. (2015). Experimental cyclic behaviour of RC columns with plain bars and proposal for Eurocode 8 formula improvement. *Engineering Structures*, 88, 22–36. <https://doi.org/10.1016/j.engstruct.2015.01.033>.
- [9] Xie, Y., Ebad Sichani, M., Padgett, J. E., DesRoches, R. (2020). The promise of implementing machine learning in earthquake engineering: A state-of-the-art review. *Earthquake Spectra*, 36(4), 1769–1801, <https://doi.org/10.1177/8755293020919419>.
- [10] Luo, H., Paal, S. G. (2019). A locally weighted machine learning model for generalized prediction of drift capacity in seismic vulnerability assessments. *Computer-Aided Civil and Infrastructure Engineering*, 34(11), 935–950, <https://doi.org/10.1111/mice.12456>.
- [11] Luo, H., Paal, S. G. (2022). Artificial intelligence-enhanced seismic response prediction of reinforced concrete frames. *Advanced Engineering Informatics*, 52, 101568. <https://doi.org/10.1016/j.aei.2022.101568>.
- [12] Aladsani, M., Burton, H., Abdullah, S., Wallace, J. (2022). Explainable Machine Learning Model for Predicting Drift Capacity of Reinforced Concrete Walls. *ACI Structural Journal*, 119, 191–204. <https://doi.org/10.14359/51734484>
- [13] Yahyapour, A., Azhari, S., Hamidia, M. (2025). Intelligent multi-criteria seismic damage evaluation of RC columns using two-stage unsupervised and supervised learning. *Advanced Engineering Informatics*, 69, 104099. <https://doi.org/10.1016/j.aei.2025.104099>.
- [14] Kadhim, A.-O., Craifaleanu, I.-G. (2025). Predicting the plastic rotational capacity of reinforced concrete elements using machine learning. *Modelling in Civil and Environmental Engineering*, 20(1), 20–33, <https://doi.org/10.2478/mmce-2025-0002>.
- [15] Kadhim, A.-O., Craifaleanu, I.-G., Lozincă, E. (2025). Stacked ensemble and SHAP-based approach for predicting plastic rotational capacity in RC columns. *Frontiers in Built Environment*, 11, 1693218, <https://doi.org/10.3389/fbuil.2025.1693218>.
- [16] Berry, M., Parrish, M., Eberhard, M. (2004). *PEER Structural Performance Database User's Manual, Version 1.2*. Pacific Earthquake Engineering Research Center, University of California, Berkeley.
- [17] Moehle, J., Elwood, K. (2003). Collapse performance prediction for reinforced concrete frame structures. *Proceedings of the Pacific Conference on Earthquake Engineering*.
- [18] Smola, A. J., Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14(3), 199–222, <https://doi.org/10.1023/B:STCO.0000035301.49549.88>.
- [19] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32, <https://doi.org/10.1023/A:1010933404324>.
- [20] Chen, T., Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>.
- [21] Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, 5(2), 241–259. [https://doi.org/10.1016/S0893-6080\(05\)80023-1](https://doi.org/10.1016/S0893-6080(05)80023-1).
- [22] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.