Application of machine learning to reduce model uncertainty and enable resource-efficient design of flat slabs

T. Lux[1)], J. Sundheim[1)], T. Feiri[1)], U. Wiens[2)], and M. Ricker[1)]
[1)]Chair of Structural Concrete, Technical University of Dortmund,
44227 Dortmund, Germany, til.lux@tu-dortmund.de
[2)]German Committee for Structural Concrete, 10787 Berlin, Germany, info@dafstb.de

Abstract: The construction sector has a significant share of global CO_2 emissions due to its high material and resource consumption. Since material and resource consumption can be significantly influenced by the structural design of components, strategies to promote resource-efficient and optimised designs are required providing that the target reliability levels of structures prescribed in building design codes (e.g., DIN EN 1990+NA) are strictly satisfied. Structural components are designed in accordance with building design codes, such as DIN EN 1992-1-1+NA. Since the mechanical behaviour of some components can hardly be analytically described—e.g., punching shear of flat slabs without shear reinforcement—ultimate limit state evaluations partly rely on empirical equations. These equations exhibit significant model uncertainty, meaning that the ratio of actual to predicted load-bearing capacity is subject to statistical scatter. In punching shear verifications, model uncertainty has a great influence on the probability of failure. Previous studies have suggested that the model uncertainty of machine learning models can be lower than that of the empirical design equations specified in international codes. In this work, a machine-learning model (Extreme Gradient Boosting – *XGBoost*) was trained and utilised to predict the punching shear resistance of flat slabs without punching shear reinforcement. This model was employed in a reliability analysis with decoupled load and resistance side. Monte Carlo simulations were used to determine the probability that the actual load-bearing capacity, determined by the machine-learning model, falls below the design resistance according to DIN EN 1992-1-1+NA. The results indicate that predicting the actual load-bearing capacity with a trained model of lower uncertainty leads to a higher reliability level. This higher reliability can enable reductions in safety margins, for example, by lowering partial safety factors, thereby increasing the potential for resource-efficient designs.

Keywords: model uncertainty, punching shear, machine learning, structural reliability

1. Introduction

The construction industry accounts for approximately 37 % of global CO_2 emissions, largely due to its intensive consumption of materials and resources; of this total, about 6% is attributable to the embodied emissions of the most widely used construction materials, namely concrete, steel, and aluminum (United Nations Environment Programme, 2023). Since structural design decisions strongly affect this consumption, developing strategies that enable more resource-efficient structural components is essential (Häkkinen et al., 2015). At the same time, society expects structures to be highly reliable, as structural failure can have severe implications for human safety and significant economic consequences (e.g., Spaethe, 1992). Therefore, any

efforts to reduce material usage and enhance efficiency must ensure that the socially accepted reliability level—as recommended in modern structural codes—is maintained.
In contrast to emission reduction using alternative materials (e.g., low-carbon concretes or textile-reinforced concrete), which has been receiving substantial research attention, the potential for emission reduction through decreasing material volumes has been comparatively neglected (Fang et al., 2023). One approach to enhance material efficiency is to exploit existing safety reserves—hereby designated as *hidden safety margins*, in design standards—for example, by applying improved and less conservative load or resistance models (Teichgräber et al., 2022).
Machine **L**earning (ML) models can identify complex, nonlinear relationships between structural input parameters and load-bearing resistance, making them well suited to developing such improved resistance models. This potential is particularly relevant for structural components whose mechanical behaviour is difficult to describe analytically. Punching shear failure of flat slabs without shear reinforcement is a prime example: The failure mechanism is governed by a complex interaction of geometric, material, and loading parameters for which no closed-form analytical solution exists, and a large number of experimental datasets are publicly available (e.g., *fib* punching shear database (Ruiz & Ospina, 2025)) enabling robust model training and validation. The empirical design equations used to determine the resistance of such components inherently exhibit *model uncertainty*—defined here as the statistical scatter in the ratio between observed and predicted load-bearing capacity ($\theta_R = R_{\mathrm{obs}}/R_{\mathrm{pred}}$).
While hidden safety margins in punching shear have previously been estimated using analytical resistance models through reliability-based assessments (e.g., Ricker et al., 2021), these approaches are constrained by the model uncertainty inherent to the underlying design equations. Since data-driven ML models have been demonstrated to achieve lower model uncertainty than empirical code equations for punching shear resistance (Liang et al., 2022), their use as resistance models in reliability analyses offers a more accurate basis for quantifying such margins. However, to the authors' knowledge, no study has yet employed an ML-based resistance model for this purpose in the context of punching shear of flat slabs. This gap motivates the present study. In this study, an e**x**treme **g**radient **boo**sting model (XGBoost) is trained and used as a resistance model within a reliability analysis framework—employing **M**onte **C**arlo **S**imulation (MCS)—to quantify the actual load-bearing capacity and identify hidden safety margins for flat slabs without punching shear reinforcement. Exploiting these hidden safety margins enables more material-efficient designs, thereby reducing CO_2 emissions without compromising the reliability required by structural codes.

2. Reliability in Civil Engineering

2.1. Overview on reliability and safety margin

Reliability in structural engineering is defined as the ability of a structure or structural component to fulfil the specified requirements within its intended service life (DIN EN 1990:2021). To quantify reliability, DIN EN 1990:2021 introduces the reliability index β, defined as the negative standard normal quantile of the probability of failure P_{f}:

$$\beta = -\Phi^{(-1)}(P_{\mathrm{f}}) \tag{1}$$

The probability of failure is understood here as the notional probability that a requirement at the **U**ltimate **L**imit **S**tate (ULS) or the **S**erviceability **L**imit **S**tate (SLS) is not satisfied, arising from the statistical variability of the basic variables[1] entering the calculation. This notional probability of failure—and the corresponding reliability index β—serves as a measure of structural reliability and must be distinguished from the actual probability of failure, which is substantially influenced by human error (Vrouwenvelder et al., 2025). European design standards aim to ensure that structures satisfy requirements for safety and serviceability throughout their service life, i.e., that β meets or exceeds the specified target value. Target reliability indices are prescribed in DIN EN 1990:2021 as a function of consequence class and reference period, reflecting a codified balance between safety and economy.

In design practice, compliance with the target reliability is achieved through the partial safety factor concept, in which characteristic values of actions are amplified and those of resistances are reduced. Since partial safety factors are calibrated for a wide range of structural configurations and load cases, the resulting reliability often exceeds the target level for any given structure. The difference between the reliability index achieved and the prescribed target value constitutes the hidden safety margin—the quantity this study seeks to identify and quantify.

2.2. Reliability-based methods

Over the decades, a range of methods has been developed to approximate or determine structural failure probabilities. Reliability analysis methods are generally classified by their level of approximation. Level II methods, such as the **F**irst and **S**econd **O**rder **R**eliability **M**ethod (FORM/ SORM), operate in standardised normal space and approximate the failure surface at the design point (i.e., most probable failure point), offering computational efficiency at the cost of accuracy for strongly nonlinear limit state functions. Level III methods, such as MCS and variance reduction techniques as **I**mportance **S**ampling (MC-IS), and **Su**bset **S**imulation (SuS), estimate failure probabilities directly through repeated sampling and are considered exact in the limit of sufficiently large sample sizes, at the expense of greater computational demand (Melchers, 1999).

The fundamental task in a reliability analysis is to evaluate the probability that a structural component fails to meet its performance requirements, expressed through the limit state function:

$$g(\boldsymbol{X}) = g(X_1, X_2, \dots, X_n) \tag{2}$$

where $\boldsymbol{X} = (X_1, X_2, \dots, X_n)^T$ is the vector of basic random variables representing uncertain quantities such as material strength, geometry, and loading. Failure is defined as the event $g(\boldsymbol{X}) \leq 0$, and the corresponding probability of failure P_f is:

$$P_\mathrm{f} = P(g(\boldsymbol{X}) \leq 0) \tag{3}$$

A reliability analysis proceeds through the following key steps (Ditlevsen & Madsen, 1999). First, a target reliability level β_t is selected based on the consequence class and reference period[2]. Second, the limit state function is formulated to distinguish between safe and failed states. Third, the basic variables are identified and characterised by appropriate probability distributions and statistical parameters, accounting for uncertainties in resistance, loading, and model accuracy. Fourth, the reliability index β is computed using the chosen analysis method. Finally, a sensitivity analysis is performed to assess the relative influence of individual random variables on the computed reliability level.

---

[1] In reliability analysis, the term refers to random variables characterised by specified probability distributions and associated statistical parameters, such as mean and variance.

[2] Target reliability level are given in DIN EN 1990:2021, e.g., β_t = 3.8 for reliability class 2 and a 50-year reference period.

3. Flat slabs without punching shear reinforcement according to DIN EN 1992-1-1:2011+NA(D)

In the load transfer region between a flat slab and its supporting columns, the simultaneous occurrence of high bending moments and shear forces can give rise to two distinct failure modes: (i) flexural failure and (ii) local shear failure. The latter involves a three-dimensional, conical failure surface forming around the column—a mechanism referred to as punching shear. Two cases are distinguished depending on whether shear reinforcement is present; this study considers exclusively flat slabs without punching shear reinforcement.

The mechanical behaviour of flat slabs without shear reinforcement is complex to describe analytically, and ultimate limit state assessments therefore rely in part on empirical formulations. To improve the understanding of the underlying failure mechanisms, numerous experimental investigations have been conducted since the mid-20th century (e.g., Guandalini & Muttoni, 2004; Hallgren et al., 1998; Rankin & Long, 1987). The design equations for punching shear resistance have evolved alongside this growing body of experimental evidence. A new generation of European design standards is currently under development, incorporating recent advances in reinforced concrete research as reflected in the *fib* Model Code 2020 (Fédération internationale du béton (*fib*), 2023). The investigations presented in this study, however, refer to the currently applicable version of Eurocode 2 for engineering practice in Germany—DIN EN 1992-1-1:2011, DIN EN 1992-1-1/NA:2013, and DIN EN 1992-1-1/NA/A1:2015 (hereinafter also referred to as Eurocode 2 (EC2))—which is rooted in the CEB-FIP model code 1990 (Euro-International Committee for Concrete & Fédération internationale de la précontrainte, 1998).

According to this standard, punching shear resistance must be verified along the critical control perimeter u_1, located at a distance of $2d$ from the periphery of the loaded area, by satisfying:

$$v_{\mathrm{Ed}} = \frac{\beta_e \cdot V_{\mathrm{Ed}}}{u_1 \cdot d} \leq v_{\mathrm{Rd,c}} \tag{4}$$

where V_{Ed} is the applied shear force, d is the effective depth of the slab (distance from the extreme compression fibre to the centroid of the longitudinal tensile reinforcement), and β_e is a factor[3] accounting for an uneven shear stress distribution along u_1 due to eccentric loading.

In turn, the design shear stress v_{Ed} is compared against the punching shear resistance without shear reinforcement $v_{\mathrm{Rd,c}}$, determined as:

$$v_{\mathrm{Rd,c}} = C_{\mathrm{Rd,c}} \cdot k \cdot (100 \cdot \rho_{\mathrm{l}} \cdot f_{\mathrm{ck}})^{1/3} + k_1 \cdot \sigma_{\mathrm{cp}} \geq v_{\mathrm{min}} + k_1 \cdot \sigma_{\mathrm{cp}} \tag{5}$$

The parameters in Eq. (5) are defined as follows: $C_{\mathrm{Rd,c}}$ is an empirical coefficient equal to $0.18/\gamma_{\mathrm{C}}$ for general cases and $0.18/\gamma_{\mathrm{C}} \cdot (0.1 \cdot u_0/d + 0.6)$ for interior columns with $u_0/d < 4$, where $\gamma_{\mathrm{C}} = 1.5$ is the partial safety factor for concrete, and u_0 is the column perimeter. The term $k = 1 + (200/d)^{0.5} \leq 2.0$ (d in mm) considers the size effect of the effective depth, ρ_{l} is the mean flexural reinforcement ratio considering a slab width equal to the column width plus $3d$ on each side given by $\rho_{\mathrm{l}} = \left(\rho_{\mathrm{lx}} \cdot \rho_{\mathrm{ly}}\right)^{0.5}$, and f_{ck} is the characteristic cylinder concrete compressive strength. The additional term $k_1 \cdot \sigma_{\mathrm{cp}}$ can be used to account for membrane forces in the slab, for example when dealing with a prestressed beam or slab. The term v_{min} is the minimum punching shear capacity given by Eq. (6):

$$v_{\mathrm{min}} = \max\left(min\left(0.0525 - 0.015 \cdot \frac{d - 600}{200}; 0.0525\right); 0.0375\right) / \gamma_{\mathrm{C}} \cdot k^{3/2} \cdot f_{\mathrm{ck}}^{1/2} \tag{6}$$

[3] Note that β in the context of DIN EN 1992-1-1:2011 is unrelated to the reliability index β commonly used in structural reliability analysis.

4. Training and Validation of a data-driven machine learning model

4.1. Machine Learning Algorithm

Since the objective of the machine learning model is to predict a continuous target variable—the load-bearing capacity of flat slabs without punching shear reinforcement—from labelled experimental data, supervised regression is the appropriate learning paradigm (Bishop, 2006).
The model type was chosen based on the findings of (Liang et al., 2022), in which the XGBoost model yielded more accurate predictions than other machine-learning algorithms. Gradient boosting methods are generally well-suited to tabular datasets of moderate size—as is typical of experimental databases in structural engineering—due to their robustness to outliers, ability to handle heterogeneous input features, and strong predictive performance (Friedman, 2001). In XGBoost, an ensemble of decision trees is constructed sequentially, with each tree fitted to the residuals of its predecessor, thereby progressively reducing prediction error. For a full algorithmic description, the reader is referred to (Chen & Guestrin, 2016). All model training and predictions were performed using the *xgboost* package (Chen et al., 2026) in the open-source statistical computing environment R (R Core Team, 2025).

4.2. Database

Supervised machine learning requires a comprehensive database of input parameters and associated target variables for model training and validation (Bishop, 2006). In this study, the target variable is the punching shear resistance of flat slabs without shear reinforcement, expressed as the shear stress capacity along the critical control perimeter. The database must contain all relevant input parameters, including parameters influencing the load-bearing capacity that are not explicitly considered in the code formulation.
The experimental data used for training and validation of the XGBoost model are drawn from the *fib* punching shear database (Ruiz & Ospina, 2025), an internationally recognised repository of punching shear tests on flat slabs without shear reinforcement. The database comprises 468 tests across 73 experimental series and provides the input parameters and measured resistances required for model development.

4.3. Preparation of Data

4.3.1. *Filtering*

From the original *fib* database of 468 documented tests, a subset was selected based on the following criteria. First, only tests in which a distinct punching failure was recorded (database entry "P") were retained. Second, only specimens with circular, square, or rectangular column geometries were included. Third, a minimum reinforcement yield strength of $f_y \geq 300$ MPa was required to avoid bias in the training process arising from non-standard reinforcement configurations. Of the 468 tests, 397 satisfied all three criteria and were used for model training and validation.

4.3.2. *Feature Selection*

A defined set of predictor variables (*features*) was compiled for model training based on the parameters known to influence punching shear behaviour. These include the concrete cylinder compressive strength f_c, the effective depth d, the geometric longitudinal reinforcement ratio ρ_l, the yield strength f_y of longitudinal the reinforcement, the distance from the slab edge to the load introduction area a, the critical control perimeter

u_1, the shear slenderness $a_\lambda = a/d$, and the specific column perimeter ratio u_0/d. Column shape was represented using one-hot encoding, in which each geometry category (circular, square, rectangular) is expressed as a binary indicator variable, enabling categorical features to be incorporated directly into the splitting criteria of decision tree-based models.
The feature set was determined through an iterative selection process guided by feature importance scores and error metrics of the trained model. While additional input variables may also be considered, the inclusion of an excessive number of features should be avoided. The selection should therefore be restricted to relevant variables in order to reduce the risk of overfitting and to prevent unnecessary increases in model complexity during the training process.

4.3.3. *Feature Transformation*
The feature distributions in the training dataset exhibit noticeable skewness[4] (Figure 1).

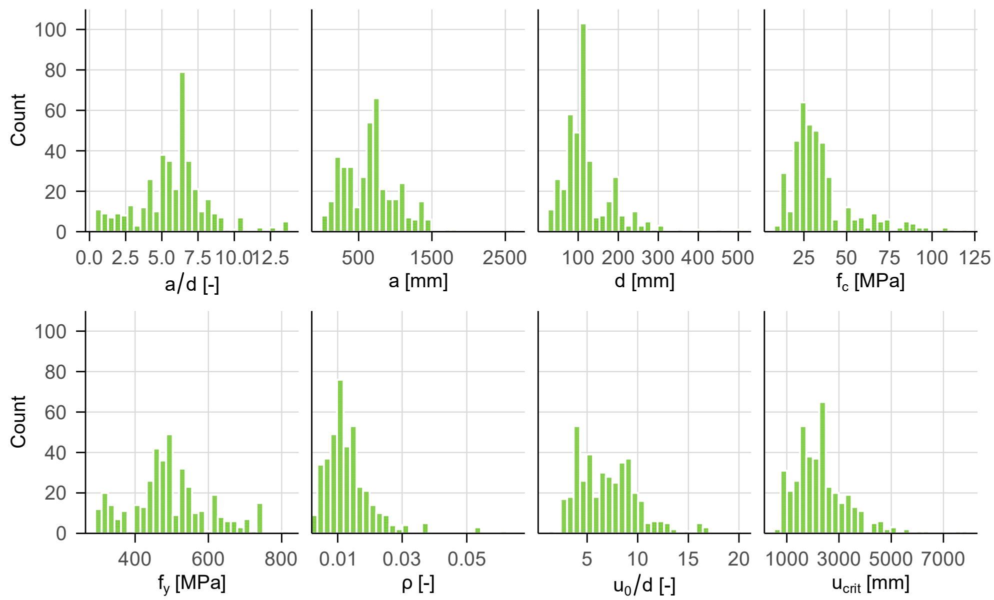

Figure 1. Histograms of the features in the filtered database

Skewed distributions can negatively affect model performance, which is why logarithmic transformations are commonly applied to reduce skewness prior to training (Kuhn & Johnson, 2013). Accordingly, all non-categorical features x_j with a skewness coefficient $\gamma_{1,j} > 1$ were log-transformed:

$$x_{\mathrm{trans,j}} = \begin{cases} \log(x_\mathrm{j}) & \text{for } x_\mathrm{j} \in \{d, f_\mathrm{c}, \rho_l, u_{\mathrm{crit}}\} \\ x_\mathrm{j} & \text{for } x_\mathrm{j} \notin \{d, f_\mathrm{c}, \rho_l, u_{\mathrm{crit}}\} \end{cases} \tag{7}$$

Subsequently, all features were standardised using *z*-score normalisation:

[4] Measure of the asymmetry of a probability distribution around its mean.

$$x_{\text{scale,j}} = \frac{x_{\text{trans,j}} - \bar{x}_{\text{trans,j}}}{s_{x_{\text{trans,j}}}} \tag{8}$$

Although z-score normalisation is not required for tree-based models such as XGBoost—which are invariant to monotonic feature transformations—it was applied here since an ANN was trained using the same features in the underlying study. Note that applying the trained model to new observations requires access to the mean $\bar{x}_{trans,j}$ and standard deviation $s_{x_{trans,j}}$ of the transformed training features.

4.3.4. *Target Variable Definition*

The choice of target variable significantly influences model performance and should be chosen based on the structure of the training data. To remove geometric scale dependence, the target is defined not as the ultimate failure load V_{u}; instead, it is defined as the punching shear stress along the critical control perimeter, normalised by the concrete compressive strength f_{c} to reduce the dominance of this single parameter on the output distribution. A log-transformation is applied to reduce skewness in the target distribution, consistent with the pre-processing applied to the input features. The target variable y is therefore in accordance with (Mangalathu et al., 2021) defined as:

$$y = \log\left(\frac{V_u}{u_{\text{crit}} \cdot d \cdot f_{\text{c}}}\right) \tag{9}$$

The target variable used here performs well for the available dataset; however, alternative target definitions are also possible.

4.4. Model Setup and Training

The filtered database of 397 tests was randomly partitioned into a training set (70%, 277 tests) and a held-out test set (30%, 120 tests). The hyperparameters governing model architecture were fixed prior to training based on (Mangalathu et al., 2021): learning rate `eta=0.045`, maximum tree depth `max_depth=5`, training instance subsample ratio `subsample=0.55`, and column subsample ratio per tree `colsample_bytree=0.9`. For a detailed description of each parameter and its effect on model behaviour, the reader is referred to the XGBoost documentation (Chen et al., 2026). The optimal number of boosting iterations was determined using 5-fold cross-validation (i.e., `nfold=5`) on the training set via the `xgb.cv()`, function, minimising the root mean square error (RMSE). A maximum of 300 boosting rounds (i.e., `nrounds=300`) was permitted, with early stopping applied if no improvement in RMSE was observed over 30 consecutive rounds (i.e., `early_stopping_rounds=30`). The lowest cross-validation RMSE was achieved at 296 boosting iterations (i.e., `nrounds=296`), which was adopted as the final model configuration.

4.5. Validation and Model Uncertainty

The model performance was evaluated on a held-out test dataset (n = 120) using four error metrics. To facilitate interpretation, the target variable was transformed back to the load-bearing capacity V_u:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(V_{u,i} - \widehat{V_{u,i}}\right)^2} = 56.3 \tag{10}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|V_{u,i} - \widehat{V_{u,i}}\right| = 32.0 \tag{11}$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{V_{u,i} - \widehat{V_{u,i}}}{V_{u,i}}\right| = 7.71 \tag{12}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(V_{u,i} - \widehat{V_{u,i}}\right)^2}{\sum_{i=1}^{n}\left(V_{u,i} - \bar{V_u}\right)^2} = 0.979 \tag{13}$$

where n is the number of tests in the test dataset, $V_{u,i}$ is the observed ultimate failure load for test i, $\widehat{V_{u,i}}$ is the corresponding model prediction for test i, and $\bar{V_u}$ is the mean of the observed ultimate failure loads. The model uncertainty θ_R (index R for Resistance) is defined as the ratio of the observed to the predicted load-bearing capacity $\theta_R = R_{\text{obs}}/R_{\text{pred}}$. This quantity exhibits statistical dispersion across the test dataset and is treated as a random variable characterised by an underlying probability distribution. In structural reliability, model uncertainty is commonly represented by a Lognormal distribution, which is strictly non-negative and consistent with the multiplicative nature of resistance uncertainties (e.g., *Probabilistic Model Code*, 2001). The adequacy of this assumption was assessed visually via the histogram displayed in Figure 2.

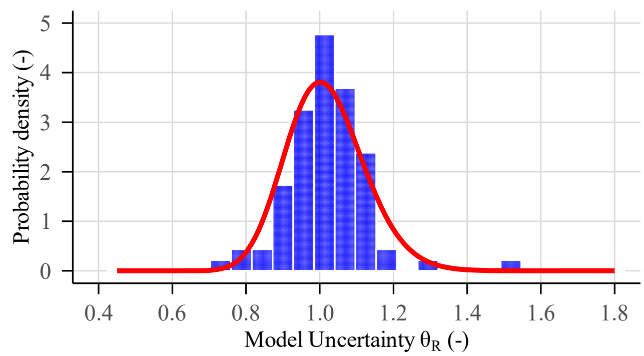

Figure 2. Histogram of the model uncertainty θ_R and fitted log-normal distribution

The lognormal distribution parameters[5] are derived from the sample mean $\bar{\theta}_{R,XGB}$ = 1.017 and coefficient of variation CoV($\theta_{R,XGB}$) = 0.106.
A mean model uncertainty close to unity indicates that the XGBoost model predicts the punching shear resistance without systematic bias. The coefficient of variation of 10.6% is low relative to typical values reported for empirical design equations (e.g., Liang et al., 2022; Ricker et al., 2021), suggesting that the data-driven model captures the underlying resistance mechanism with high predictive accuracy. This is further substantiated by direct comparison with the model uncertainty of the design equation according to DIN EN 1992-1-1:2011 assessed in Section 4.6.

[5] $\sigma_{\ln} = \sqrt{\ln(1 + COV(\theta_R)^2)}$, $\mu_{\ln} = \ln(\overline{\theta_R}) - 0{,}5 \cdot \sigma_{\ln}^2$

4.6. COMPARISON WITH THE DESIGN MODEL ACCORDING TO DIN EN 1992-1-1

For comparison with the model uncertainty of the model specified in DIN EN 1992-1-1:2011, the same test dataset as used for the XGBoost model is considered. However, additional filters are applied to avoid biasing the model uncertainty of EC2 by unusual or impractical configurations. Tests with an effective depth d < 100 mm and characteristic cylindric concrete compressive strength f_{ck} > 104 MPa are therefore excluded in accordance with (Hegger et al., 2021). Of the 120 tests in the original test dataset, 82 tests remain that are used to compare the model uncertainty of the XGBoost model with that of DIN EN 1992-1-1:2011. Note that, according to Section 9.3.1.1 of DIN EN 1992-1-1/NA:2013, a minimum slab thickness of 70 mm is sufficient for slabs without punching shear reinforcement. However, in practice, the construction of flat slabs with a thickness smaller than 150 mm is rare in practice; therefore, a minimum value of d = 100 mm was assumed for the effective depth in this study (Hegger et al., 2021). Based on this reduced test dataset, both the error metrics of the model uncertainty for EC2 and XGBoost according to Eqs. (8)–(11) as well as the point estimates of the distribution parameters were calculated and compared (see Table I).

Table I. Comparison of error metrics and Point estimators of XGBoost and EC2

Model	No. Tests	RMSE [kN]	MAE [kN]	MAPE [%]	R^2	$\bar{\theta}$	CoV_θ
XGBoost	82	66.17	39.26	7.25	0.972	1.017	0.104
EC2	82	117.7	74.45	13.98	0.939	1.146	0.166

The design equation according to EC2 exhibits substantially higher model uncertainty than the XGBoost model, both in terms of mean value and coefficient of variation. The distributions of both model uncertainties are compared in Figure 3.

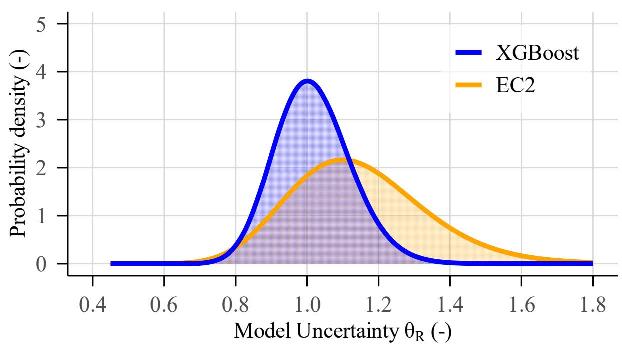

Figure 3. Fitted Lognormal distributed model uncertainty θ_R of the trained XGBoost model and the model according to Eurocode 2

4.7. INTERVAL ESTIMATORS OF MODEL UNCERTAINTIES

The model uncertainty parameters $\mu_{\ln}$ and $\sigma_{\ln}$ were estimated from a test dataset of n = 82 observations. Given this limited sample size, point estimates alone are insufficient to characterise the distributional parameters with confidence; interval estimators are therefore applied to quantify the uncertainty in the estimated parameters. Confidence intervals at the 95 % confidence level were calculated following (Deutscher Ausschuss für Stahlbeton e.V. (DAfStb), 2025). The resulting confidence intervals are reported in Table II.

For the subsequent reliability analysis, the conservative bounds of the confidence intervals are adopted as inputs to the MCS, ensuring that parameter uncertainty is propagated into the computed reliability index.

Table II. Confidence intervals of the distribution parameters of θ_R

Model	Stochastic parameters	Point Estimate		Interval Estimators (95 %)		
		LN-Value	LN-Value transformed	LN-Value	LN-Value transformed	Corrected parameters
EC2	$\hat{\mu}_{\theta_{R,EC2}}$	0.122	1.146	[0.092, +∞]	[1.111, +∞]	1.111
	$\hat{\sigma}_{\theta_{R,EC2}}$	0.165	0.191	[0, 0.190]	[0, 0.219]	0.206[6]
	$\mathrm{CoV}(\theta_{R,EC2})$	-	0.166	-	[0, 0.192]	0.185[6]
XGB	$\hat{\mu}_{\theta_{R,XGB}}$	0.012	1.017	[0, +∞]	[0.998, +∞]	0.998
	$\hat{\sigma}_{\theta_{R,XGB}}$	0.104	0.106	[0, 0.120]	[0, 0.122]	0.109[6]
	$\mathrm{CoV}(\theta_{R,XGB})$	-	0.107	-	[0, 0.120]	0.109[6]

5. Application of the trained model in reliability analyses

5.1. Numerical Analysis

The reliability of flat slabs without punching shear reinforcement designed according to DIN EN 1992-1-1:2011 has been extensively investigated in the literature (e.g., Ricker et al., 2021). To enable a direct comparison between the reliability indices obtained using the XGBoost resistance model and those based on the analytical design equation according to DIN EN 1992-1-1:2011, the reference example from (Ricker et al., 2021) is adopted here. This example is used to quantify the influence of the data-driven resistance model on the computed reliability index β and to identify the resulting hidden safety margin (Figure 4).

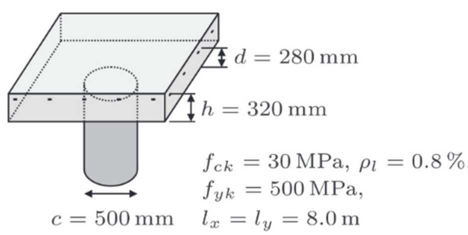

Figure 4. Considered example of a flat slab without punching shear reinforcement, taken from (Ricker et al., 2021)

To investigate the influence of individual parameters on the reliability index, each parameter is varied individually from the reference state (Figure 4) while all other parameters are held constant. The reliability

[6] Correction of σ_{θ_R} and $\mathrm{CoV}(\theta_R)$ by taking into account the test uncertainties CoV_e acc. to (Sykora et al., 2015):

$CoV(\theta_R) = \sqrt{CoV_{observed}^2 - CoV_e^2}$, with $CoV_e = 0.05$

analyses follow the approach of (Ricker et al., 2021), in which the load and resistance sides of the limit state function are treated separately. The ultimate limit state function is formulated as:

$$g(\boldsymbol{X}) = V_{\mathrm{R}}(\boldsymbol{X}) - V_{\mathrm{Rd,c}} = 0 \tag{14}$$

where $V_{\mathrm{R}}(\boldsymbol{X})$ is the actual resistance as a function of the random variables vector $\boldsymbol{X} = \{X_1, X_2, \ldots, X_i\}$, and V_{Rd} is the design value of resistance. Failure is defined as the event $g(\boldsymbol{X}) \leq 0$, i.e., the realised resistance falling below the design resistance value. Since only the resistance side is evaluated, the computed reliability index β_{R} corresponds to the resistance component of the total reliability index, related through the sensitivity factor α_R as: $\beta_{\mathrm{R}} = \alpha_R \cdot \beta$. Adopting the value $\alpha_R = 0.8$ prescribed in DIN EN 1990:2021—consistent with the assumption that resistance and load variability contribute in fixed proportion to total uncertainty— the total reliability index is recovered by scaling $\beta = \beta_{\mathrm{R}}/0.8$. Without this correction, the resistance-only analysis would systematically underestimate β relative to the full reliability problem.

5.2. Limit state Function and Basic Variables

The failure domain defined by the punching shear design provision of DIN EN 1992-1-1 is expressed as:

$$g(\boldsymbol{X}) = \theta_{\mathrm{R}} \cdot v_{\mathrm{R,c}}(d, c, \rho_{\mathrm{l}}, f_{\mathrm{c}}) \cdot u_1(d, c) \cdot d - V_{\mathrm{Rd,c}} \leq 0 \tag{15}$$

where $\theta_{\mathrm{R}}$is the random variable representing model uncertainty of the design equation; $v_{\mathrm{R,c}}(\cdot)$ is the punching shear resistance as a function of the following basic variables: effective depth d, column dimension c, flexural reinforcement ratio ρ_{l}, and concrete compressive strength f_{c}. The term $u_1(d, c)$ is the critical control perimeter at a distance of $2d$ from the column face. $V_{\mathrm{Rd,c}}$ is the deterministic design resistance according to DIN EN 1992-1-1:2011, as defined in Section 3. Substituting the resistance formulation according to this code, for interior circular columns into Eq. (15) yields:

$$g_{\mathrm{EC2}}(\boldsymbol{x}) = \theta_{\mathrm{R}} \left[C_{Rd,c} \left[\min\left(2.0;\ \sqrt{1 + \frac{200}{d}} \right) \right] \cdot (100 \cdot \rho_{\mathrm{l}} \cdot f_{\mathrm{c}})^{1/3} \right] \cdot \left(2 \cdot \pi \cdot \left(\frac{c}{2} + 2 \cdot d \right) \right) \cdot d - V_{\mathrm{Rd,c}} = 0 \tag{16}$$

with $C_{Rc} = 0.18 \cdot \min(1;\ 0.1 \cdot \frac{\pi \cdot c}{d} + 0.6)$.

The limit state function is subsequently reformulated by replacing the EC2 resistance model with predictions from the trained XGBoost model:

$$g_{\mathrm{XGB}}(\boldsymbol{x}) = \theta_{\mathrm{R,XGB}} \cdot \exp\left(y_{\mathrm{XGB}}\left(f_{\mathrm{c}}, d, \rho_{\mathrm{l}}, f_{\mathrm{y}}, u_{\mathrm{crit}}, a, \frac{a}{d}, \frac{u_0}{d} \right) \right) \cdot f_{\mathrm{c}} \cdot u_{\mathrm{crit}} \cdot d - V_{\mathrm{Rd,c}} = 0 \tag{17}$$

where a/d is included as an independent feature consistent with the XGBoost feature set defined in Section 4.3, derived from the random variables a and d. The model uncertainty $\theta_{\mathrm{R,XGB}}$ was characterised in Section 4.5; interval estimates of its Lognormal distribution parameters are adopted here to avoid underestimating uncertainty arising from the limited validation subset. The stochastic models for the basic variables c, f_{c}, d, f_{y}, and a were adopted from (*Probabilistic Model Code*, 2001). The stochastic model for the longitudinal reinforcement cross section follows (Muttoni, 2023). The derived quantities $u_{\mathrm{crit}} = 2 \cdot \pi \left(\frac{c}{2} + 2 \cdot d \right)$, $u_0 = \pi \cdot c$ and $\rho_{\mathrm{l}} = A_{\mathrm{s}}/d$ are computed from the basic random variables at each simulation step. The stochastic models are summarised in Table III.

Table III. Stochastic models of basic variables						
Basic variable X		Distribu-tion Type	Unit	Nominal value	Expected Value E(X)	Standard Deviation SD(X)
Cyl. concrete compressive strength	f_c	LST[7]	N/mm²	f_{ck}	-	-
Effective depth	d	N	mm	d_{nom}	$d_{nom} + 10$	10
Column diameter	c	N	mm	c_{nom}	$c_{nom} + 0.003 \cdot c_{nom}$	$4 + 0.006 \cdot c_{nom}$
Longitudinal reinforcement area	A_s	N	mm²	$A_{s,nom}$	$0.97 \cdot A_{s,nom}$	$0.02 \cdot A_{s,nom}$
Shear span	a	N	mm	a_{nom}	$a_{nom} + 0.003 \cdot a_{nom}$	$4 + 0.006 \cdot a_{nom}$
Yield strength of reinforcement	f_y	N	N/mm²	f_{yk}	$500 + 2 \cdot$ SD(X)	30
Model uncertainty EC2	θ_{EC2}	LN	-	-	1.111	0.206
Model uncertainty XGB	θ_{XGB}	LN	-	-	0.998	0.109
LST: Log-student-t distribution, N: Lormal distribution, LN: Lognormal distribution						

5.3. Reliability Methods

The reliability analyses were performed using subset simulation, which combines conditional sampling with **M**arkov **C**hain **M**onte **C**arlo (MCMC) techniques to estimate small failure probabilities efficiently—an approach that is substantially less computationally demanding than direct MCS for the probability levels relevant to structural reliability (10 000 samples per simulation level, level of probability $p_0 = 0.10$, and confidence level $\alpha = 0.05$) (Au & Beck, 2001). The analyses were implemented using the open-source package *TesiproV* (Nille-Hauf et al., 2026) in R (R Core Team, 2025).

5.4. Results and Discussion

Figure 5 shows the computed reliability index β as a function of the varied input parameters for both limit state functions investigated. Across all parameter variations, the XGBoost model consistently yields higher reliability indices than the DIN EN 1992-1-1:2011 formulation, reflecting its lower model uncertainty. The magnitude of the increase in β is parameter-dependent: for example, the effect is more pronounced at lower concrete compressive strengths than at higher values (Figure 5b). This sensitivity reflects the degree to which each input variable contributes to prediction uncertainty in the XGBoost model and warrants further investigation through a dedicated sensitivity analysis, for example using SHAP values.

---

[7] The stochastic model of the concrete compressive strength was proposed by (Rackwitz, 1983) as a Log-student-t (LST) distribution: $F_x(x) = F_{t_{\nu'}}\left[\frac{\ln(x-m')}{s'} \cdot \sqrt{\frac{n'}{n'+1}}\right]$, with m', n', s' and ν' being prior parameters given for the different concrete classes in (*Probabilistic Model Code*, 2001). For ready-mixed concrete, the parameters are given as follows:

Concrete grade	Prior Parameters			
	m'	n'	s'	ν'
C15	3.40	3.0	0.14	10
C25	3.65	3.0	0.12	10
C35	3.85	3.0	0.09	10
C45	3.98	3.0	0.07	10

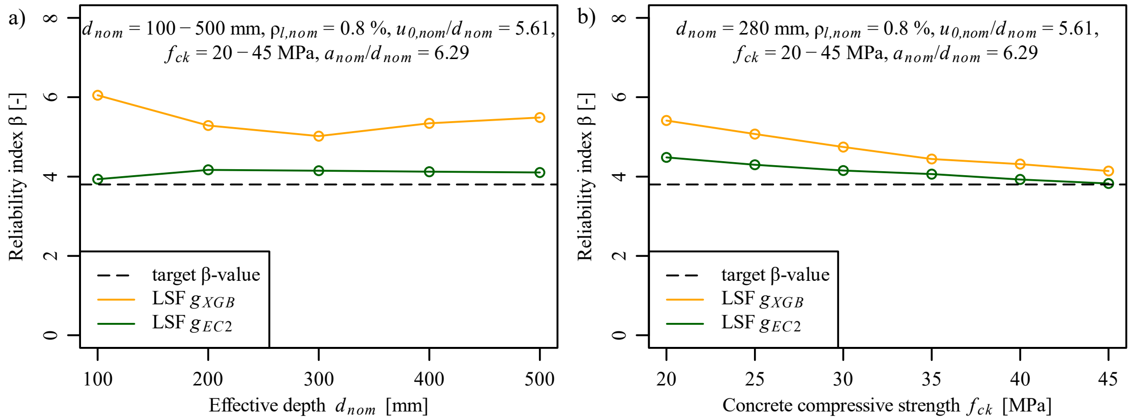

Figure 5. Reliability index β in variation with (a) effective depth d [mm], and (b) concrete compressive strength f_{ck} [MPa]

The results further reveal physical influences captured by the XGBoost model that are not represented in the code verification format. Figure 6a shows that the model predicts an increasing reliability index for small shear slenderness values, while the code formulation does not include this parameter and therefore yields a constant reliability index. The beneficial trend predicted by the model is physically plausible and consistent with the upcoming new generation of Eurocode 2, where the shear slenderness will be explicitly included in the design equation of DIN EN 1992-1-1:2025.

Figure 6b demonstrates that for specific column perimeters $u_0/d < 4$, the XGBoost model predicts a substantially higher resistance than the code design model. Under the current National Annex, the empirical factor $C_{\mathrm{Rd,c}}$ is reduced for $u_0/d < 4$, a correction not reflected in the trained XGBoost model. This discrepancy may result either from the limited number of corresponding tests in the training dataset or from an overly conservative reduction in the code formulation. The available data do not allow a clear distinction between these explanations, which represents a limitation of the present study and suggests the need for targeted experimental investigations.

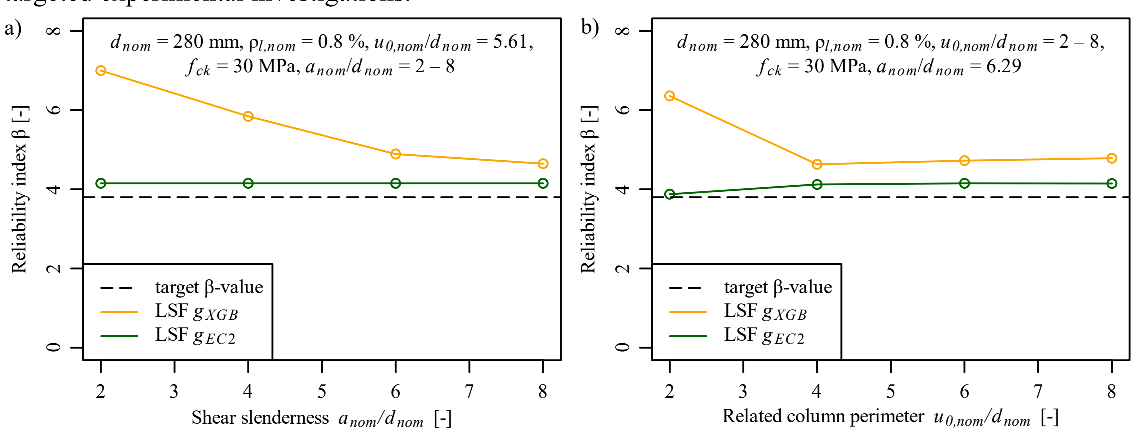

Figure 6. Reliability index β in variation with (a) shear slenderness a/d [-], and (b) related column perimeter u_0/d [-]

Figure 7 presents the reliability index β as a function of the mean longitudinal reinforcement ratio ρ_l. For reinforcement ratios up to approximately 1.5%, the XGBoost model consistently yields higher reliability indices than the code-based limit state function. At ρ_l = 2%, however, the XGBoost model predicts a lower resistance than the code design equation, causing the two reliability curves to converge or causing β_XGB to approach β_EC2. This observation is physically significant: DIN EN 1992-1-1:2011 applies an upper cap of $\rho_\mathrm{l} = 0.02$ in the resistance formula, beyond which no further increase in punching shear resistance is assumed. The XGBoost model suggests that this plateau is reached at a lower reinforcement ratio, in the range $0.015 < \rho_\mathrm{l} < 0.02$, implying that the code formulation may overestimate the contribution of high reinforcement ratios to resistance. This finding supports the need for the code cap, while also suggesting that the effective limit may be more conservative than currently prescribed.

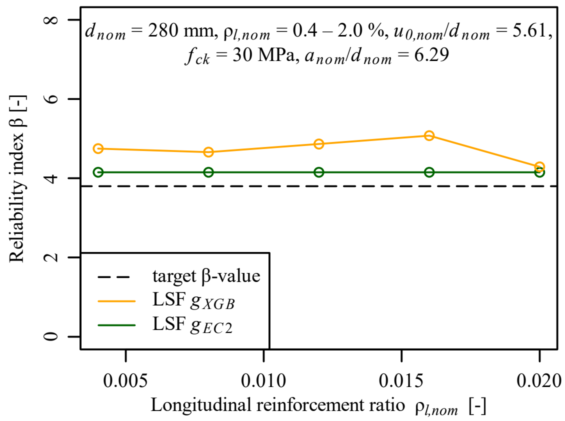

Figure 7. Reliability index β in variation with longitudinal reinforcement ratio ρ_l [-]

Further experimental data for specimens with ρ_l > 1.5% is required to investigate this observation more rigorously, as the current training dataset may underrepresent this range (Figure 7). A targeted sensitivity analysis using SHAP values could additionally clarify the role of ρ_l relative to other input features in this reinforcement range.

6. Conclusions

Based on this investigation, the following conclusions can be drawn:

- The trained XGBoost model predicts the load-bearing capacity of flat slabs without punching shear reinforcement with lower bias and lower scatter than the design equation of DIN EN 1992-1-1:2011, as evidenced by a mean model uncertainty closer to unity ($\mu_{\theta_{R,EC2}} = 1.146$ *vs.* $\mu_{\bar{\theta}_{R,XGB}}$ = 1.017) and a substantially reduced coefficient of variation (COV($\theta_{R,EC2}$) = 0.166 *vs.* COV($\bar{\theta}_{R,XGB}$) = 0.106).

- The lower scatter of the XGBoost model uncertainty reduces the probability of very low load-bearing capacity realisations compared to the code model. This directly translates into higher reliability indices in the reliability analyses: flat slabs without punching shear reinforcement designed according to DIN

EN 1992-1-1:2011 exhibit hidden safety margins that arise solely from the relatively high model uncertainty of the code resistance equation. When the more accurate XGBoost model is used as the resistance model, these margins become apparent.

- The reliability analyses further reveal physical relationships between input parameters and load-bearing capacity that are not explicitly captured in the code design mode—either as deliberate simplifications or due to limitations in the empirical calibration basis. Notable examples include the beneficial effect of small shear slenderness and the apparent overestimation of resistance at high longitudinal reinforcement ratios in the code formulation.

Several limitations of this study should be noted. The XGBoost model is trained on 397 tests from the *fib* punching shear database, representing idealised laboratory conditions and excluding long-term effects, cyclic loading, and construction imperfections. Some parameter ranges—particularly high longitudinal reinforcement ratios and small relative column perimeters (u_0/d< 4)—are underrepresented, so conclusions for these regions remain uncertain. In addition, the reliability analyses are performed solely within the framework of DIN EN 1992-1-1:2011 and its German National Annex, meaning the identified safety margins are specific to this design format and not directly transferable to other codes.

Future work could address these limitations. Targeted experiments in underrepresented parameter ranges would improve model reliability and help clarify observed behaviours such as those for u_0/d< 4. A SHAP-based sensitivity analysis could provide a systematic decomposition of the XGBoost predictions and enable a more rigorous comparison with the functional relationships assumed in DIN EN 1992-1-1:2011. Finally, the most consequential extension of this work would be a reliability-based design optimisation that translates the identified safety margins into reduced cross-sections—directly fulfilling the material efficiency objective stated in the introduction and quantifying the associated reduction in CO_2 emissions.

Acknowledgements

The authors wish to express their gratitude to the German Committee for Structural Concrete (*Deutscher Ausschuss für Stahlbeton e.V., DAfStb*) for funding this investigation (DAfStb-Projekt V521). The authors also gratefully acknowledge the computing time provided on the Linux HPC cluster at Technical University Dortmund (LiDO3), partially funded in the course of the Large-Scale Equipment Initiative by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) as project 271512359.

References

Au, S.-K., & Beck, J. L. (2001). Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics*, *16*(4), 263–277. https://doi.org/10.1016/S0266-8920(01)00019-4

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

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

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y., Li, Y., Yuan, J., & Cortes, D. (2026). *xgboost: Extreme Gradient Boosting* (p. 3.2.1.1) [Dataset]. https://doi.org/10.32614/CRAN.package.xgboost

Deutscher Ausschuss für Stahlbeton e.V. (DAfStb). (2025). *Verfahren zur Herleitung von Sicherheitsbeiwerten im Massivbau unter Verwendung probabilistischer Methoden*.

Ditlevsen, O., & Madsen, H. O. (1999). *Structural reliability methods* (Repr). Wiley.
Euro-International Committee for Concrete & Fédération internationale de la précontrainte (Eds.). (1998). *CEB-FIP model code 1990: Design code* (Repr). Telford.
Fang, D., Brown, N., De Wolf, C., & Mueller, C. (2023). Reducing embodied carbon in structural systems: A review of early-stage design strategies. *Journal of Building Engineering*, *76*, 107054. https://doi.org/10.1016/j.jobe.2023.107054
Fédération internationale du béton (fib). (2023). *Model Code for Concrete Structures 2020*.
Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, *29*(5), 1189–1232.
Guandalini, S., & Muttoni, A. (2004). *Punching test on symmetrical reinforced concrete slabs without shear reinforcement (Essais de poinçonnement symétrique des dalles en béton armé sans armature à l'effort tranchant)* (Versuchsbericht 00.03-R1). Institute de Structures, Laboratoire de Construction en Béton (IS-Beton), École Polytechnique Fédérale de Lausanne,.
Häkkinen, T., Kuittinen, M., Ruuska, A., & Jung, N. (2015). Reducing embodied carbon during the design process of buildings. *Journal of Building Engineering*, *4*, 1–13. https://doi.org/10.1016/j.jobe.2015.06.005
Hallgren, M., Kinnunen, S., & Nylander, B. (1998). Punching Shear Tests on Column Footings. *Nordic Concrete Research*, *21*(3), 1–22.
Hegger, J., Ricker, M., Adam, V., Feiri, T., & Nille-Hauf, K. (2021). *Überprüfung der Zuverlässigkeit der für die nächste Generation von EN 1992-1-1 vorgesehenen neuen Bemessungsansätze gegen Durchstanzen ohne Durchstanzbewehrung und Querkraft ohne Querkraftbewehrung* (Abschlussbericht F-2017-034).
Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. Springer New York. https://doi.org/10.1007/978-1-4614-6849-3
Liang, S., Shen, Y., & Ren, X. (2022). Comparative study of influential factors for punching shear resistance/failure of RC slab-column joints using machine-learning models. *Structures*, *45*, 1333–1349. https://doi.org/10.1016/j.istruc.2022.09.110
Mangalathu, S., Shin, H., Choi, E., & Jeon, J.-S. (2021). Explainable machine learning models for punching shear strength estimation of flat slabs without transverse reinforcement. *Journal of Building Engineering*, *39*, 102300. https://doi.org/10.1016/j.jobe.2021.102300
Melchers, R. E. (Robert E.) (with Internet Archive). (1999). *Structural reliability analysis and prediction*. Chichester ; New York : John Wiley. http://archive.org/details/structuralreliab0000melc_w1w5
Muttoni, A. (2023). *Background document to clause 4.3.3 and Annex A Partial safety factors for materials* (EPFL-IBETON 16-06-R11). EPFL, Ecole Polytechnique Fédérale de Lausanne.
Nille-Hauf, K., Lux, T., Feiri, T., & Ricker, M. (2026). *TesiproV — Probabilistic Reliability Analysis Framework*.
Probabilistic Model Code. (2001). Joint Committee on Structural Safety.
R Core Team. (2025). _R: A Language and Environment for Statistical Computing_. *R Foundation for Statistical Computing*. Vienna, Austria. <https://www.R-project.org/>
Rackwitz, R. (1983). Predictive distribution of strength under control. *Material & Constructions*, *16*(94), 259–267.
Rankin, G. I. B., & Long, A. E. (1987). Predicting the punching strength of conventional slab-column specimens. *Proceedings of ICE, Civil Engineering*, *82*, 327–346.
Ricker, M., Feiri, T., Nille-Hauf, K., Adam, V., & Hegger, J. (2021). Enhanced reliability assessment of punching shear resistance models for flat slabs without shear reinforcement. *Engineering Structures*, *226*, 111319. https://doi.org/10.1016/j.engstruct.2020.111319
Ruiz, M. F., & Ospina, C. E. (2025, June 4). *Punching Shear Database* [Fédération internationale du béton]. Fédération internationale du béton. Fédération Internationale Du Béton. https://www.fib-international.org/commissions/databases.html
Spaethe, G. (1992). *Die Sicherheit tragender Baukonstruktionen* (2., neubearb. Aufl.). Springer.
Sykora, M., Holicky, M., Prieto, M., & Tanner, P. (2015). Uncertainties in resistance models for sound and corrosion-damaged RC structures according to EN 1992-1-1. *Materials and Structures*, *48*(10), 3415–3430. https://doi.org/10.1617/s11527-014-0409-1
Teichgräber, M., Köhler, J., & Straub, D. (2022). Hidden safety in structural design codes. *Engineering Structures*, *257*, 114017. https://doi.org/10.1016/j.engstruct.2022.114017
United Nations Environment Programme. (2023). *Building Materials and the Climate: Constructing a New Future*. United Nations.
Vrouwenvelder, T., Beck, A., Proske, D., Faber, M., Köhler, J., Schubert, M., Straub, D., & Teichgräber, M. (2025). Interpretation of probability in structural safety – A philosophical conundrum. *Structural Safety*, *113*, 102473. https://doi.org/10.1016/j.strusafe.2024.102473