



## Ensemble Learning Models for Prediction of Punching Shear Strength in RC Slab-Column Connections

Omid Habibi <sup>1\*</sup>, Tarik Youssef <sup>2</sup>, Hamed Naseri <sup>3</sup>, Khalid Ibrahim <sup>4</sup>

<sup>1</sup> Department of Building, Civil and Environmental Engineering, Concordia University, Montreal H3G 2W1, Canada.

<sup>2</sup> Faculty of Engineering, L'Université Française d'Égypte, Al-Shorouk City 11837, Egypt.

<sup>3</sup> Department of Civil, Geological, and Mining Engineering, Polytechnique Montreal, Montreal H3T 1J4, Canada.

<sup>4</sup> Department of Structural Engineering, Faculty of Engineering, Ain Shams University, Cairo 11535, Egypt.

Received 29 December 2023; Revised 24 April 2024; Accepted 07 May 2024; Published 26 May 2024

### Abstract

In reinforced concrete (RC) structures, accurate prediction of the punching shear strength (PSS) of slab-column connections is imperative for ensuring safety. The existing equations in the literature show variability in defining parameters influencing PSS. They neglect potential variable interactions and rely on a limited dataset. This study aims to develop an accurate and reliable model to predict the PSS of slab-column connections. An extensive dataset, including 616 experimental results, was collected from earlier studies. Six robust ensemble machine learning techniques—random forest, gradient boosting, extreme gradient boosting, adaptive boosting, gradient boosting with categorical feature support, and light gradient boosting machines—are employed to predict the PSS. The findings indicate that gradient boosting stands out as the most accurate method compared to other prediction models and existing equations in the literature, achieving a coefficient of determination of 0.986. Moreover, this study utilizes techniques to explain machine learning predictions. A feature importance analysis is conducted, wherein it is observed that the reinforcement ratio and compressive strength of concrete demonstrate the highest influence on the PSS output. SHapley Additive exPlanation is conducted to represent the influence of variables on PSS. A graphical user interface for PSS prediction was developed for users' convenience.

**Keywords:** RC Slab-Column Connection; Punching Shear Strength; Machine Learning; Feature Importance Analysis; SHAP.

### 1. Introduction

The connection between reinforced concrete (RC) slabs and columns is a crucial feature of structural systems that requires a meticulous and precise design process. Punching shear and flexural failure are common failure modes in RC slab-column connections. When a concentrated load is exerted onto an RC slab, this leads to the creation of a shear cone around the column, ultimately causing punching failure (see Figure 1) [1]. Punching failure happens in a way that makes the slab crack along critical planes that go through the slab from its compression to tension surfaces in an oblique direction [2, 3]. This failure is immense, sudden, and detrimental; no apparent signs were observed before the collapse. Hence, the reliable prediction of the punching shear capacity (PSC) of RC slab-column connections is important in the design of RC slab-column connections.

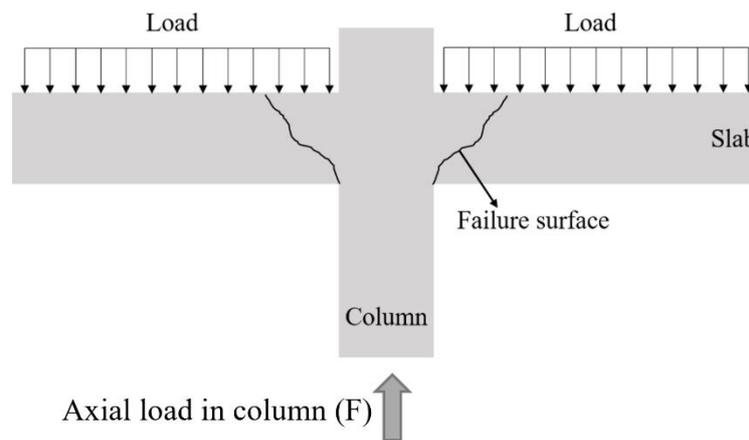
\* Corresponding author: [omid.habibi@concordia.ca](mailto:omid.habibi@concordia.ca)

 <http://dx.doi.org/10.28991/CEJ-SP2024-010-01>



© 2024 by the authors. Licensee C.E.J, Tehran, Iran. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Several codes and literature resources offer empirical equations/models to estimate the PSC of RC slab-column connections [4–6]. These equations are typically developed based on experimental records using traditional analyses (e.g., linear regression or trend curve models). Traditional regression analyses typically involved statistical fitting based on individually assessing the connection between PSC and each design variable. Consequently, potential interactions between different variables were not taken into account [7]. A significant debate surrounds the identification of the optimal predictive model for PSC in RC slab-column connections [8, 9]. Notably, design codes showed considerable disparities when compared to experimentally measured PSC [9]. It is important to note that the data used to develop any formula is limited to a fraction of the experiments conducted prior to the date of its development. To obtain a reliable prediction for PSC of RC slab-column connections, it is essential to have a comprehensive and updated dataset wherein modern/accurate prediction methods apply.



**Figure 1. Punching shear failure in slab-column connections**

Conventional experimental tests are the established method for evaluating various characteristics of concrete structures [10, 11]. Recently, machine learning (ML) techniques have found extensive applications across diverse domains within the field of structural engineering, aiming to enhance prediction performance (i.e., accuracy) [12–14]. Machine learning, a subset of artificial intelligence (AI), can analyze extensive datasets and identify patterns, aiding in predicting how structures perform in various situations. ML models can address the complex relationships among the numerous input variables, offering more precise predictions compared to traditional regression analyses [7]. The incorporation of ML methods offers a notably more cost-effective approach. This is particularly true when compared to conducting large-scale experimental studies, especially in scenarios where expensive equipment and time-intensive processes are involved. The above-mentioned advantages of ML models highlight the importance of applying machine learning to predict the PSC of slab-column connections.

Accordingly, a few studies have been conducted to assess the PSC of slab-column connections using ML techniques. As such, an artificial neural network (ANN) was applied to estimate the PSC of slab-column connections using 244 data observations collected from the literature [15]. The study assessed the impact of slab effective depth, reinforcement ratio, and compressive strength of concrete on the punching shear strength (PSS) of slab-column connections. The results suggested that ANN could outperform the available design code equations (e.g., the ACI 318-08 [16] equation) when comparing prediction accuracy. Chetchotisak et al. [17] collected 342 experimental results to estimate the punching shear capacity of the slab-column connections using multiple linear regression (MLR) and ANN models. Results showed that the MLR approach exhibits higher accuracy and simplicity in the prediction of PSS. Similarly, Tran et al. [3] employed MLR and ANN to predict the PSS of the two-way concrete slab-column connections using 218 data observations. Input variables included slab thickness, column section width, effective slab depth, reinforcement ratio, concrete compressive strength, and reinforcement yield strength. Herein, ANN demonstrated higher prediction accuracy compared to MLR and design code equations. Faridmehr et al. [18] introduced a hybrid model combining the Bat algorithm with ANN (Bat-ANN) to estimate the PSS of slab-column connections. Comparing the prediction accuracy of the introduced model with available design codes suggested the superior performance of the Bat-ANN model in terms of prediction accuracy. A further sensitivity analysis was performed using the hybrid Bat-ANN model, highlighting the influential role of the reinforcement ratio in PSS. Naseri Nasab et al. [19] estimated the PSS of RC slabs connected to circular and square columns through ANN models with different configurations, utilizing 164 experimental observations gathered from the literature. The findings indicated that the optimized arrangement of ANN models surpasses existing models in the literature, demonstrating a superior coefficient of determination.

Fuzzy-based models have also been used to estimate the PSS of slab-column connections. As such, Choi et al. [20] developed a fuzzy-based model to forecast the PSS in interior slab-column connections subjected to concentric loads and to address interactions among various punching shear modeling variables. This model substantially enhanced the prediction accuracy of PSS in slab-column connections compared to design codes. Akbarpour and Akbarpour [21] utilized ANN and the adaptive neuro-fuzzy inference system (ANFIS) to estimate the PSS of two-way RC slabs based on 189 experimental records gathered from the literature. Comparing the models' predictions revealed that the accuracy of the ANN was superior to that of the ANFIS.

The capability of other ML models has been considered in previous studies. As such, Mangalathu et al. [22] examined the effectiveness of ML models in predicting PSS, utilizing a dataset of 380 experimental points. Their results demonstrated that extreme gradient boosting (XGBoost) outperforms existing literature equations in terms of prediction accuracy. Cao [23] introduced a novel strategy by integrating XGBoost with diverse hyperparameter optimization techniques, such as forensic-based investigation (FBI), Bayesian optimization (BO), grid search (GS), random search (RS), and default setting (DS), for predicting punching shear strength (PSS) in RC flat slab-column connections. The study utilized three datasets, each with 207, 241, and 380 data points. The results highlighted the FBI-XGBoost model as the most reliable method for accurately estimating punching shear strength in RC flat slab-column systems. Wu and Zhou [24] utilized a hybrid PSO-SVR model by combining support vector regression (SVR) and particle swarm optimization (PSO) to predict PSS in two-way reinforced concrete slabs. The hybrid model exhibited superior performance, achieving a higher correlation coefficient and smaller error indices than the original SVR model.

As evidenced by the studies reviewed above, it is concluded that a limited number of data observations have been generally used to predict the PSS of slab-column connections. The prediction performance of machine learning techniques significantly depends on the number of data samples used in the modeling [25]. Therefore, an effective strategy for enhancing the accuracy of current prediction models in estimating PSS is to utilize a comprehensive dataset. Further, most of the previous studies applied ANNs, fuzzy-based techniques, and linear regression models. However, other robust prediction models, particularly ensemble learning techniques, have not received enough attention to predict punching shear strength. In other research areas, ensemble learning techniques have demonstrated higher prediction accuracy compared to conventional ML models [26-28], particularly those models utilized for predicting the punching shear strength of slab-column connections (e.g., ANN). For instance, the light gradient boosting machine (LightGBM) exhibited superior performance regarding both prediction accuracy and computational efficiency when compared to a range of prediction techniques, including K-nearest neighbors (KNN) [29], back-propagation ANN [30], and support vector machine [31]. As another ensemble learning method, gradient boosting with categorical features support (CatBoost) demonstrated higher prediction accuracy than conventional machine learning techniques, such as ANN, gaussian naïve Bayes, decision tree (DT), multi-layered perceptron (MLP), and logistic regression (LR) [28, 32]. However, the application of these powerful ensemble learning methods to the punching shear strength of slab-column connections has been limited. Moreover, limited research has investigated the interpretability of intricate ML models, specifically in identifying the most influential parameters on the PSS of slab-column connections and understanding the impact of variations in these parameters on PSS variation.

To address the above-mentioned limitations, this study aims to predict punching shear strength with the highest possible accuracy, employing ensemble machine learning models. A comprehensive dataset comprising 616 experimental results, considered among the most extensive in the author's knowledge, has been meticulously collected from the literature. This dataset is intended to contribute to the development of reliable and accurate prediction models. Different ensemble learning models are applied for the prediction process. From the applied techniques, LightGBM, gradient boosting (GB), and CatBoost, which have demonstrated great prediction accuracy in other areas of structural engineering [26-30] are used for the first time to predict the punching shear strength of RC slab-column connections. The performance of these techniques is compared with existing literature equations. This performance comparison encompasses nearly all pertinent equations found in the literature, demonstrating the efficacy of powerful ensemble learning techniques in predicting the punching shear strength of slab-column connections. The other objective of this study is to consider the influence of different variables on the PSS of slab-column connections, a fundamental aspect of optimizing structural design. For this purpose, feature importance selection is applied using the permutation model to find the most influential parameters in PSS. Additionally, a SHapley Additive exPlanation (SHAP) dependence analysis is conducted to assess how changes in each variable contribute to variations in PSS. Finally, a graphical user interface (GUI) is developed to enhance the practicality and accessibility of this research. The methodology used in this study is summarized in Figure 2. In the following sections, the data preparation process is first presented. Subsequently, the applied machine learning techniques are summarized. Then, the results of the machine learning technique are discussed. Consequently, the outcomes of a feature importance analysis are presented. Afterward, the results of SHAP are presented. Finally, the developed GUI is displayed.

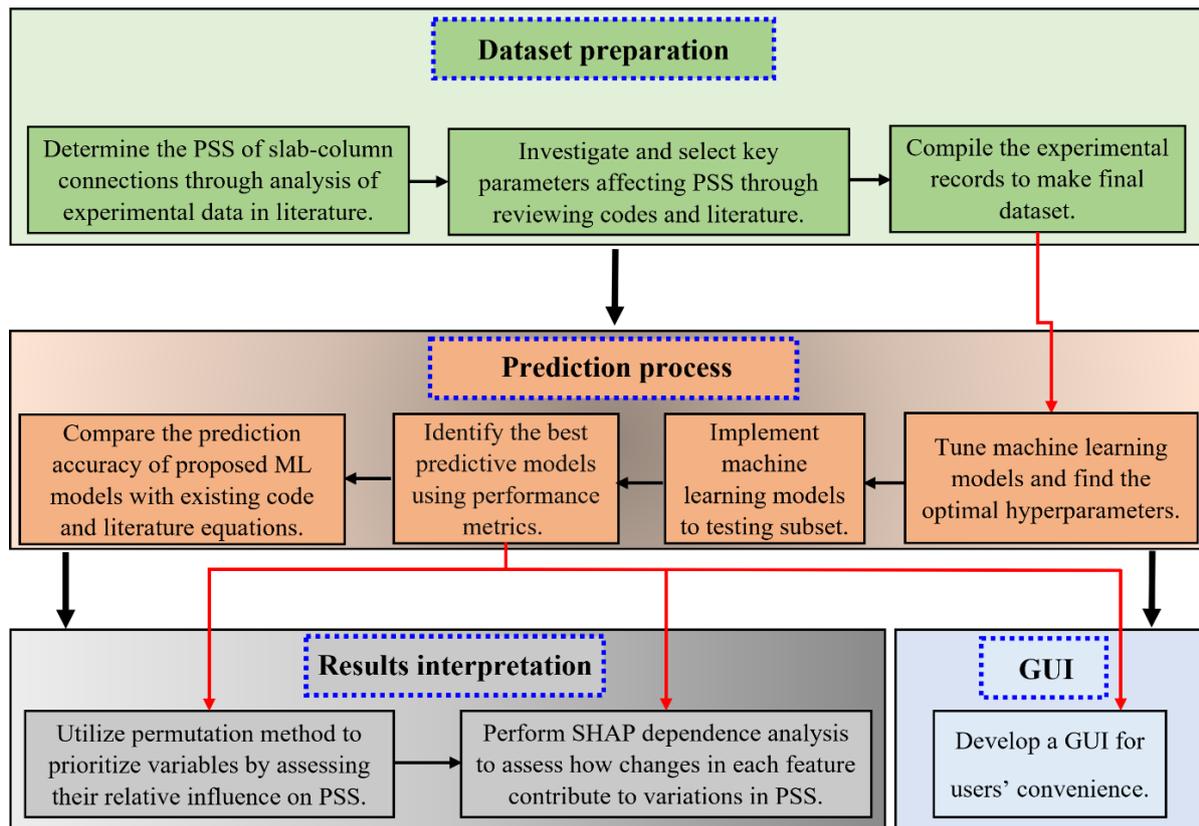


Figure 2. Flowchart of the research methodology

## 2. Data and Variables

To develop reliable ML models, comprehensive databases are required. The applied database comprising the aforementioned 616 experiments is collected from the literature [33-72] (uploaded to <https://github.com/Omid1373/PSC>). Concrete compressive strength ( $f'_c$ ), reinforcement ratio ( $\rho$ ), steel yield strength ( $f_y$ ), effective depth ( $d$ ), and critical perimeter ( $b_0$ ) (according to ACI 318-14 [4] and ACI 318-19 [73]) are considered parameters affecting the PSS. These parameters are in accordance with design code equations and the literature survey [4, 5, 15, 39, 74-78]. The input variables of  $f'_c$ ,  $\rho$ ,  $1/d$ ,  $f_y$ , and  $d/b_0$  are chosen due to their direct relevance to the PSS ( $v_c$ , output variable), aligning with the mechanical aspects of punching shear as outlined in existing literature [15, 76, 78] (see Table 1). Statistical inferences about these variables are presented in Table 2. The distribution plots of variables are demonstrated in Figure 3.

Table 1. Description of variables of the study

Variables	Unit	Variable type	Description
$f'_c$	MPa	Input	Concrete compressive strength
$\rho$	%	Input	Reinforcement ratio
$1/d$	(1/mm)	Input	Reverse of effective depth
$f_y$	MPa	Input	Yielding stress of steel
$d/b_0$	---	Input	Effective depth to critical perimeter ratio
$v_c$	MPa	Output	Nominal shear strength

Table 2. Statistical information of input variables

Criteria	$f'_c$ (MPa)	$\rho$ (%)	$1/d$ (1/mm)	$f_y$ (MPa)	$d/b_0$	$v_c$ (MPa)
Mean	41.017	1.278	0.011	520.118	0.091	2.684
Median	33.500	1.060	0.009	494.000	0.086	2.539
Minimum	9.500	0.280	0.002	250.000	0.027	0.380
Maximum	130.100	11.700	0.032	900.000	0.230	8.590
STD	21.960	1.100	0.006	122.152	0.029	1.029
COV	0.535	0.861	0.528	0.235	0.319	0.383

Note: STD= Standard Deviation COV= Coefficient of Variance.

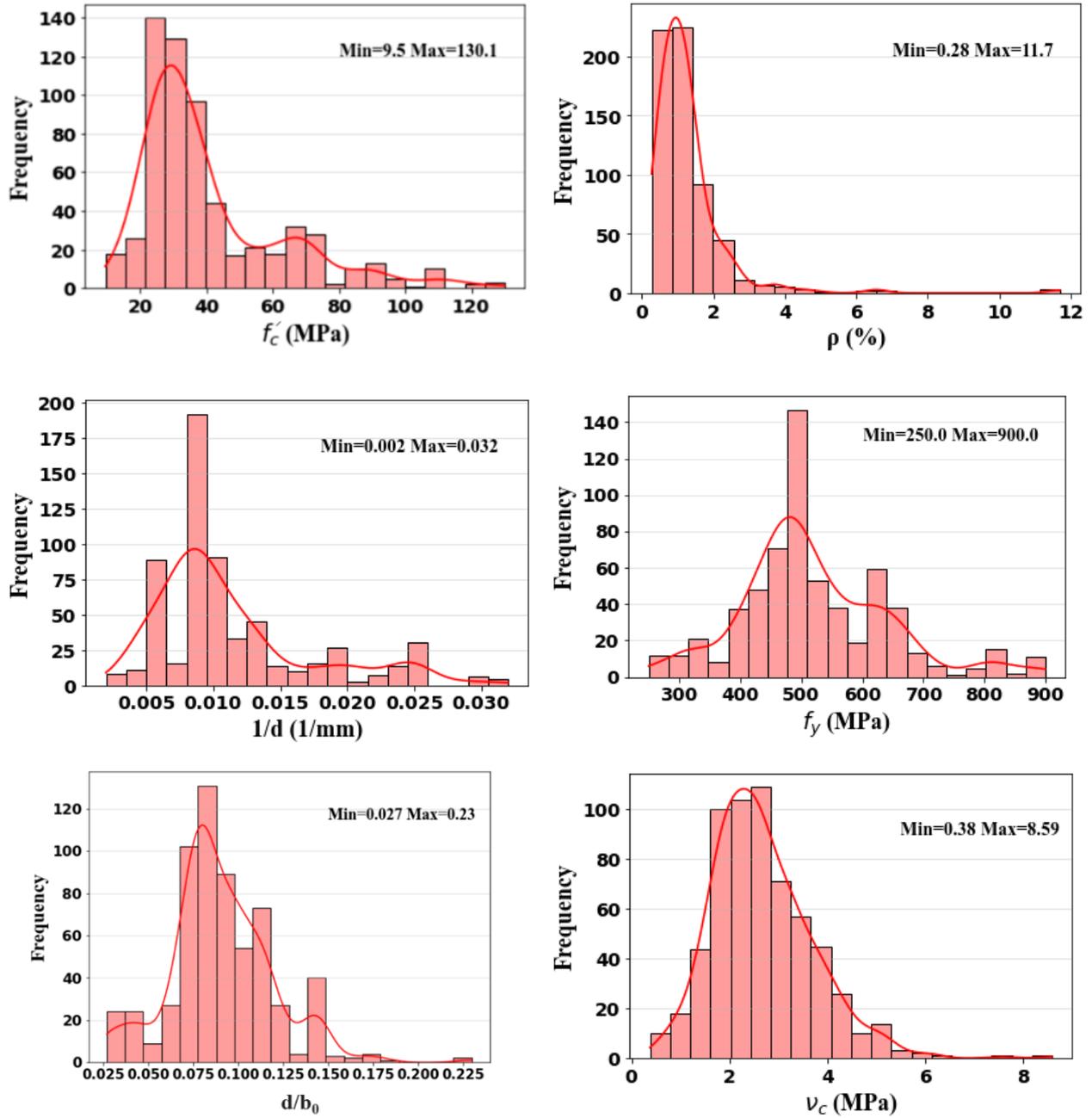
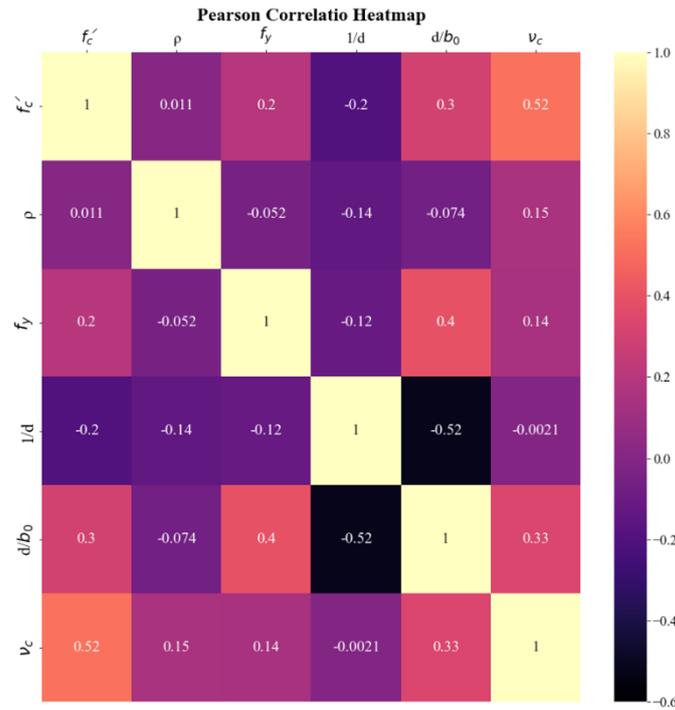


Figure 3. Input variable range

To evaluate the correlation of variables, their mutual correlation is calculated using the Pearson equation [79]:

$$\text{Pearson Correlation} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

where  $x_i$  and  $y_i$  are the  $i^{th}$  values of variables  $x$  and  $y$ .  $\bar{x}$  and  $\bar{y}$  are the average values of variables  $x$  and  $y$ , and  $n$  is the number of samples in the dataset (here, 616). Using this equation, the correlation of variables is calculated, and the heatmap of the variables' correlation is illustrated in Figure 4. The heatmap of variable correlations displays the patterns and relations between variables in the dataset. The values in the correlation heatmap range between -1 and 1, wherein “-1” signifies perfect negative correlation; the “zero” integer signifies that there is no correlation; and “1” implies perfect positive correlation. The highest correlation is between  $\nu_c$  and  $f'_c$  as well as  $1/d$  and  $d/b_0$ , with a correlation magnitude of 0.52.



**Figure 4. Correlation matrix**

### 3. Machine Learning Models

The six ML algorithms are used to predict the RC slab-column connections PSS, including ensemble models of random forest (RF), gradient boosting (GB), adaptive boosting (AdaBoost), gradient boosting with categorical features support (CatBoost), extreme gradient boosting (XGBoost), and light gradient boosting machine (LightGBM). An overview of the ML models is presented below:

#### 3.1. Random Forest (RF)

RF is a robust machine learning method used for a variety of tasks, including classification and regression. In the case of RF, the base model is a decision tree (DT), a structure resembling a tree, wherein each internal node signifies a variable, and each leaf node represents a prediction. The purpose of the DT is to partition the data observations into smaller sets according to the input variable values until each set contains only one target value [80]. The RF algorithm creates multiple DTs by randomly selecting sets of the training data and variables at each iteration, a technique known as bagging, which helps to avoid overfitting and improve the model's prediction ability. During the prediction phase, the RF algorithm takes the average of the predictions from all the individual DTs to obtain the final prediction ( $\hat{y}$ ), as follows (see Equation 2):

$$\hat{y} = \frac{1}{N} \sum_{i=1}^N DT_i(x) \quad (2)$$

where,  $N$  is the total number of the decision trees and  $DT_i(x)$  is the prediction of  $i^{th}$  decision tree for input data. This process, known as ensemble averaging, functions to reduce the variance and improve model accuracy [81].

#### 3.2. Gradient Boosting (GB)

GB is one of the variants of ensemble methods where multiple weak models are created (i.e., decision tree) and then combined for enhanced performance. Each weak learner is trained, and the error residual of them is measured using a loss function. In GB, each new tree tries to reduce the errors generated by the previous ones and reduces the system's loss function through the gradient descent iterative process. This works by fitting a tree to the residuals (i.e., the differences between experimental and predicted values) of the previous tree. This iterative process continues until the desired level of accuracy is achieved [82].

#### 3.3. Adaptive Boosting (AdaBoost)

AdaBoost is a robust technique that is attractive for a variety of machine learning tasks [83]. Like other ensemble methods, AdaBoost constructs a strong predictor by combining multiple weak learners. In AdaBoost, each weak learner is assigned equal weights at the beginning and is also iteratively trained on the data. After each iteration, the weights of

the misclassified observations are increased, enabling them to receive more attention in the subsequent iteration. This iterative process focuses on improving the accuracy of the overall ensemble model by giving more importance to previously misclassified data points. This learning approach might appear similar to the gradient descent utilized in GB. However, in contrast to adjusting a solitary predictor's parameter to minimize the cost function, AdaBoost progressively incorporates predictors into the ensemble, leading to gradual enhancement. This process is repeated until a predefined number of weak learners are trained or until the desired level of accuracy is achieved. In the final model, each weak learner's output is weighted according to its performance, with more accurate learners receiving higher weights. AdaBoost then combines these weak learners to form a robust predictor [84].

### 3.4. Gradient Boosting with Categorical Features Support (CatBoost)

CatBoost is a powerful and flexible machine learning algorithm that can be employed for an extensive range of applications, particularly in areas where categorical data is common. CatBoost mainly uses Bayesian estimators to avoid overfitting [85]. CatBoost employs a specialized technique to handle categorical variables efficiently. Instead of directly converting them to binary variables, CatBoost uses an ordered boosting scheme combined with target-based statistics. This approach enables CatBoost to directly model categorical variables without the need for one-hot encoding. By doing so, CatBoost significantly reduces computational complexity and saves on training times while maintaining accurate predictions. In addition, CatBoost uses a balanced and symmetric tree structure to generate the ensemble, ensuring that the splits are consistent across all nodes at the same depth [86].

### 3.5. Extreme Gradient Boosting (XGBoost)

XGBoost is another version of the gradient boosting algorithm developed by Chen & Guestrin [87]. Similar to GB, XGBoost also employs a set of weak learners (e.g., decision trees) to create an ensemble model that is more accurate and powerful than any individual model. The overall prediction of the XGBoost model is obtained by summing the predictions of all the weak models:

$$\hat{y} = \sum_{k=1}^K f_k(x) \quad (3)$$

where  $\hat{y}$  is the final prediction,  $f_k$  represents the prediction of the  $k^{th}$  weak model for the input  $x$ , and  $K$  is the number of weak models. In a mathematical context, XGBoost aims to minimize an objective function that comprises two primary components: a loss function measuring the variance between predicted and actual values, and a regularization term that manages the intricacy of the ensemble. Through the minimization of this function, XGBoost builds a series of decision trees that collaborate harmoniously to yield precise forecasts. The formulation of the objective function is outlined as follows:

$$\ell(\phi) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (4)$$

$$\Omega(f) = \gamma K + 0.5\lambda \sum_{j=1}^K (w_j)^2 \quad (5)$$

where  $\ell$  is the objective function,  $\phi = \{f_1, f_2, \dots, f_k\}$  shows the set of weak models,  $l$  is the loss function,  $y_i$  is the true target,  $\hat{y}_i$  is the predicted target, and  $\Omega$  is the regularization term. Additionally,  $\gamma$  and  $\lambda$  denote the complexity of the leaf node as well as the regularization term, respectively, and  $w_j$  stands for the vector of scores of leaves of  $j^{th}$  tree. By minimizing this function and effectively combining the predictions, XGBoost produces accurate results. XGBoost is intended to deliver better precision and quicker calculation speeds by utilizing several algorithmic improvements, including regularization techniques, customized loss functions, and parallel processing. These enhancements enable XGBoost to process large and intricate datasets more effectively, contributing to its widespread use in various machine learning tasks [87].

### 3.6. Light Gradient Boosting Machine (LightGBM)

LightGBM is known for its high performance and efficiency, which applies to tasks involving both regression and classification [88]. It is also an ensemble method that combines multiple weak learners to create a robust predictive model. As explained for XGBoost, LightGBM also optimizes an objective function by iteratively adding weak models to the ensemble. The key advantage of LightGBM is its ability to improve memory efficiency and computational costs through parallel learning and distributed computing, and its scalability, which can handle large datasets with numerous input variables. LightGBM employs a histogram-based approach to find the best split points, becoming significantly faster than the traditional approach of searching for the best split points during tree construction [89, 90].

## 4. Model Implementation and Results

### 4.1. Optimizing ML Methods and Hyperparameters

To assess how effectively the ML model performs and generalizes, a crucial procedure entails partitioning the dataset into random training and testing subsets. This approach enables training the model on the designated training subset and then assessing its predictive performance on the testing subset. To achieve this, the dataset of this study is randomly partitioned into distinct training and testing subsets. Specifically, 80% of the dataset is designated for training purposes, while the testing subset comprises the remaining 20%. The input variable range differs greatly, as demonstrated in Figure 3, which may reduce the efficiency of ML predictions [91]. To bring the input scales to uniformity, input variables are normalized using Equation 6:

$$\bar{x} = \frac{x-\mu}{\sigma} \tag{6}$$

where  $\bar{x}$ ,  $x$ ,  $\mu$ , and  $\sigma$  are standardized values, original values, the mean and standard deviation of original input values. The standardization is performed using the standard-scaler function available in the scikit-learn package in Python [92].

To determine the best hyperparameters for each model, random search, grid search, and K-fold cross-validation are simultaneously used. Random search is a method in machine learning for optimizing hyperparameters; it randomly samples hyperparameter combinations within defined ranges, assessing performance to discover the best set for a given model. Grid search is a reliable and commonly used method in machine learning that searches through all possible combinations of given hyperparameters to find optimal hyperparameter values [93]. Initially, hyperparameters are selected by conducting a randomized exploration through the process of random search. Following that, an optimization procedure is executed, employing a grid-search methodology to carefully fine-tune the models and obtain optimal hyperparameters, giving the highest model’s prediction accuracy. Ten-fold cross-validation improves the prediction performance and prevents probable overfitting [94]. Ten-fold cross-validation involves splitting the available data into ten equal-sized subsets or "folds." The model is trained on nine of these folds and evaluated on the remaining one. This model is implemented ten times, with each fold applied as the evaluation set once. The final evaluation metric is then calculated as the average of the ten evaluation results. The performance of ten-fold cross-validation is illustrated in Figure 5 [95]. The final hyperparameters of all ML models are presented in Table 3.

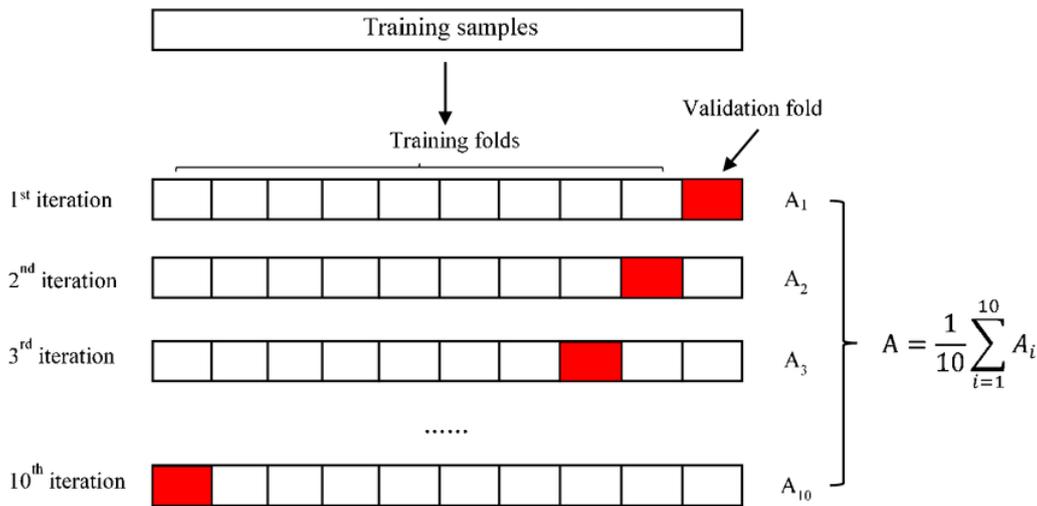


Figure 5. Ten-fold cross-validation method

Table 3. Optimal hyperparameters of candidate ML models

ML models	Optimal hyperparameters
RF	'Bootstrap': False, 'minimum samples leaf': 1, 'criterion': 'mse', 'minimum samples split': 5, 'number of estimators': 500, 'maximum features': 'sqrt'.
GB	'Subsample': 0.75, 'minimum samples split': 8, 'maximum depth': 7, 'number of estimators': 50, 'minimum samples leaf': 1, 'maximum features': 'sqrt'.
AdaBoost	'Base estimator': random forest regressor (), 'learning rate': 0.05, 'number of estimators': 100
CatBoost	'Depth': 4, 'iterations': 450, 'coefficient at the L2 regularization term': 1, 'learning rate': 0.1, 'random strength': 2, 'rsm': 0.75.
XGBoost	'Learning rate': 0.25, 'maximum depth': 6, 'reg alpha': 1, 'reg lambda': 1, 'minimum child weight': 7, 'subsample': 1, 'subsample bytree': 0.25, 'gamma': 0.
LightGBM	'Learning rate': 0.06, 'maximum number of bins': 200, 'maximum depth': 9, 'minimum data in leaf': 15, 'number of iterations': 550, 'number of leaves': 30, 'reg alpha': 1, 'reg lambda': 0.

## 4.2. Performance Indicators

The performance of ML methods is compared using different performance indicators, including the coefficient of determination ( $R^2$ ), root mean squared error (RMSE), and mean absolute error (MAE). The performance indicators equations are as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - p_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (7)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - p_i)^2} \quad (8)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - p_i| \quad (9)$$

where the  $y_i$  is the experimental value, the  $p_i$  is the prediction value obtained through machine learning algorithms,  $n$  is the number of data samples, and  $\bar{y}$  is the average value of experimental results.

## 4.3. Prediction Accuracy of ML Methods

All the ensemble models demonstrated a coefficient of determination ( $R^2$ ) greater than 0.92 for the training set and greater than 0.82 for the test set. According to Table 4, the GB, XGBoost, and LightGBM methods show the highest  $R^2$  values and the lowest RMSE, as well as the lowest MAE values on the test subset in comparison to the other methods. The  $R^2$  of GB, XGBoost, and LightGBM is 0.844, 0.854, and 0.856, respectively. The RMSE of GB, XGBoost, and LightGBM is 0.338, 0.331, and 0.325 MPa, respectively. Further, the MAE of GB, XGBoost, and LightGBM is 0.259, 0.260, and 0.255 MPa, respectively. Execution time is one of the key indices evaluating the computational cost of prediction techniques. Table 4 provides an overview of the average execution time of ML models, offering insights into their computational efficiency. AdaBoost exhibits a considerably higher execution time of 23.621 seconds compared to other ML models. The execution times of GB, XGBoost, and LightGBM are 0.084, 0.151, and 0.378 seconds, respectively.

**Table 4. Performance evaluation of ML models**

Criteria	Data	RF	GB	AdaBoost	CatBoost	XGBoost	LightGBM
$R^2$	Train	0.927	0.964	0.974	0.962	0.958	0.951
	Test	0.820	0.844	0.831	0.831	0.854	0.856
RMSE (MPa)	Train	0.287	0.201	0.168	0.206	0.216	0.235
	Test	0.369	0.338	0.350	0.354	0.331	0.325
MAE (MPa)	Train	0.177	0.126	0.119	0.140	0.132	0.145
	Test	0.264	0.259	0.263	0.263	0.260	0.255
Execution time	-	1.784	0.084	23.621	0.799	0.151	0.378

Considering all the performance indicators and computational costs, it can be postulated that GB, XGBoost, and LightGBM are the best options for predicting the nominal PSS. Their combination of high prediction accuracy and reasonable computational costs makes them the most suitable options for our current application.

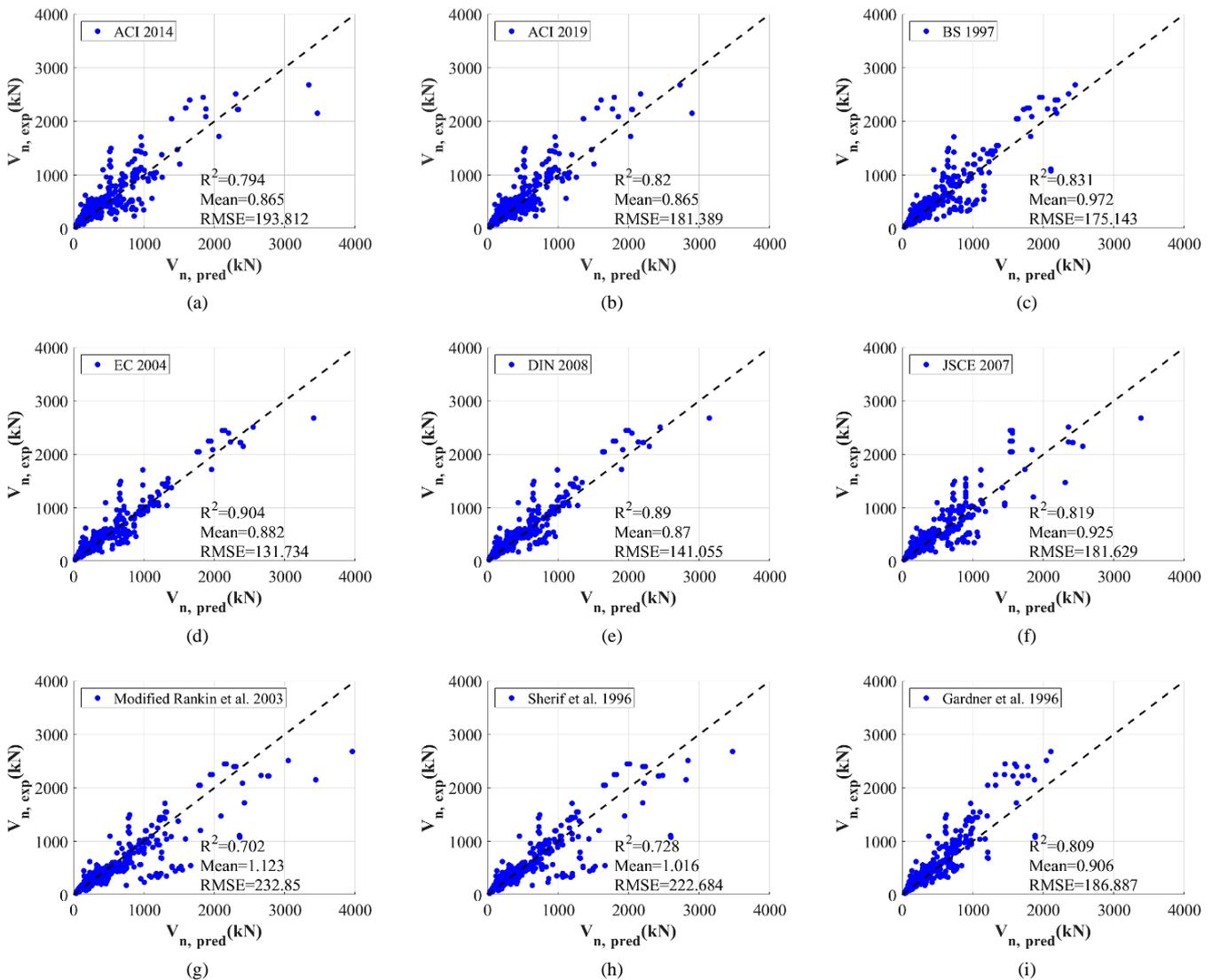
The statistical indicators for GB, XGBoost, and LightGBM are illustrated in Table 5. The proposed ML models, including GB, XGBoost, and LightGBM, exhibit a closely aligned mean of predicted to experimental PSS ratio ( $v_{c,(Pred)}/v_{c,(Exp)}$ ) ( $1.021 \pm 0.132$ ,  $1.019 \pm 0.132$ , and  $1.012 \pm 0.131$ , respectively). The latter ratio value is approximately unity (1), indicating a high correlation between predicted and experimental results.

**Table 4. Statistical indicators for proposed ML models for PSS predictions**

ML models	Ratio of predicted to experimental PSS ( $v_{c,(Pred)}/v_{c,(Exp)}$ )					
	Training set			Testing set		
	Mean	STD	COV	Mean	STD	COV
GB	1.009	0.073	0.073	1.021	0.132	0.129
XGBoost	1.009	0.080	0.079	1.019	0.132	0.129
LightGBM	1.011	0.084	0.083	1.012	0.131	0.129

### 4.4. Comparison between Prediction Accuracy of Proposed ML Models and Existing Equations

This study also evaluates and compares the performance of several machine learning models against existing code equations and equations proposed by previous researchers (see Figure 6). The codes and literature equations are listed in Table 6. So are the critical primeters used in these equations, as demonstrated in Table 7. Because of variations in the definitions of critical perimeter used in the codes and previous research, the punching shear strength ( $v_c$ ) values (in MPa) are converted to punching shear capacity ( $V_n$ ) values (in kN). This conversion allows for direct comparisons between the experimental results and the PSC obtained from ML algorithms across the entire dataset. From Figure 6, the EC-2-1-1 [5] equation is found to be the most accurate in predicting PSC among existing equations, with a highest  $R^2$  value of 0.904 and a lowest RMSE value of 131.734 kN, followed by DIN 1045-1 [96] with an  $R^2$  of 0.89 and an RMSE of 141.055 kN. Despite the inherent simplicity of ACI 318-14 [4], it yields a good correlation with experimental results with an  $R^2$  of 0.794. Considering the size factor in ACI 318-19 [73] improves the coefficient of determination by 3.2 percent compared to ACI 318-14 [4]. However, the proposed ML models of GB, XGBoost, and LightGBM provide a significantly improved set of predictions compared to existing equations, with  $R^2$  values of 0.986, 0.986, and 0.983, respectively. Additionally, the RMSE values for GB, XGBoost, and LightGBM are 50.124, 50.472, and 55.412 kN, respectively, representing less than half of the RMSE observed in the codes' formulas. The mean values of the proposed ML models are found to be close to unity (1), indicating a high proximity of predictions to experimental results. Overall, the findings of this study suggest that ML models can be effectively used to improve the prediction accuracy of PSC, surpassing the accuracy of existing codes [4, 5, 73, 74, 77, 96] and literature equations [39, 75, 76, 78]. The ML approach complements previous experimental and analytical studies, providing an accelerated and reliable method for predicting the behavior of new specimens with varying parameter values.



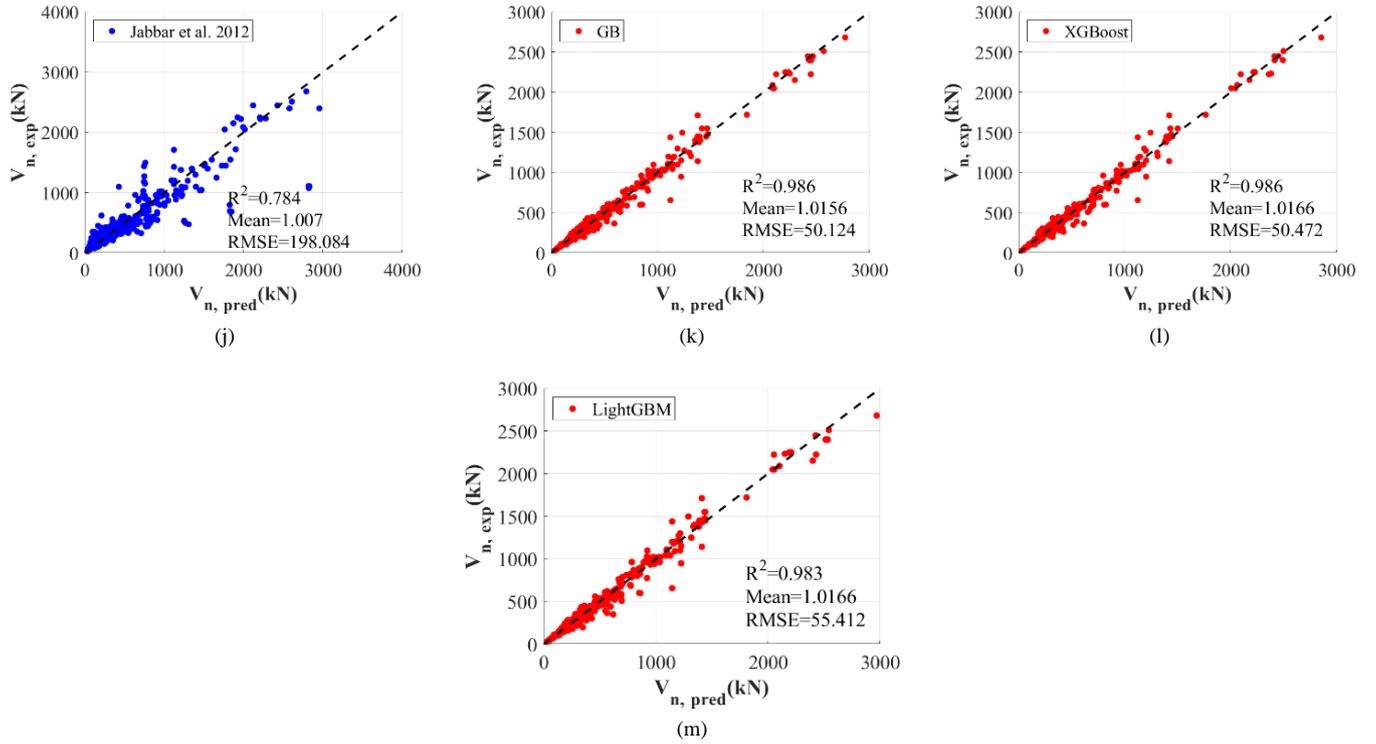
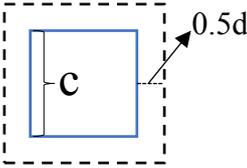
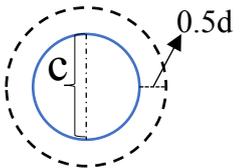
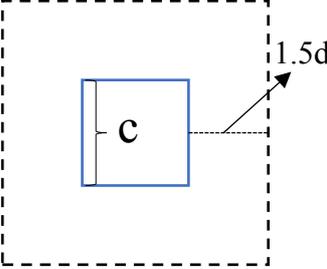
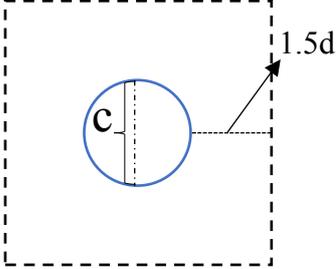
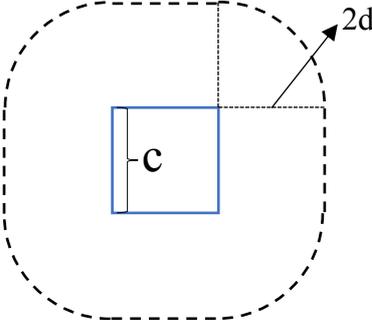
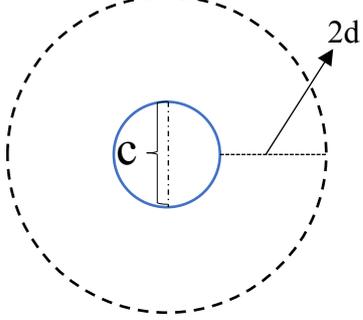
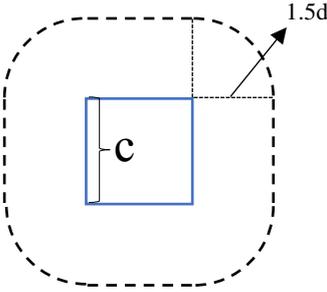
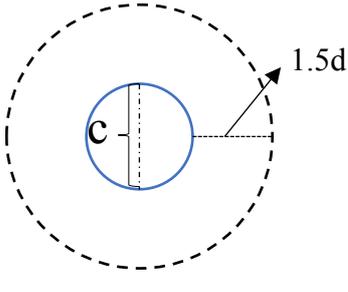


Figure 6. PSC predictions: (a) ACI 2014 (b) ACI 2019 (c) BS (d) EC (e) DIN (f) JSCE (g) Modified Rankin et al. (h) Sherif et al. (i) Gardner et al. (j) Jabbar et al. (k) GB (l) XGBoost (m) LightGBM

Table 5. Code and previous researchers' equation for punching shear

Codes	Punching shear capacity equations
ACI 318-14 (2014) [4]	$V_n = \frac{1}{3}\sqrt{f_c}b_0d$ (10)
ACI 318-19 (2019) [73]	$V_n = \min(\frac{1}{3}, \frac{1}{6}(1 + \frac{2}{\beta})), \frac{1}{12}(2 + \frac{\alpha_s d}{b_0})\lambda_s \sqrt{f_c}b_0d$ $\beta = \text{the ratio of length to width of column, } \lambda_s = \sqrt{\frac{z}{1+0.004d}}$ $\alpha_s = 40$ (For interior columns)
BS 8110-1 (1997) [74]	$V_n = 0.79 \sqrt[3]{\frac{f_{cu}}{25} (\frac{400}{d})^{(1/4)} (100\rho)^{(1/3)}} b_0d$ (12) $f_{cu} > 25\text{MPa, } (\frac{400}{d})^{(1/4)} > 1$
EC 2-1-1 (2004) [5]	$V_n = 0.18(f_c)^{(1/3)}(1 + \sqrt{\frac{200}{d}})(100\rho)^{(1/3)}b_0d$ (13) $\rho < 0.02, 1 + \sqrt{\frac{200}{d}} \leq 2$
DIN 1045-1 (2008) [96]	$V_n = 0.21(f_c)^{(1/3)}(1 + \sqrt{\frac{200}{d}})(100\rho)^{(1/3)}b_0d$ (14) $\rho < 0.02, 1 + \sqrt{\frac{200}{d}} \leq 2$
JSCE (2007) [77]	$V_n = 0.2\sqrt{f_c}(\frac{1000}{d})^{(1/4)}(100\rho)^{(1/3)}(1 + \frac{1}{1 + 0.25\frac{u}{d}})b_0d$ (15) $0.2\sqrt{f_c} < 1.2, (\frac{1000}{d})^{(1/4)} < 1.5, (100\rho)^{(1/3)} < 1.5$ $u = \text{column's perimeter}$
Rankin et al. (2003) [39, 97]	$V_n = 0.78(f_c)^{(1/3)}(100\rho)^{(1/4)}b_0d$ (16)
Sherif et al. (1996) [75]	$V_n = 0.7(f_c)^{(1/3)}(100\rho)^{(1/3)}b_0d$ (17)
Gardner et al. (1996) [76]	$V_n = 0.79(f_c)^{(1/3)}(\rho f_y)^{(1/3)}(1 + \sqrt{\frac{200}{d}})\sqrt{\frac{d}{u}}ud$ (18)
Jabbar et al. (2012) [78]	$V_n = 0.9(f_c)^{(1/3)}(\frac{200}{d})^{(1/4)}(f_y\rho)^{(1/2)}(\frac{d}{b_0})^{(1/2)}b_0d$ (19)

**Table 6. Critical perimeter for different code and literature equations**

Codes and research	$b_0$ (Square columns)	$b_0$ (Circular columns)
ACI 318-14, 19 [4, 73], JSCE [77], Rankin et al. [39, 97], Sherif et al. [75], Jabbar et al. [78]	 $b_0 = 4(c + d)$	 $b_0 = \pi(c + d)$
BS 8110-1 (1997) [74]	 $b_0 = 4(c + 3d)$	 $b_0 = 4(c + 3d)$
EC 2-1-1 (2004) [98]	 $b_0 = 4(c + \pi d)$	 $b_0 = \pi(c + 4d)$
DIN 1045-1 (2008) [96]	 $b_0 = 4c + 3\pi d$	 $b_0 = \pi(c + 3d)$

The deviation between the predicted results using proposed ML models and existing equations and experimental records is depicted in Figure 7. The ML models, including GB, XGB, and LightGBM, exhibit remarkable prediction accuracy compared to existing equations, with 97.0%, 95.0%, and 96.5% of their predictions falling within the 0-25% error range, respectively; meanwhile, 2.5%, 4.1%, and 2.8% of the predictions lie within the 25-50% error range. Less than 1% of the predicted outcomes fall within an error range exceeding 50%. The punching shear equations of ACI 318-14 [4] and ACI 318-19 [73] exhibit higher deviations from experimental findings when compared to other equations and ML models, with 46% of the predicted outcomes falling within the 0-25% error range.

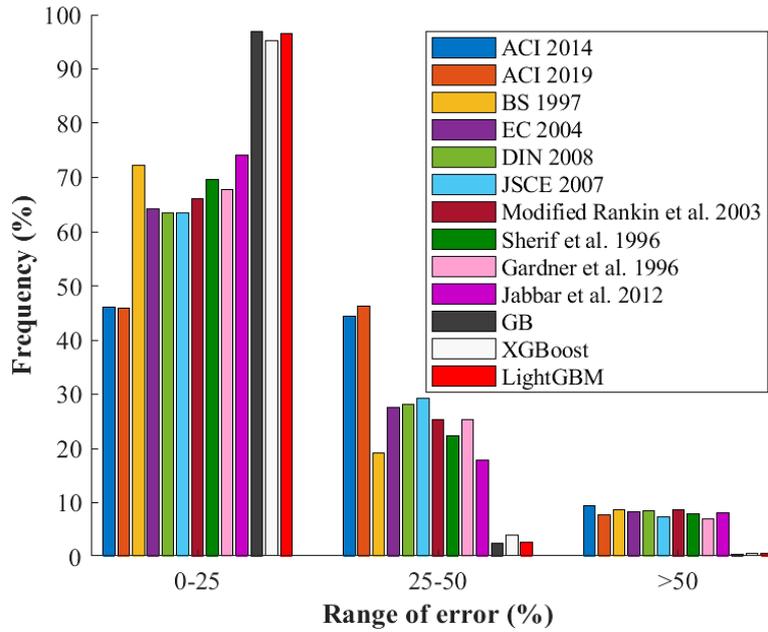


Figure 7. Comparison of error range between proposed ML models and existing equations

### 5. Feature Importance Analysis

Feature importance analysis using the permutation method is conducted to overcome the black-box nature of ML algorithms so as to evaluate the relative influence of each variable on the PSS of RC slab-column connections. This method (i.e., permutation) was introduced by Berman et al. [80]. The permutation method operates by randomly shuffling the values of a variable and observing the subsequent change in the model's prediction error as follows:

Primarily, a dataset  $A$  consisting of some samples and variables and an output variable  $Y$  are presented. Training a machine learning model follows wherein  $f(A)$  predicts  $Y$  based on  $A$ . The model's prediction error is computed based on the original data, denoted as  $E(A)$ , using a performance metric such as RMSE or other regression accuracy criteria. A randomly shuffle follows where the values of each variable  $j$  in  $A$  create a new dataset  $A'$ ; the shuffled variable values are denoted as  $(A_j)'$ . The model's prediction error is computed on the shuffled data  $A'$ , denoted as  $E((A_j)')$ . The importance score for variable  $j$  is then defined as:

$$\text{Importance score } (j) = E(A) - E((A_j)') \tag{20}$$

By repeating this process multiple times and calculating the average of the resulting importance scores, one can obtain an accurate measure of the variable's importance. If the model relies heavily on that variable, shuffling it will have a noticeable impact on the prediction. Feature selection analysis (see Figure 8) reveals that all input parameters influence the PSS. The most influential parameters are, herein, the reinforcement ratio ( $\rho$ ) and compressive strength of concrete ( $f'_c$ ); predominantly included in all existing equations. A higher reinforcement ratio contributes to resisting the widening of cracks and increases the ductility of RC slab-column connections, postponing sudden punching shear failure. Moreover, higher concrete strength delays the formation of cracks and minimizes crack width [15, 18]. The parameters of  $(d/b_0)$ ,  $(1/d)$ , and  $(f_y)$  follow in order of decreasing influence. Notably, the impact of  $f_y$  on PSS is found to be the lowest, which can be related to the fact that steel reinforcement seldom reaches its yielding strength during punching shear failure. Notably, even though the input parameters were selected based on literature, including ACI 318-14 [4], EC 2-1-1 [5], DIN 1045-1 [96], Zhang et al. [97], Elshafey et al. [15] and Jabbar et al. [78], there is the potential influence of other parameters on the PSS of RC slab-column connections. However, this study's remarkably high prediction accuracy ( $R^2 = 0.986$ ) suggests that incorporating additional parameters may not significantly enhance the prediction accuracy.

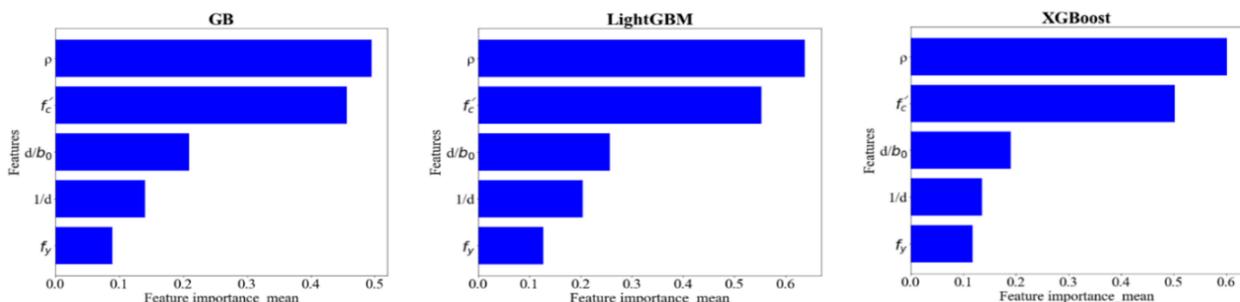


Figure 8. Feature importance analysis using proposed ML models

### 6. SHAP Dependence Analysis

SHAP dependence analysis is a powerful tool in machine learning that helps improve model interpretability and understanding. It isolates the impact of individual predictor variables on the target variable while holding all other predictors constant. This analysis is implemented, and its outcomes are depicted in Figure 9. As can be seen, there is a clear correlation between  $f'_c$  and PSS. This correlation arises because PSS is dependent on the tensile strength of concrete [15, 99], which, in turn, is non-linearly influenced by its compressive strength. Based on SHAP dependence analysis, the correlation between  $f'_c$  and PSS is for  $f'_c$  up to 80 MPa (see Figure 9-a). Furthermore, PSC increases as the reinforcement ratio ( $\rho$ ) increases. This applies to reinforcement ratios up to 2% (see Figure 9-b), confirming the limitations of EC 2-1-1 [5] and DIN 1045-1 [96]. PSC is not highly influenced by steel yield stress; the latter is not reached prior to the occurrence of punching failure. With an effective depth decrease, nominal PSS increases. Marzouk et al. [100] found that as the slab thickness and effective depth increase, the brittleness of the concrete slab increases, leading to lower shear strength. Nominal PSS is non-linearly related to the depth-critical perimeter ratio ( $d/b_0$ ). This is also supported by literature studies and codes [76-78].

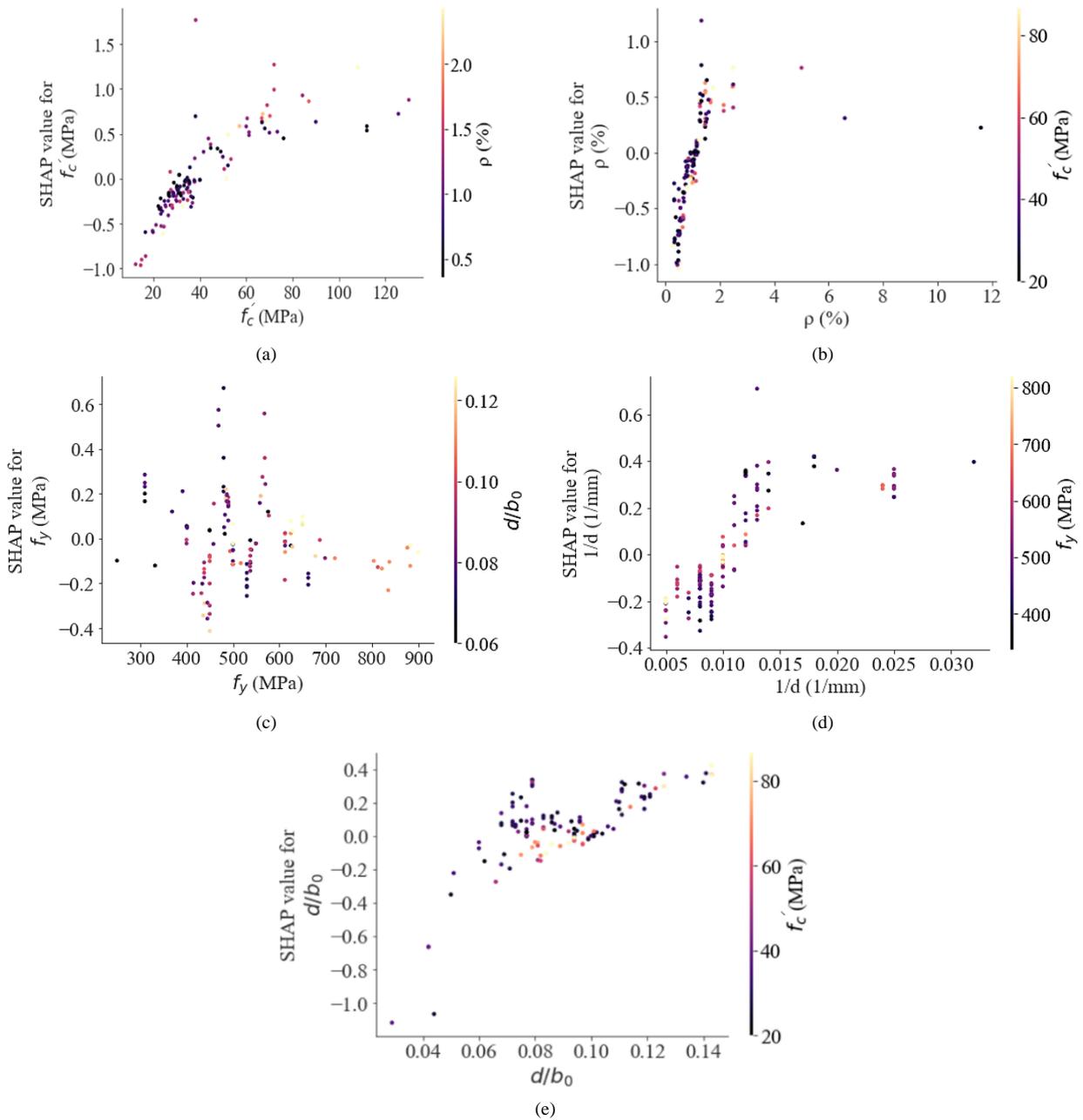


Figure 9. SHAP dependence plots

According to SHAP dependence analysis, the equation for calculating the nominal PSS is assumed as follows:

$$v = c_1 f'_c{}^{c_2} \rho^{c_3} \left(\frac{c_4}{d}\right)^{c_5} \left(\frac{d}{b_0}\right)^{c_6} \quad 0 < c_i \leq 1, i = 2, 3, 4, 5, 6 \tag{21}$$

## 7. Development of Graphical User Interface (GUI)

A GUI is developed using proposed ML models for predicting the PSC of slab-column connections (see Figure 10). This user-friendly interface simplifies the PSC prediction process, ensuring accessibility for all users regardless of their level of programming expertise. In this interface, the user can simply enter numeric values of  $f'_c$ ,  $\rho$ ,  $1/d$ ,  $f_y$ , and  $d/b_0$  to obtain the PSC of slab-column connections.

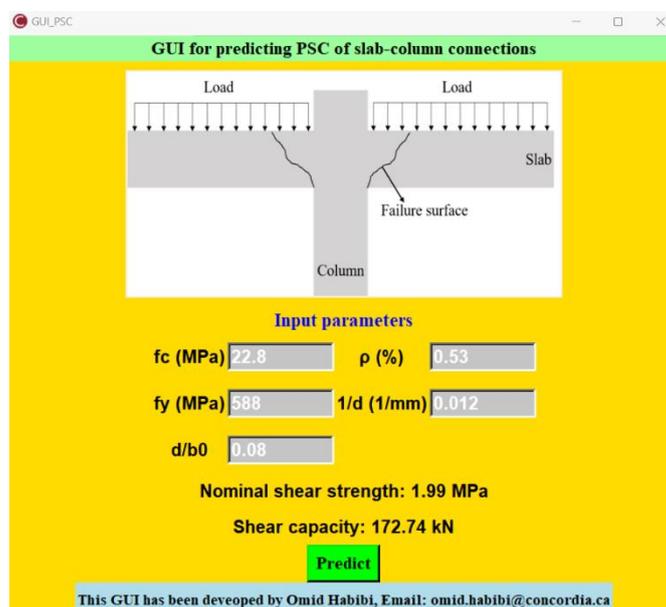


Figure 10. GUI for predicting the PSC

## 8. Conclusions

This study aims to investigate the punching shear strength of reinforced concrete slab-column connections using machine learning algorithms. A comprehensive dataset, including 616 experimental results, is collected from previous studies. Six ensemble machine learning algorithms, including RF, GB, AdaBoost, CatBoost, XGBoost, and LightGBM, are utilized to suggest the most accurate model for estimating the punching shear strength. The major findings of this study are as follows:

- Results demonstrate that GB, XGBoost, and LightGBM deliver the highest  $R^2$  and lowest RMSE and MAE compared to other machine learning methods for both training and testing subsets. These three methods display low computational cost and, in turn, efficiency. These models are proposed for predicting punching shear strength.
- The accuracy of the proposed ML models in predicting the punching shear capacity of slab-column connections is compared with existing formulas in codes and the literature. The BS EN 1992-2 [5] yields the highest performance among existing equations, achieving an  $R^2$  of 0.904 and an RMSE of 131.734 kN. For the simple formula of ACI 318-14 [4], a relatively reasonable correlation ( $R^2 = 0.794$ ) is obtained. Incorporating the size factor in ACI 318-19 [73] enhances the coefficient of determination by 3.2% compared to ACI 318-14 [4]. However, the proposed ML models manifest a remarkable enhancement in predicting punching shear capacity, surpassing the accuracy of existing equations. Among these models, GB exhibits the highest level of prediction accuracy. The  $R^2$  values are notably higher than existing equations at 0.986, 0.986, and 0.983 for GB, XGBoost, and LightGBM, respectively. The corresponding RMSE values are 50.124, 50.472, and 55.412 kN, respectively.
- A feature importance analysis is conducted to investigate the importance of each variable in punching shear strength, confirming the strongest influence is related to reinforcement ratio ( $\rho$ ) and concrete compressive strength ( $f'_c$ ).
- The influence of steel yield strength ( $f_y$ ) on punching shear strength is insignificant; for the pre-designed slab-column connections, steel rebar generally does not reach its yield stress prior to punching failure.
- SHAP dependence analysis is conducted to better understand the effect of each variable on punching shear strength. It is concluded that  $\rho$ ,  $f'_c$ ,  $1/d$  and  $d/b_0$  are non-linearly proportional to nominal punching shear strength. This trend is observed for  $\rho$  extending up to 2%, thereby affirming the constraints outlined in EC 2-1-1 [5] and DIN 1045-1 [96].
- One of the limitations of this study is that it does not consider the crack and failure patterns of concrete. Analyzing crack patterns requires experimental investigation. Hence, it is suggested that the crack and failure pattern of concrete be documented using images of punching shear failures, and image processing techniques (e.g., deep learning) be applied for prediction and modeling in future studies.

## 9. Declarations

### 9.1. Author Contributions

Conceptualization, O.H., T.Y., and K.I.; methodology, O.H.; software, O.H. and H.N.; formal analysis, O.H. and K.I.; investigation, O.H. and K.I.; data curation, T.Y.; writing—original draft preparation, O.H.; writing—review and editing, O.H., T.Y., H.N., and K.I.; supervision, T.Y. and K.I. All authors have read and agreed to the published version of the manuscript.

### 9.2. Data Availability Statement

The dataset utilized in this study has been uploaded to <https://github.com/Omid1373/PSC>, and the machine learning codes are available on request from the corresponding author.

### 9.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

### 9.4. Conflicts of Interest

The authors declare no conflict of interest.

## 10. References

- [1] Rochdi, E. H., Bigaud, D., Ferrier, E., & Hamelin, P. (2006). Ultimate behavior of CFRP strengthened RC flat slabs under a centrally applied load. *Composite Structures*, 72(1), 69–78. doi:10.1016/j.compstruct.2004.10.017.
- [2] Nguyen, K. L., Trinh, H. T., & Pham, T. M. (2024). Prediction of punching shear strength in flat slabs: ensemble learning models and practical implementation. *Neural Computing and Applications*, 36(8), 4207–4228. doi:10.1007/s00521-023-09296-0.
- [3] Tran, V. L., & Kim, S. E. (2021). A practical ANN model for predicting the PSS of two-way reinforced concrete slabs. *Engineering with Computers*, 37(3), 2303–2327. doi:10.1007/s00366-020-00944-w.
- [4] ACI 318-14. (2014). Building Code Requirements for Structural Concrete (ACI 318-14): An ACI Standard; Commentary on Building Code Requirements for Structural Concrete (ACI 318R-14). American Concrete Institute (ACI), Michigan, United States.
- [5] BS EN 1992-2. (2005). Eurocode 2: Design of concrete structures. Concrete bridges. Design and detailing rules. European Committee for Standardization, Brussels, Belgium.
- [6] JSCE NO. 15. (2010). Standard Specifications for Concrete. Structures – 2007: Design. Japan Society of Civil Engineers (JSCE), Tokyo, Japan.
- [7] Nguyen, H. D., Truong, G. T., & Shin, M. (2021). Development of extreme gradient boosting model for prediction of punching shear resistance of r/c interior slabs. *Engineering Structures*, 235, 112067. doi:10.1016/j.engstruct.2021.112067.
- [8] Mellios, N., Uz, O., & Spyridis, P. (2023). Data-based modeling of the punching shear capacity of concrete structures. *Engineering Structures*, 275, 115195. doi:10.1016/j.engstruct.2022.115195.
- [9] El-said, A., Deifalla, A. F., Yousef, S. E. A. S., El-Sayed, T. A., Tawfik, M., & Ayash, N. M. (2023). Code provisions evaluation for the punching shear capacity of R.C footings without reinforcement for punching shear. *Case Studies in Construction Materials*, 18, e02182. doi:10.1016/j.cscm.2023.e02182.
- [10] Habibi, O., Khaloo, A., & Abdoos, H. (2021). Seismic behavior comparison of RC shear walls strengthened using FRP composites and steel elements. *Scientia Iranica*, 28(3 A), 1–27. doi:10.24200/sci.2020.55328.4170.
- [11] Khaloo, A., Borhani, M. H., Habibi, O., Tabatabaeian, M., & Askari, S. M. (2024). Experimental and numerical investigation on the performance of GFRP-confined expansive concrete-filled unplasticized polyvinyl chloride tubes. *Journal of Thermoplastic Composite Materials*, 37(3), 983–1011. doi:10.1177/08927057231190558.
- [12] Adeli, H. (2001). Neural networks in civil engineering: 1989-2000. *Computer-Aided Civil and Infrastructure Engineering*, 16(2), 126–142. doi:10.1111/0885-9507.00219.
- [13] Salehi, H., & Burgueño, R. (2018). Emerging artificial intelligence methods in structural engineering. *Engineering Structures*, 171, 170–189. doi:10.1016/j.engstruct.2018.05.084.
- [14] Thai, H.-T. (2022). Machine learning for structural engineering: A state-of-the-art review. *Structures*, 38, 448–491. doi:10.1016/j.istruc.2022.02.003.
- [15] Elshafey, A. A., Rizk, E., Marzouk, H., & Haddara, M. R. (2011). Prediction of punching shear strength of two-way slabs. *Engineering Structures*, 33(5), 1742–1753. doi:10.1016/j.engstruct.2011.02.013.

- [16] ACI 318-08. (2008). Building Code Requirements for Structural Concrete (ACI 318-08): An ACI Standard; Commentary on Building Code Requirements for Structural Concrete (ACI 318R-08). American Concrete Institute (ACI), Michigan, United States.
- [17] Chetchotisak, P., Ruengpim, P., Chetchotsak, D., & Yindeesuk, S. (2018). Punching Shear Strengths of RC Slab-Column Connections: Prediction and Reliability. *KSCE Journal of Civil Engineering*, 22(8), 3066–3076. doi:10.1007/s12205-017-0456-6.
- [18] Faridmehr, I., Nehdi, M. L., & Hajmohammadian Baghban, M. (2022). Novel informational bat-ANN model for predicting punching shear of RC flat slabs without shear reinforcement. *Engineering Structures*, 256, 114030. doi:10.1016/j.engstruct.2022.114030.
- [19] Naseri Nasab, M., Jahangir, H., Hasani, H., Majidi, M. H., & Khorashadizadeh, S. (2023). Estimating the punching shear capacities of concrete slabs reinforced by steel and FRP rebars with ANN-Based GUI toolbox. *Structures*, 50, 1204–1221. doi:10.1016/j.istruc.2023.02.072.
- [20] Choi, K. K., Reda Taha, M. M., & Sherif, A. G. (2007). Simplified punching shear design method for slab-column connections using fuzzy learning. *ACI Structural Journal*, 104(4), 438–447. doi:10.14359/18774.
- [21] Akbarpour, H., & Akbarpour, M. (2017). Prediction of punching shear strength of two-way slabs using artificial neural network and adaptive neuro-fuzzy inference system. *Neural Computing and Applications*, 28(11), 3273–3284. doi:10.1007/s00521-016-2239-2.
- [22] 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. doi:10.1016/j.jobe.2021.102300.
- [23] Cao, M. T. (2023). Advanced soft computing techniques for predicting punching shear strength. *Journal of Building Engineering*, 79, 107800. doi:10.1016/j.jobe.2023.107800.
- [24] Wu, Y., & Zhou, Y. (2023). Prediction and feature analysis of punching shear strength of two-way reinforced concrete slabs using optimized machine learning algorithm and Shapley additive explanations. *Mechanics of Advanced Materials and Structures*, 30(15), 3086–3096. doi:10.1080/15376494.2022.2068209.
- [25] Ng, W., Minasny, B., de Sousa Mendes, W., & Melo Demattê, J. A. (2020). The influence of training sample size on the accuracy of deep learning models for the prediction of soil properties with near-infrared spectroscopy data. *Soil*, 6(2), 565–578. doi:10.5194/soil-6-565-2020.
- [26] Degtyarev, V. V. (2022). Machine Learning Models for Predicting Bond Strength of Deformed Bars in Concrete. *ACI Structural Journal*, 119(5), 43–56. doi:10.14359/51734833.
- [27] Shahani, N. M., Kamran, M., Zheng, X., Liu, C., & Guo, X. (2021). Application of gradient boosting machine learning algorithms to predict uniaxial compressive strength of soft sedimentary rocks at Thar coalfield. *Advances in Civil Engineering*, 2021, 1–19. doi:10.1155/2021/2565488.
- [28] Ibrahim, A. A., Ridwan, R. L., Muhammed, M. M., Abdulaziz, R. O., & Saheed, G. A. (2020). Comparison of the CatBoost Classifier with other Machine Learning Methods. *International Journal of Advanced Computer Science and Applications*, 11(11), 738–748. doi:10.14569/IJACSA.2020.0111190.
- [29] Liu, Y., Lyu, C., Khadka, A., Zhang, W., & Liu, Z. (2020). Spatio-Temporal Ensemble Method for Car-Hailing Demand Prediction. *IEEE Transactions on Intelligent Transportation Systems*, 21(12), 5328–5333. doi:10.1109/TITS.2019.2948790.
- [30] Cai, W., Wei, R., Xu, L., & Ding, X. (2022). A method for modelling greenhouse temperature using gradient boost decision tree. *Information Processing in Agriculture*, 9(3), 343–354. doi:10.1016/j.inpa.2021.08.004.
- [31] Shanguan, Q., Fu, T., Wang, J., Fang, S., & Fu, L. (2022). A proactive lane-changing risk prediction framework considering driving intention recognition and different lane-changing patterns. *Accident Analysis and Prevention*, 164, 106500. doi:10.1016/j.aap.2021.106500.
- [32] Xia, Y., He, L., Li, Y., Liu, N., & Ding, Y. (2020). Predicting loan default in peer-to-peer lending using narrative data. *Journal of Forecasting*, 39(2), 260–280. doi:10.1002/for.2625.
- [33] Sih, N.-S. (1957). Shearing Strength of Reinforced Concrete Slabs. *Journal of the Structural Division*, 83(1), 29–58. doi:10.1061/jsdeag.0000081.
- [34] Kinnunen, S., & Nylander, H. S. E. (1960). Punching of concrete slabs without shear reinforcement. *Elander, Mölnlycke, Sweden*.
- [35] Mowrer, R. D., & Vanderbilt, M. D. (1967). Shear Strength of Lightweight Aggregate Reinforced Concrete Flat Plates. *ACI Journal Proceedings*, 64(11). doi:10.14359/7601.

- [36] Hanson, N. W., & Hanson, J. M. (1968). Shear and moment transfer between concrete slabs and columns. Portland Cement Association, Research and Development Laboratories, Chicago, United States.
- [37] Kinnunen S, Nylander H, T. P. (1978). Investigations on punching at the division of building statics and structural engineering. *Nordisk Betong*, 3, 25–27.
- [38] Regan, P., Walker, P., & Zakaria, K. (1979). Tests of reinforced concrete flat slabs, CIRIA Project No. RP 220. Polytechnic of Central London, London, United Kingdom.
- [39] Rankin, G. I. B., & Long, A. E. (2019). Punching strength of conventional slab-column specimens. *Engineering Structures*, 178(2), 37–54. doi:10.1016/j.engstruct.2018.10.014.
- [40] Gardner, N. J. (1990). Relationship of the punching shear capacity of reinforced concrete slabs with concrete strength. *ACI Structural Journal*, 87(1), 66–71. doi:10.14359/2932.
- [41] Lovrovich, J. S., & McLean, D. I. (1990). Punching shear behavior of slabs with varying span-depth ratios. *ACI Structural Journal*, 87(5), 507–511. doi:10.14359/2616.
- [42] McLean, D. I., Phan, L. T., Lew, H. S., & White, R. N. (1990). Punching shear behavior of lightweight concrete slabs and shells. *ACI Structural Journal*, 87(4), 386–392. doi:10.14359/2735.
- [43] Marzouk, H., & Hussein, A. (1991). Punching shear analysis of reinforced high-strength concrete slabs. *Canadian Journal of Civil Engineering*, 18(6), 954–963. doi:10.1139/191-118.
- [44] Alexander, S. D. B., & Simmonds, S. H. (1992). Tests of column-flat plate connections. *ACI Structural Journal*, 89(5), 495–502. doi:10.14359/2948.
- [45] Marzouk, H., & Hussein, A. (1991). Experimental investigation on the behavior of high-strength concrete slabs. *ACI Structural Journal*, 88(6), 701–713. doi:10.14359/1261.
- [46] Tomaszewicz, A. (1993). Punching shear capacity of reinforced concrete slabs. High Strength Concrete SP2-Plates and Shells. Report 2.3. Report No. STF70A93082. SINTEF, Trondheim, Norway.
- [47] Fang, I. K., Lee, J. H., & Chen, C. R. (1994). Behavior of partially restrained slabs under concentrated load. *ACI Structural Journal*, 91(2), 133–139. doi:10.14359/4557.
- [48] Collins, M. P., Mitchell, D., Adebar, P., & Vecchio, F. J. (1996). A general shear design method. *ACI Structural Journal*, 93(1), 36–45. doi:10.14359/9838.
- [49] Hallgren, M. (1998). Punching shear capacity of reinforced high-strength concrete slabs. Ph.D. Thesis, Royal Institute of Technology, Stockholm, Sweden.
- [50] Marzouk, H., & Jiang, D. (1997). Experimental investigation on shear enhancement types for high-strength concrete plates. *ACI Structural Journal*, 94(1), 49–58. doi:10.14359/460.
- [51] Ghannoum, C. M. (1998). Effect of high-strength concrete on the performance of slab-column specimens. Master Thesis, McGill University, Montreal, Canada.
- [52] Marzouk, H., Emam, M., & Hilal, M. S. (1998). Effect of high-strength concrete slab on the behavior of slab-column connections. *ACI Structural Journal*, 95(3), 227–237. doi:10.14359/9713.
- [53] Broms, C. E. (2000). Elimination of flat plate punching failure mode. *ACI Structural Journal*, 97(1), 94–101. doi:10.14359/838.
- [54] Li, K. K. L. (2000). Influence of size on punching shear strength of concrete slabs. Master Thesis, McGill University, Montreal, Canada.
- [55] McHarg, P. J., Cook, W. D., Mitchell, D., & Yoon, Y. S. (2000). Improved transmission of high-strength concrete column loads through normal strength concrete slabs. *ACI Structural Journal*, 97(1), 157–165. doi:10.14359/845.
- [56] Osman, M., Marzouk, H., & Helmy, S. (2000). Behavior of high-strength lightweight concrete slabs under punching loads. *ACI Structural Journal*, 97(3), 492–498. doi:10.14359/4644.
- [57] Guandalini, S., & Muttoni, A. (2004). Symmetrical punching tests on slabs without transverse reinforcement. Test Report, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland.
- [58] Sundquist, H., & Kinnunen, S. (2004). The effect of column head and drop panels on the punching capacity of flat slabs. *Bulletin*, 82.
- [59] Abdel Hafez, A. M. (2005). Punching shear behavior of normal and high-strength concrete slabs under static loading. *Journal of Engineering Sciences*, 33(4), 1215–1235.
- [60] Ozden, S., Ersoy, U., & Ozturan, T. (2006). Punching shear tests of normal- and high-strength concrete flat plates. *Canadian Journal of Civil Engineering*, 33(11), 1389–1400. doi:10.1139/L06-089.

- [61] Birkle, G., & Dilger, W. H. (2008). Influence of slab thickness on punching shear strength. *ACI Structural Journal*, 105(2), 180–188. doi:10.14359/19733.
- [62] Lee, J.-H., Yoon, Y.-S., Lee, S.-H., Cook, W. D., & Mitchell, D. (2008). Enhancing Performance of Slab-Column Connections. *Journal of Structural Engineering*, 134(3), 448–457. doi:10.1061/(asce)0733-9445(2008)134:3(448).
- [63] Metwally, I. M., Issa, M. S., & El-Betar, S. A. (2008). Punching shear resistance of normal and high strength reinforced concrete flat slabs. *Civ Eng Res Mag*, 30(3), 982-1003.
- [64] Marzouk, R., & Rizk, E. (2009). Punching analysis of reinforced concrete two-way slabs. Research Report RCS01, Faculty of Engineering and Applied Science, Memorial University of Newfoundland St. John's, Newfoundland, Canada.
- [65] Inácio, M. M. G., Almeida, A. F. O., Faria, D. M. V., Lúcio, V. J. G., & Ramos, A. P. (2015). Punching of high strength concrete flat slabs without shear reinforcement. *Engineering Structures*, 103, 275–284. doi:10.1016/j.engstruct.2015.09.010.
- [66] Einpaul, J., Bujnak, J., Ruiz, M. F., & Muttoni, A. (2016). Study on influence of column size and slab slenderness on punching strength. *ACI Structural Journal*, 113(1), 135–146. doi:10.14359/51687945.
- [67] Francesconi, L., Pani, L., & Stochino, F. (2016). Punching shear strength of reinforced recycled concrete slabs. *Construction and Building Materials*, 127, 248–263. doi:10.1016/j.conbuildmat.2016.09.094.
- [68] Teng, S., Chanthabouala, K., Lim, D. T. Y., & Hidayat, R. (2018). Punching shear strength of slabs and influence of low reinforcement ratio. *ACI Structural Journal*, 115(6), 1816. doi:10.14359/51701089.
- [69] Urban, T., Gołdyn, M., Krawczyk, Ł., & Sowa, Ł. (2019). Experimental investigations on punching shear of lightweight aggregate concrete flat slabs. *Engineering Structures*, 197, 109371. doi:10.1016/j.engstruct.2019.109371.
- [70] Shaaban, I. G., Hosni, A. H., Montaser, W. M., & El-Sayed, M. M. (2020). Effect of premature loading on punching resistance of reinforced concrete flat slabs. *Case Studies in Construction Materials*, 12, 320. doi:10.1016/j.cscm.2019.e00320.
- [71] Sahoo, S., & Singh, B. (2021). Punching shear capacity of recycled-aggregate concrete slab-column connections. *Journal of Building Engineering*, 41, 102430. doi:10.1016/j.jobe.2021.102430.
- [72] Qian, K., Li, J. S., Huang, T., Weng, Y. H., & Deng, X. F. (2022). Punching shear strength of corroded reinforced concrete slab-column connections. *Journal of Building Engineering*, 45, 103489. doi:10.1016/j.jobe.2021.103489.
- [73] ACI 318-19. (2019). *Building Code Requirements for Structural Concrete (ACI 318-19): An ACI Standard; Commentary on Building Code Requirements for Structural Concrete (ACI 318R-19)*. American Concrete Institute (ACI), Michigan, United States.
- [74] BS 8110-1-1997. (1997). *Structural use of concrete, part 1: Code of practice for design and construction*. British Standards Institution (BSI), London, United Kingdom.
- [75] Sherif, A. G., & Dilger, W. H. (1996). Critical review of the CSA A23.3-94 punching shear strength provisions for interior columns. *Canadian Journal of Civil Engineering*, 23(5), 998–1011. doi:10.1139/196-907.
- [76] Gardner, N. J., & Shao, X. Y. (1996). Punching shear of continuous flat reinforced concrete slabs. *ACI Structural Journal*, 93(2), 218–228. doi:10.14359/1494.
- [77] Japan Society of Civil Engineers (JSCE). (2007). *Standard Specifications for Concrete Structures*. Japan Society of Civil Engineers (JSCE), Tokyo, Japan.
- [78] Jabbar, A. S. A., Alam, M. A., & Mustapha, K. N. (2012). A new equation for predicting punching shear strength of R/C flat plates. *Proceedings National Graduate Conference*, 8-10 November, 2012, University Tenaga Nasional, Putrajaya Campus, Kajang, Malaysia.
- [79] Freedman, D., Pisani, R., & Purves, R. (2020). *Statistics: Fourth international student edition*. W. W. Norton & Company, New York, United States.
- [80] Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32. doi:10.1023/A:1010933404324
- [81] Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, 88(11), 2783–2792. doi:10.1890/07-0539.1.
- [82] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232. doi:10.1214/aos/1013203451.
- [83] Drucker, H. (1997). Improving regressors using boosting techniques. *Proceedings of the Fourteenth International Conference on Machine Learning*, 8-12 July, 1997, San Francisco, United States.
- [84] chapiro, R. E. (2013). Explaining AdaBoost. *Empirical Inference*, 37–52. doi:10.1007/978-3-642-41136-6\_5.

- [85] Dhananjay, B., & Sivaraman, J. (2021). Analysis and classification of heart rate using CatBoost feature ranking model. *Biomedical Signal Processing and Control*, 68, 102610. doi:10.1016/j.bspc.2021.102610.
- [86] CatBoost. (2024). CatBoost is a machine learning algorithm that uses gradient boosting on decision trees. Available online: <https://catboost.ai/docs/> (accessed on May 2024).
- [87] Chen, T., & Guestrin, C. (2016). XGBoost. *Proceedings of the 22<sup>nd</sup> ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. doi:10.1145/2939672.2939785.
- [88] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 2017-December, 3147–3155.
- [89] Zhou, Z., Wang, M., Huang, J., Lin, S., & Lv, Z. (2022). Blockchain in Big Data Security for Intelligent Transportation with 6G. *IEEE Transactions on Intelligent Transportation Systems*, 23(7), 9736–9746. doi:10.1109/TITS.2021.3107011.
- [90] Yang, S., & Zhang, H. (2018). Comparison of Several Data Mining Methods in Credit Card Default Prediction. *Intelligent Information Management*, 10(05), 115–122. doi:10.4236/iim.2018.105010.
- [91] Shirzadi Javid, A. A., Naseri, H., & Etebari Ghasbeh, M. A. (2021). Estimating the Optimal Mixture Design of Concrete Pavements Using a Numerical Method and Meta-heuristic Algorithms. *Iranian Journal of Science and Technology - Transactions of Civil Engineering*, 45(2), 913–927. doi:10.1007/s40996-020-00352-6.
- [92] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, 2825-2830.
- [93] Naseri, H., Waygood, E. O. D., Wang, B., & Patterson, Z. (2022). Application of Machine Learning to Child Mode Choice with a Novel Technique to Optimize Hyperparameters. *International Journal of Environmental Research and Public Health*, 19(24), 16844. doi:10.3390/ijerph192416844.
- [94] Franklin, J. (2005). The elements of statistical learning: data mining, inference and prediction. *Mathematical Intelligencer*, 27(2), 83-85. doi:10.1007/BF02985802.
- [95] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York, United States. doi:10.1007/978-0-387-84858-7.
- [96] DIN 1045-1 (2008). Plain, Reinforced and Prestressed Concrete Structures. Design and Construction. Deutsches Institut für Normung (DIN), Berlin, Germany.
- [97] Zhang, Q. I. (2003). The punching strength of high strength flat slabs: Experimental study. Research Study, Memorial University of Newfoundland St. John's, Newfoundland, Canada.
- [98] BS EN 1992-1-1. (2004). Eurocode 2: Design of concrete structures: Part 1-1: General rules and rules for buildings. British Standards Institution. British Standards Institution, London, United Kingdom.
- [99] Marzouk, H., & Chen, Z. W. (1995). Fracture Energy and Tension Properties of High-Strength Concrete. *Journal of Materials in Civil Engineering*, 7(2), 108–116. doi:10.1061/(asce)0899-1561(1995)7:2(108).
- [100] Marzouk, H., Emam, M., & Hilal, M. S. (1998). Sensitivity of shear strength to fracture energy of high-strength concrete slabs. *Canadian Journal of Civil Engineering*, 25(1), 40–50. doi:10.1139/197-053.