Journal of Science and Transport Technology
Journal homepage: https://jstt.vn/index.php/en

Article info
Type of article:
Original research paper

DOI:
https://doi.org/10.58845/jstt.utt.2026.en.6.2.429-454

***Corresponding author:**
Email address:
anhnt@utt.edu.vn

Received: 12/04/2025
Received in Revised Form: 01/11/2026
Accepted: 29/05/2026

Accurate Shear Strength Prediction for RCDB via Voting XGBoost-CatBoost Model

Phuong Thao Vu[1], Thuy-Anh Nguyen[2*], Hai Minh Le[2], Hoang-Anh Vu[2]
[1]University of Transport and Communications
[2]University of Transport Technology, Hanoi 100000, Vietnam

Abstract: Accurate prediction of shear strength (SS) in reinforced concrete deep beams (RCDB) is significant for structural design, yet complexities arise from non-linear behavior which traditional methods may not sufficiently capture. This work examines the application of machine learning (ML) techniques, namely XGBoost (XGBR), CatBoost (CBR), and ensemble models (Stacking, Voting), to predict the shear strength of RCDB utilizing a database of 840 experimental results. Models were developed and evaluated using metrics including R^2, MAE, RMSE, MAPE, and A20. The Voting ensemble model, a combination of XGBoost and CatBoost, demonstrated the highest performance on the test set, achieving an R^2 value of 0.961 and an RMSE of 131.8 kN, outperforming individual models and other ML algorithms such as Support Vector Regression (SVR), Multi-Layer Perceptron (MLP), and Random Forest (RF). Additionally, the Voting model demonstrated significant improvements over traditional design standards, ACI 318 and Eurocode 2, with a higher R^2 and lower RMSE metrics. SHapley Additive exPlanations (SHAP) analysis was applied for model interpretation, showing that beam width, shear span-to-depth ratio, beam height, and concrete strength were the most influential parameters, aligning with established structural engineering principles. The results suggest that the proposed Voting ensemble model provides a highly accurate and interpretable alternative to conventional design codes for predicting the shear strength of RCDB, offering both prediction accuracy and parameter explanation.

Keywords: Reinforced Concrete Deep Beams (RCDB); Machine Learning (ML); Shear Strength (SS); Ensemble Learning; Model Interpretation.

1. Introduction

Reinforced concrete deep beams (RCDB) are structural components used in structures such as bridges, transfer girders, and industrial buildings, owing to their capacity for transferring heavy loads over short spans [1]. RCDB generally possess a small shear span-to-depth ratio, typically below 2.5 [2]. In contrast to slender beams, RCDB exhibit different shear mechanisms, mainly involving substantial arch action and nonlinear stress distribution. Shear failure in these beams involves complex patterns of stress redistribution, influenced by factors including effective depth, shear span-to-depth ratio, longitudinal and transverse reinforcement ratios, and concrete compressive strength [3, 4]. This interaction between geometric and material parameters leads to non-linear behavior that is challenging for simplified analytical models to represent accurately.

Traditional approaches for estimating RCDB shear strength primarily rely on semi-empirical models [5–7], and most of the code formulassuch as ACI 318 [8] and Eurocode 2 [9], are also semi-empirical models. In addition, based on the strut-and-tie model (STM), theoretical methods to calculate the shear strength of RCDB are proposed [10–13]. While these methods provide a basis for design, they often suffer from significant limitations. Code-based formulas are frequently found to be overly conservative for certain parameter ranges, leading to inefficient material use. Furthermore, these conventional methods struggle to comprehensively capture the complex, non-linear interactions among the numerous influencing variables, often resulting in considerable scatter when their predictions are compared against experimental data [14].

ML techniques offer an alternative approach to address the limitations found in traditional methods. These techniques utilize data to construct models for the shear strength of RCDB, potentially yielding improved accuracy and adaptability under various conditions. ML algorithms, such as support vector machines (SVMs), artificial neural networks (ANNs), genetic algorithms (GAs), and ensemble learning methods, can recognize complex, nonlinear relationships independent of predefined assumptions. These methods have been applied to the prediction of shear strength for RCDB [2,15–18]. For example, a study by Ma et al. [2] applied six ML models to a dataset of 457 RCDB tests, where the XGBoost model achieved a high correlation coefficient (R^2 = 0.917) on the test set. Other research has similarly demonstrated the performance of specific models. Nguyen et al. [19] reported that Gaussian Process Regression (GPR) provided accurate predictions, while studies by Feng et al. [15] and Tiwari et al. [20] confirmed the effectiveness of XGBoost, though Feng et al. observed a risk of overfitting when Grid Search was used for parameter tuning on a dataset of 271 samples. Gene Expression Programming (GEP) and genetic algorithms (GAs) represent other techniques applied to formulate empirical expressions. Comparative studies examining experimental results suggest that ML models developed under appropriate conditions generally provide higher prediction accuracy than traditional code provisions, including ACI 318 and Eurocode 2, as well as existing mechanistic models.

Despite the progress made in applying ML techniques to predict the shear strength of RCDB, the datasets used in many of these studies are often limited in scope and size, which can restrict the generalizability of the developed models. To address these gaps and enhance the generalizability of shear strength prediction for RCDB, the present study proposes the use of advanced ML algorithms, specifically CatBoost and XGBoost, with their hyperparameters optimized through the Grid Search technique. Subsequent to determining appropriate hyperparameter sets for the individual models, Stacking and Voting ensemble techniques were applied to combine the selected CatBoost and XGBoost models. This procedure yielded four distinct ML models for predicting the shear strength of RCDB, comprising the individual CatBoost and XGBoost models along with their corresponding Stacking and Voting ensemble versions. The dataset employed for training and evaluating these ML models consists of 840 experimental samples, compiled from reputable published studies, and includes both RCDB with web reinforcement and those without, ensuring a comprehensive representation of different reinforcement configurations.

This paper proceeds with the following organization. Section 2 details the dataset, including statistical properties, distribution analysis, and correlations between input and output variables. Section 3 outlines the methodology applied. In Section 4, the results of the ML models are presented and discussed, covering predictive performance and an

examination of the influence of input parameters on the shear strength of RCDB. Section 5 concludes the paper by summarizing the principal findings and suggesting areas for subsequent research.

2. Database description and analysis

2.1. Design standards and shear mechanisms in RCDB

The shear strength of RCDB is traditionally assessed using design standards that model load transfer through the strut-and-tie mechanism, tailored to their unique behavior with a shear span-to-depth ratio (a/d) typically below 2.5. ACI 318-19 employs the strut-and-tie model (STM), calculating nominal shear strength as [14]:

$$SS_{ACI}=0.85\beta_s f_c\, bw_s\sin\theta,\quad w_s=[1.8w_t\cos\theta+(w_{tp}+w_{bp})\sin\theta]/2 \tag{1}$$

Eurocode 2 adopts a variable-angle STM, with design shear strength [14]:

$$SS_{EU}=0.85\beta_s f_c\, bw_s\sin\theta,\quad w_s=[1.85w_t\cos\theta+(w_{tp}+w_{bp})\sin\theta]/2 \tag{2}$$

where: β_s is coefficient of strut, θ is the angle between the strut and the longitudinal axis, w_s and w_t are the widths of the strut and tie.

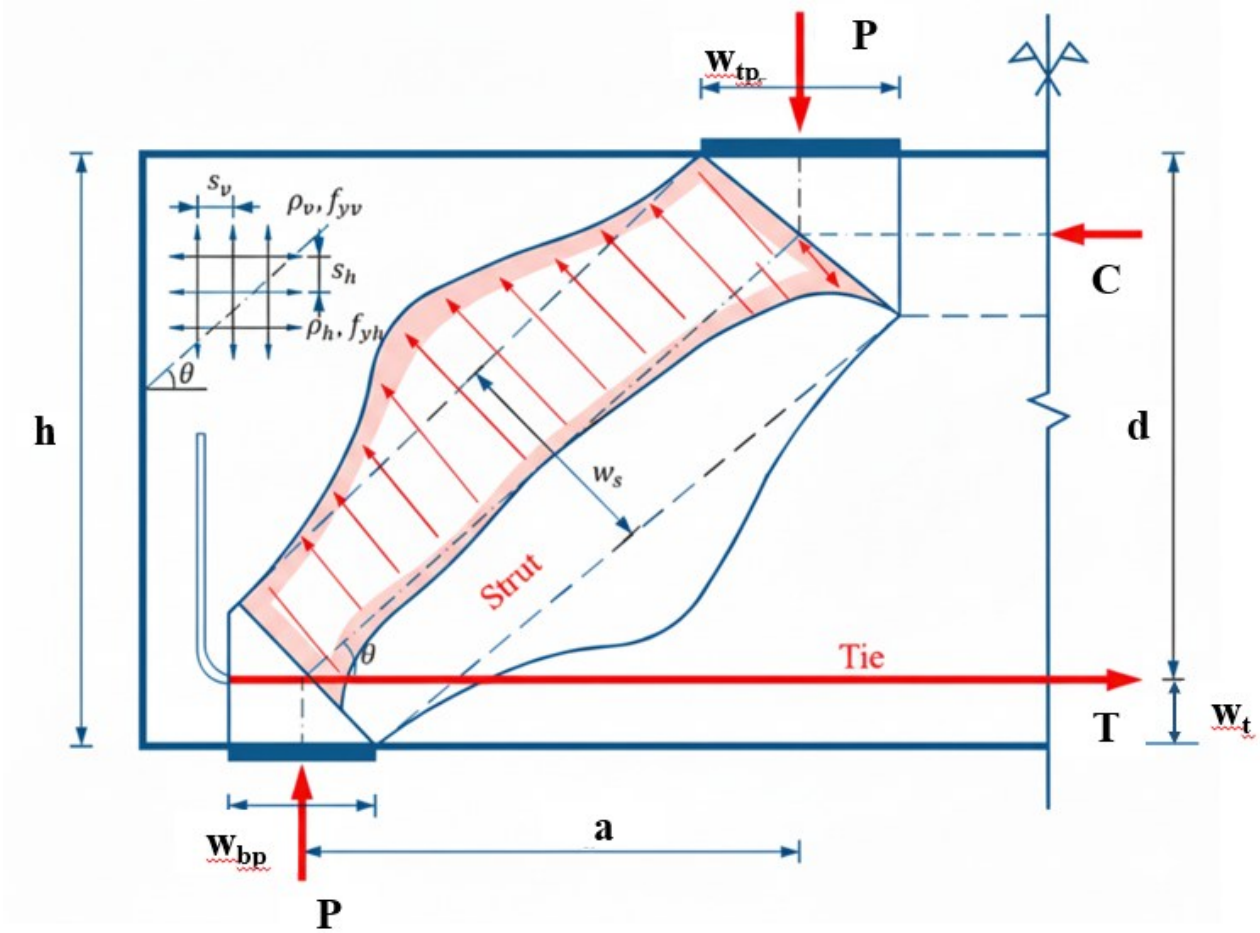

Fig. 1. Schematic diagram of the shear mechanism of RCDB

Fig. 1 depicts the shear mechanism in RCDB, which includes compressive struts and tension ties.

2.2. Dataset Description

The experimental dataset employed for developing and evaluating ML models for predicting the shear strength of RCDB is detailed in this section. The origin, scope, and essential characteristics of the data are outlined, along with the statistical distributions of the relevant parameters. This description provides a clear basis for the modeling process that follows. For the purpose of developing reliable models, a thorough dataset was compiled. This process involved gathering information from previously published, peer-reviewed research, specifically the collection

organized by Chetchotisak et al., to ensure the quality and dependability of the data [21]. The dataset assembled for this study contains 840 experimental results from RCDB tests. Among these samples, 322 are from beams without web reinforcement and 518 are from beams with web reinforcement, encompassing a range of configurations. Based on findings from various experimental and theoretical studies [12, 22, 23], the shear strength of RCDB is affected by multiple components, including the strength of concrete material, longitudinal reinforcement, and web reinforcement. Consequently, 12 features are designated as input variables, categorized into five groups: (i) geometric dimensions, including beam height (h), beam width (b), and shear span-to-depth ratio (a/d); (ii) concrete property, represented by concrete strength (f_c); (iii) bottom longitudinal reinforcement properties, comprising bottom reinforcement ratio (ρ) and strength (f_y); (iv) web reinforcement properties, encompassing vertical web reinforcement ratio (ρ_v) and strength (f_{yv}), as well as horizontal web reinforcement ratio (ρ_h) and strength (f_{yh}); and (v) top plate width (w_{tp}) and bottom plate width (w_{bp}). The corresponding output variable is established as the shear strength, derived from experimental data. Fig. 2 shows a schematic of an RCDB with its basic dimensions.

Table 1 provides a summary of the main input and output parameters utilized for the development of the ML models in this investigation. Descriptive statistics, including the minimum, median, maximum, average, and standard deviation for each variable within the 840-sample dataset, are presented. The main goal of this table is to thoroughly describe the dataset, offering information on the range, central tendency, and variability of the geometric and material properties that influence the shear strength (SS) of the RCDB under analysis. The table indicates substantial variability across the majority of parameters. Notably, the target output, Shear Strength (SS), exhibits a broad range from 20.70 kN to 6294.00 kN, accompanied by a large standard deviation (667.92 kN), suggesting varied structural capacities within the dataset. Key input parameters such as beam height (h: 152.00 - 2100.00 mm), beam width (b: 51.00 - 914.00 mm), and concrete compressive strength (f_c: 11.30 - 120.10 MPa) also display considerable ranges. The shear span-to-depth ratio (a/d) ranges from 0.27 to 2.50, confirming that the dataset primarily consists of beams classified as deep beams. Importantly, the statistics for vertical and horizontal web reinforcement ratios and strengths show minimum values of 0.00, verifying the inclusion of samples both with and without web reinforcement, as previously mentioned. The extensive statistical ranges observed emphasize the heterogeneity of the dataset and reflect the complex, multi-variable nature of shear behavior in RCDB. This diversity is crucial for training resilient ML models capable of generalizing effectively across various design scenarios. The considerable dispersion observed in shear strength values indicates the difficulty associated with accurately predicting this parameter using conventional methods. The incorporation of beams with differing reinforcement configurations (indicated by the web reinforcement statistics) allows the models to potentially learn the specific contributions of different reinforcement types to shear capacity.

Fig. 3 displays a series of box plots, providing a visual representation of the statistical distribution for each parameter. These plots serve to illustrate graphically each variable's central tendency (median), dispersion (interquartile range), overall range, and the existence of asymmetry and potential extreme values within the dataset. The box plots indicate varied distribution patterns across different parameters. Several essential variables, including beam height, beam width, concrete strength, bottom reinforcement strength, top and bottom plate widths, and notably the output variable shear strength, show distributions with positive skewness. This is indicated by the median

(red line) being closer to the lower boundary of the box (first quartile) than the upper boundary, and the presence of numerous data points extending considerably beyond the upper whisker (extreme values), particularly for shear strength. The shear span-to-depth ratio (a/d) exhibits a more symmetrical distribution around its median, although it spans the entire range up to 2.5. Parameters related to web reinforcement (vertical and horizontal ratios and strengths) show distributions largely concentrated at or near zero, with many data points clustered at the minimum value, visually confirming that a substantial portion of the dataset includes beams without such reinforcement. The difference between the mean and median in many plots further quantifies the observed asymmetry. The observed asymmetry and the prevalence of extreme values, especially in the target Shear Strength variable, suggest the non-normal nature of the data. This implies that predictive models need to be capable of accommodating such distributions and potential extreme values effectively. The distributions heavily weighted towards zero for web reinforcement parameters visually emphasize the dataset's inclusion of beams both with and without web reinforcement, which is vital for the models to learn the influence of this design choice. Overall, the plots highlight the inherent variability and complexity within the collected experimental data.

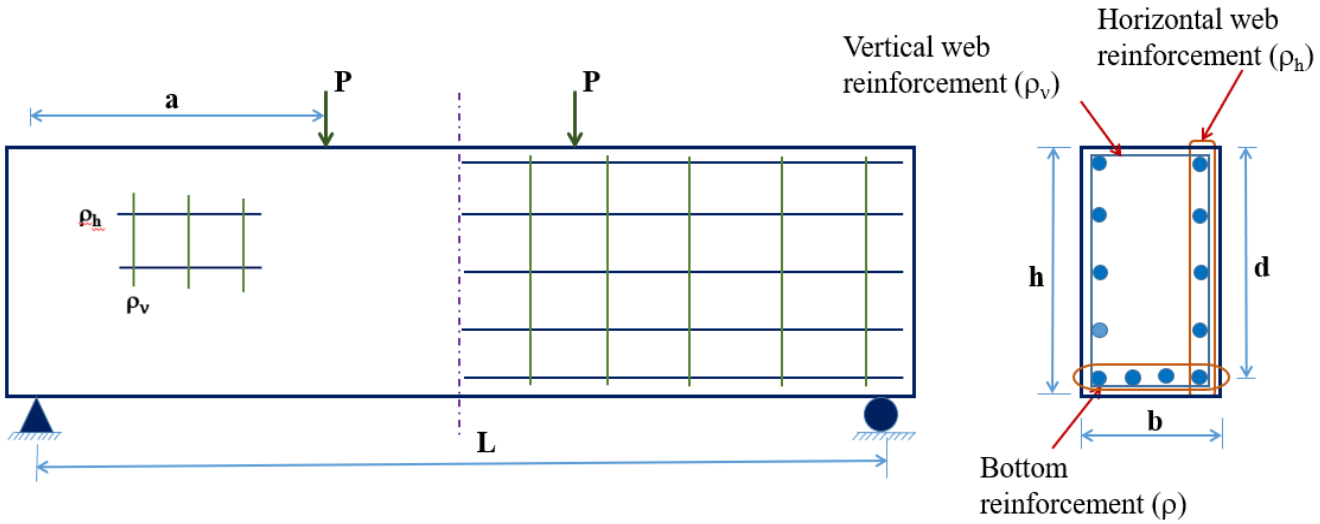

Fig. 2. Schematic illustration of RCDB

Table 1. Statistical summary of parameters used

Parameters	Symbol	Role	Min	Median	Max	Average	Std
Beam height	h (mm)	Input	152.000	460.000	2100.000	563.580	290.572
Beam width	b (mm)	Input	51.000	158.500	914.000	195.538	120.004
Shear span-to-depth ratio	a/d	Input	0.270	1.207	2.502	1.304	0.541
Concrete strength	f_c (MPa)	Input	11.300	32.000	120.100	40.178	21.594
Bottom reinforcement ratio	ρ	Input	0.003	0.019	0.113	0.020	0.011
Bottom reinforcement strength	f_{yl} (MPa)	Input	267.000	441.000	1330.000	470.268	130.563
Vertical web reinforcement ratio	ρ_v	Input	0.000	0.002	0.029	0.003	0.004

Table 1. (continued)

Parameters	Symbol	Role	Min	Median	Max	Average	Std
Vertical web reinforcement strength	f_{yv} (MPa)	Input	0.000	368.000	1051.000	272.677	226.894
Horizontal web reinforcement ratio	ρ_h	Input	0.000	0.000	0.032	0.002	0.003
Horizontal web reinforcement strength	f_{yh} (MPa)	Input	0.000	0.000	855.000	206.106	229.689
Top plate width	w_{tp} (mm)	Input	10.000	102.000	914.000	146.155	106.566
Bottom plate width	w_{bp} (mm)	Input	10.000	102.000	610.000	135.852	82.342
Shear strength	SS (kN)	Output	20.700	347.500	6294.000	562.009	667.918

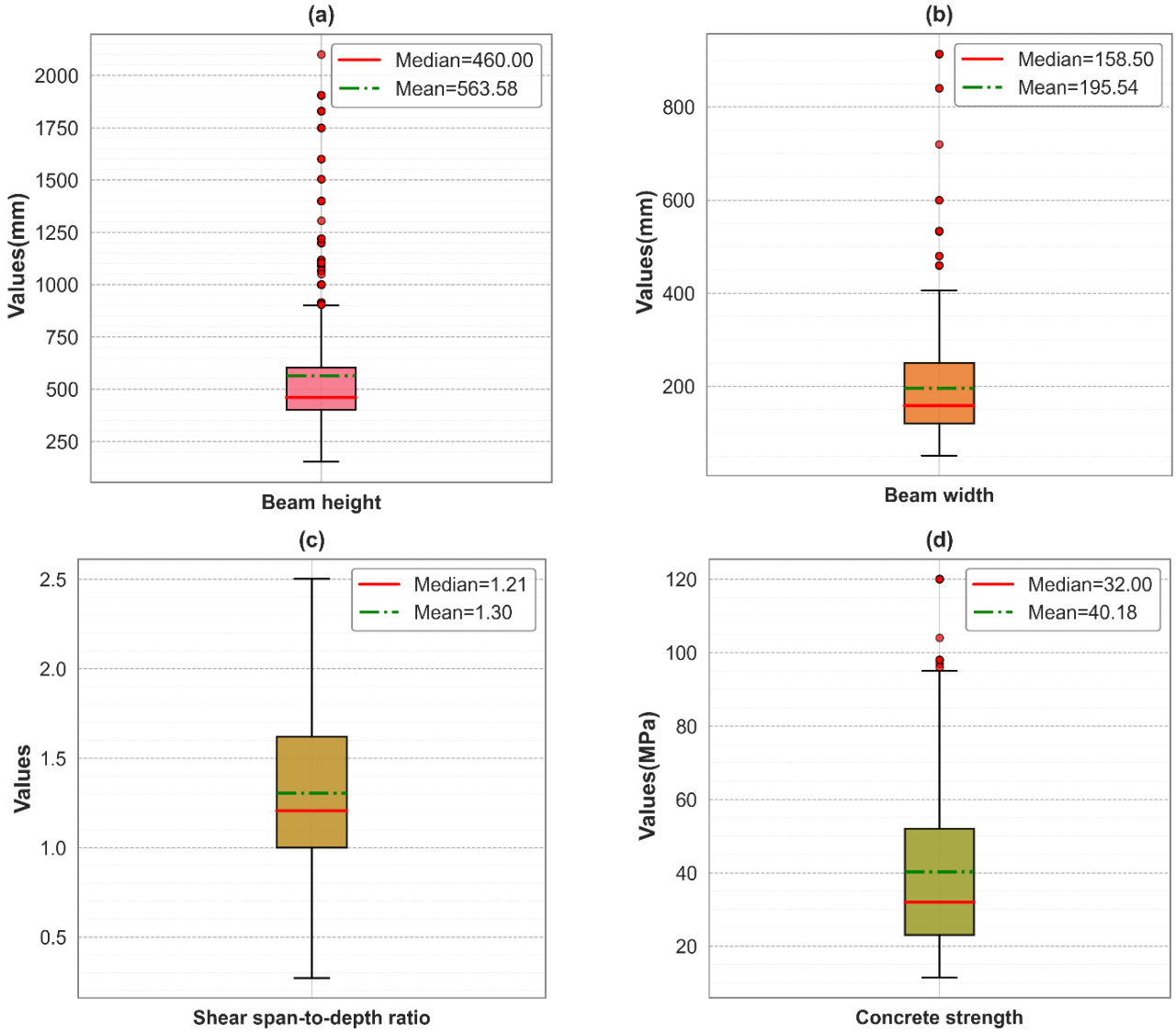

Fig. 3. Distribution plots (box plots) for input and output parameters showing median, mean, interquartile range, and outliers: (a) Beam height, (b) Beam width, (c) Shear span-to-depth ratio, (d) Concrete strength, (e) Bottom reinforcement ratio, (f) Bottom reinforcement strength, (g) Vertical web reinforcement ratio, (h) Vertical web reinforcement strength, (i) Horizontal web reinforcement ratio, (j) Horizontal web reinforcement strength, (k) Top plate width, (l) Bottom plate width, (m) Shear strength.

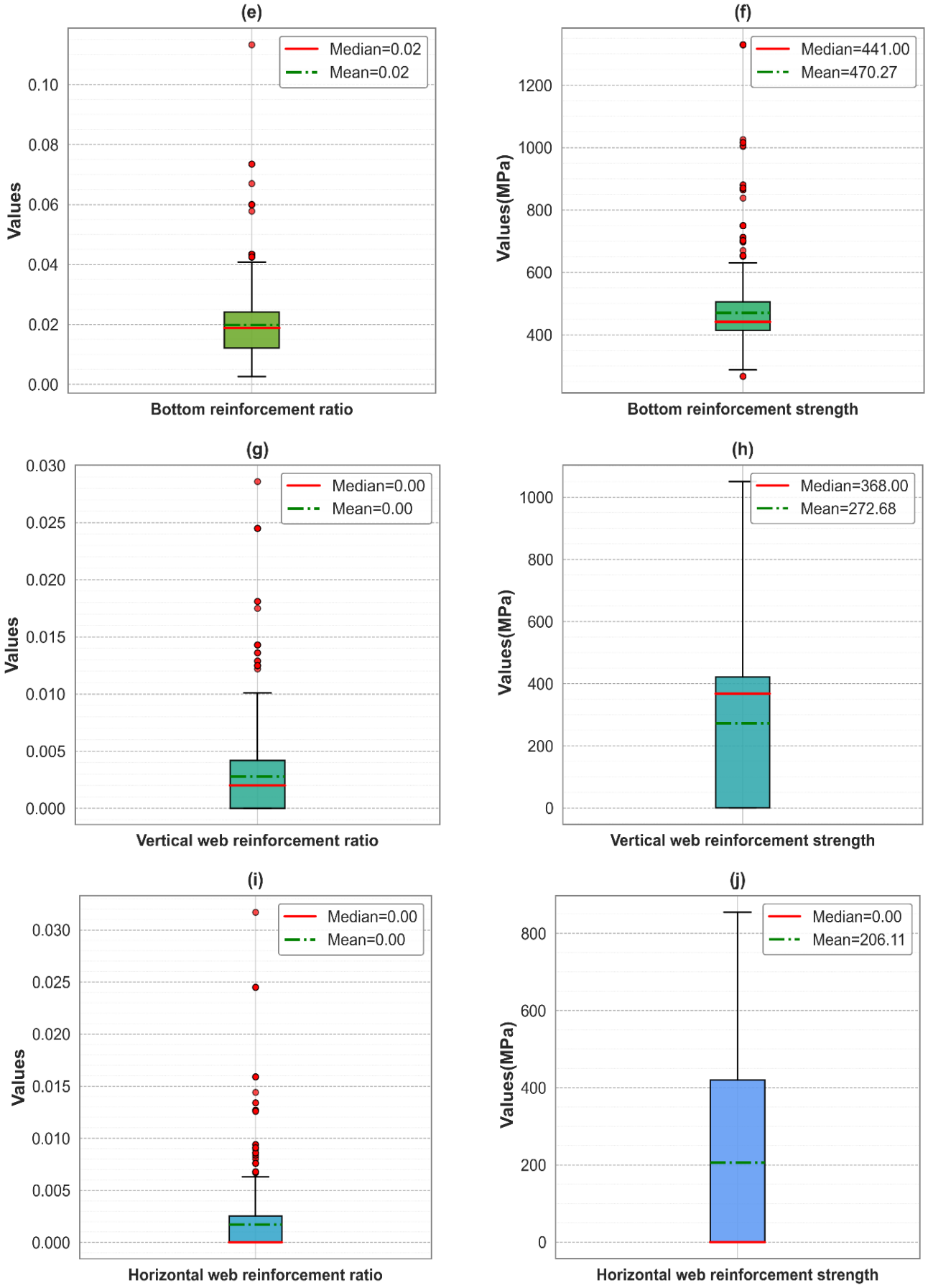

Fig. 3. (continued)

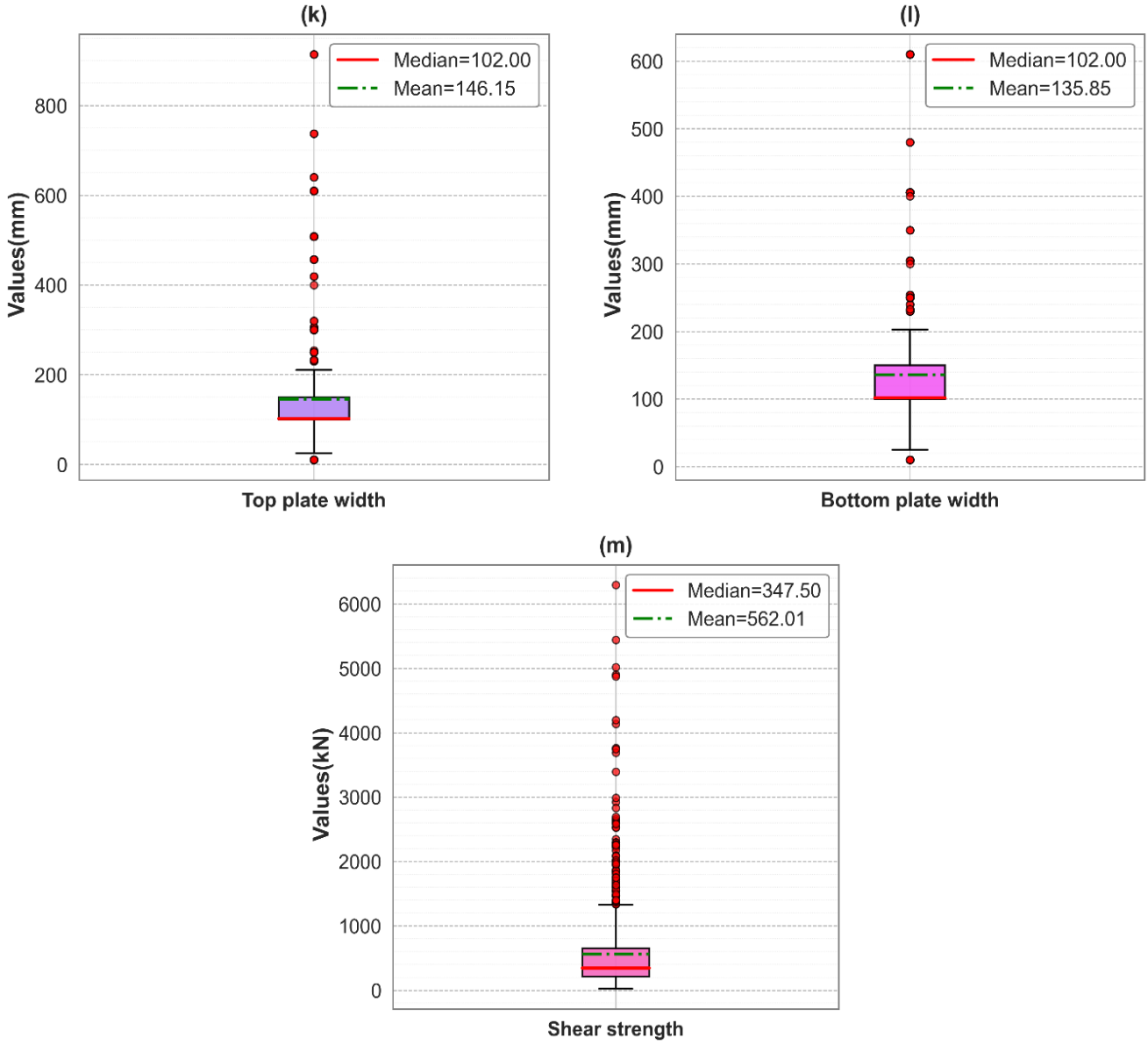

Fig. 3. (continued)

Fig. 4 displays the Pearson correlation matrix for the dataset, providing a graphical representation of the linear relationships between all pairs of input variables and the output variable (shear strength, denoted as 'V' in the figure). This matrix employs color intensity and circle size to indicate the correlation coefficient, with numerical values presented for accuracy. The aim of this analysis is to identify potential linear dependencies between variables, determine which input parameters exhibit the strongest linear association with shear strength, and detect possible multicollinearity among predictor variables.

The correlation matrix indicates several noteworthy linear relationships within the dataset. Strong positive linear correlations are observed between the shear strength ('V') and key geometric parameters: beam height ('h', r=0.68), beam width ('b', r=0.76), top plate width ('w_{tp}', r=0.72), and bottom plate width ('w_{bp}', r=0.69). Significant positive linear correlations also exist among these geometric input variables themselves, such as between 'h' and 'b' (r=0.51), 'h' and 'w_{tp}' (r=0.65), 'h' and 'w_{bp}' (r=0.74), and particularly between 'w_{tp}' and 'w_{bp}' (r=0.79). Concrete strength ('f_c') shows only a weak positive linear correlation with shear strength (r=0.11). The shear span-to-depth ratio ('a/d') exhibits a negligible linear correlation with

shear strength (r=-0.04). The parameters related to vertical and horizontal web reinforcement (ratios and strengths) generally display weak linear correlations with shear strength individually, although moderate correlations exist between some reinforcement parameters (e.g., 'f_{yv}' and 'f_{yh}', r=0.58). The strong positive linear correlations between beam dimensions and shear strength are consistent with fundamental structural engineering principles, where larger beams are generally expected to exhibit higher shear capacities. However, the relatively weak linear correlations observed for concrete strength, reinforcement properties, and especially the shear span-to-depth ratio suggest that their influence on shear strength is likely complex and non-linear. These factors are known to interact in intricate ways, affecting mechanisms such as arch action, which are not adequately represented by simple linear correlation coefficients.The moderate to strong correlations between certain input parameters (e.g., beam height and width, plate widths) indicate potential multicollinearity, which could be a consideration for some modeling techniques but is often managed effectively by the tree-based ensemble methods (XGBoost, CatBoost) employed later in this study. While the strong correlation of dimensions with strength is anticipated, the low linear correlation value for the a/d ratio is particularly notable. This ratio is widely recognized as a critical factor distinguishing deep beam behavior from slender beams, primarily through its influence on shear transfer mechanisms. The lack of a strong linear correlation here supports statements made in the introduction regarding the highly non-linear behavior of deep beams and the limitations of simpler models. Similarly, while concrete strength is undoubtedly important, its contribution might vary non-linearly depending on other factors like reinforcement and geometry.

In summary, Fig. 4 emphasizes the substantial linear contribution of beam geometry to the prediction of shear strength, while also demonstrating that linear correlation alone is insufficient to fully represent the influence of other crucial factors such as the a/d ratio and material properties. This observation supports the need for utilizing advanced, non-linear ML models that can learn intricate relationships directly from the data. With the characteristics and linear relationships within the dataset now examined, the study logically proceeds to the development and evaluation of such ML models in the following sections.

3. Methodology

3.1. Development of CatBoost and XGBoost Models

The prediction of shear strength in RCDB was undertaken using two gradient boosting algorithms: CatBoost and XGBoost.

CatBoost, originating from Yandex [24], is a gradient boosting algorithm constructed particularly for processing categorical features effectively and reducing the tendency for overfitting. One of its key innovations is ordered boosting, a permutation-based technique that addresses the issue of target leakage inherent in traditional gradient boosting methods. In ordered boosting, CatBoost trains the model on subsets of the data in a specific order, ensuring that the target value of the current sample is not used during its evaluation. This approach minimizes bias in the gradient calculations. By reducing overfitting, ordered boosting enhances the model's generalization ability, making it well-suited for predicting shear strength from diverse input parameters. Additionally, CatBoost employs oblivious decision trees (also referred to as symmetric trees), where the same splitting criterion is applied across all nodes at a given tree level. This structural constraint acts as a form of regularization, simplifying the model and reducing the risk of overfitting. Furthermore, the symmetry in tree structure accelerates computation by enabling efficient parallelization and reducing memory

usage, which is critical for processing large datasets or complex feature sets in RCDB modeling.

XGBoost [25], recognized as an optimized gradient boosting framework, is often chosen for its scalability and high performance in predictive modeling. A core strength of XGBoost lies in its objective function, which incorporates L1 (alpha) and L2 (lambda) regularization terms to penalize model complexity by constraining the magnitude of leaf weights. This regularization mechanism is particularly effective in preventing overfitting, ensuring that the model generalizes well to unseen data. Another key feature is XGBoost's use of a second-order Taylor expansion of the loss function, which includes both first-order (gradient) and second-order (Hessian) information. This approach provides a more accurate approximation of the loss landscape, enabling faster convergence and improved optimization compared to traditional gradient boosting methods that rely solely on first-order gradients. Additionally, XGBoost supports parallel processing and tree pruning, which enhance computational efficiency, making it suitable for iterative model training and hyperparameter tuning in this study. Analogous to CatBoost, XGBoost was utilized in this research to predict the shear strength of RCDB from the same input parameter set.

For the initial training of both models, the dataset comprising 840 experimental samples was divided into training and testing subsets using a 70:30 allocation ratio. This partitioning provides the majority of the data for model development, while reserving a separate, independent set for the subsequent evaluation of predictive capabilities.

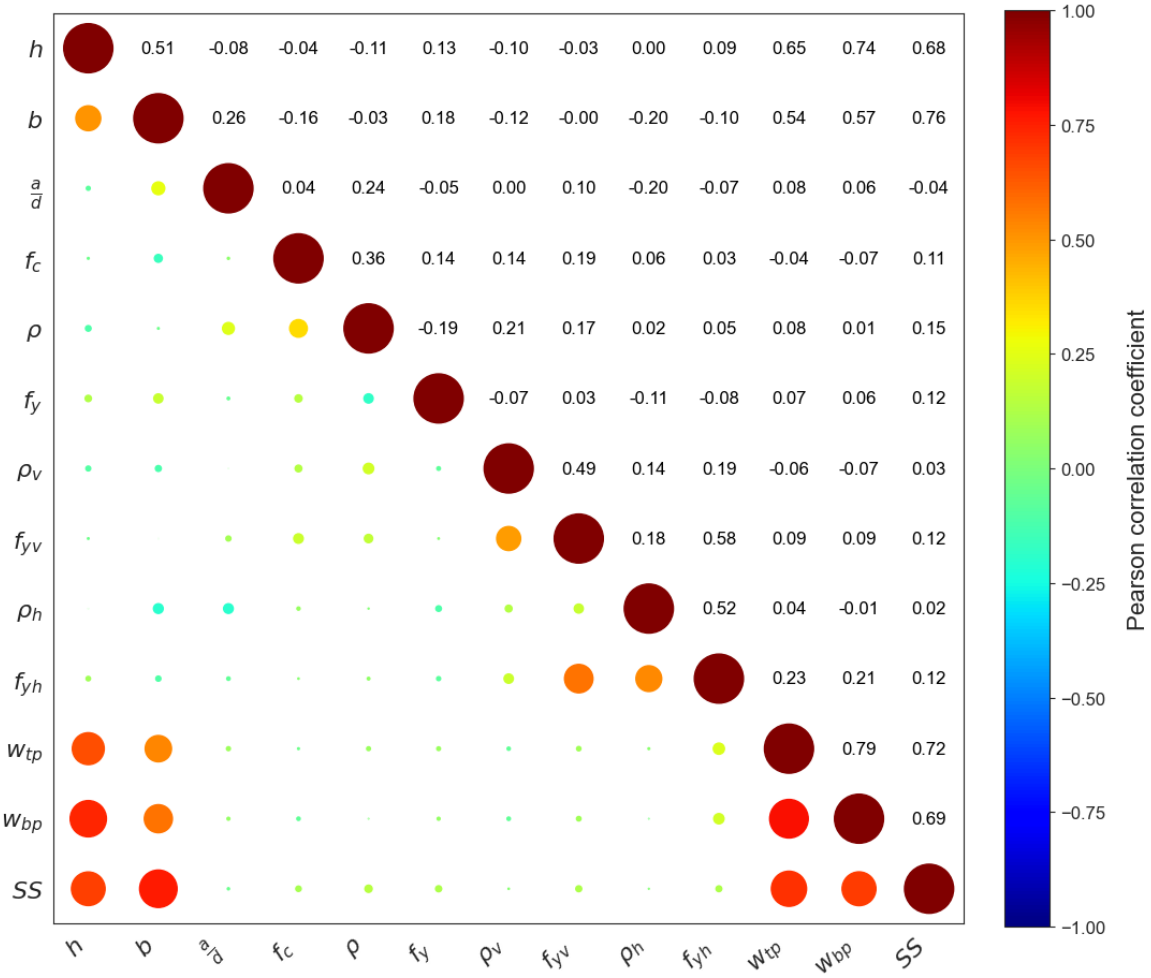

Fig. 4. Pearson correlation matrix illustrating linear relationships between input parameters and shear strength (SS)

3.2. Hyperparameter Optimization Using Grid Search

Hyperparameter adjustment for the CatBoost and XGBoost models was conducted using the Grid Search technique aiming to enhance model performance. Grid Search examines a defined grid of hyperparameter combinations, where models for each combination are trained and validated on the training set through k-fold cross-validation. In this study, 5-fold cross-validation was utilized during this process to inform the evaluation and lessen the possibility of overfitting.

For the CatBoost model, the hyperparameters subject to adjustment included learning rate, tree depth, number of iterations, and the L2 regularization parameter. The specified search ranges were: learning rate (0.01 to 0.2), tree depth (3 to 6), number of iterations (10 to 500), and L2 regularization parameter (1 to 10). Correspondingly for XGBoost, the hyperparameters adjusted were learning rate, maximum tree depth, number of estimators, and reg_lambda, using the following ranges: learning rate (0.01 to 0.2), maximum depth (3 to 6), number of estimators (10 to 500), and reg_lambda (1 to 10).

The Grid Search process determines the hyperparameter combination associated with the minimum root mean square error (RMSE) observed during cross-validation for each algorithm. The CatBoost and XGBoost models are then retrained on the entire training dataset using these identified hyperparameter configurations before proceeding to predict the shear strength of RCDB.

3.3. Implementation of Stacking and Voting Ensemble Techniques

To further improve the predictive performance and robustness of the models, the Stacking and Voting ensemble techniques are applied to combine the optimized CatBoost and XGBoost models. These ensemble methods aim to capitalize on the strengths of both models while mitigating their individual weaknesses.

In the Stacking ensemble approach, the predictions from the optimized CatBoost and XGBoost models are used as input features for a meta-learner. The Stacking process involves two levels: in the first level, the CatBoost and XGBoost models are trained on the training set, and their predictions on the validation set (obtained through cross-validation) are collected. In the second level, the meta-learner is trained on these predictions to produce the final shear strength prediction. This approach allows the Stacking ensemble to learn how to best combine the predictions of the base models, potentially improving overall accuracy. Stacking is distinguished by its meta-learner flexibility, enabling the optimal integration of diverse base model outputs, which enhances predictive precision for the complex shear behavior in RC deep beams. However, this flexibility comes with increased computational complexity due to the additional meta-learning step, and a higher risk of overfitting may arise, potentially impacting performance stability.

The Voting ensemble technique aggregates predictions from the previously tuned CatBoost and XGBoost models via averaging of their respective outputs. Such averaging can reduce overall prediction variance by combining information from the constituent base models, leading to a more robust and generalizable final model. The primary strength of Voting lies in its simplicity and effectiveness in smoothing individual prediction errors, contributing to enhanced stability against noise in the experimental data. Nevertheless, this method also incurs increased computational complexity due to the need for multiple model evaluations. Despite this limitation, the performance benefits—such as a reduced root mean square error (RMSE) - are considered to outweigh the computational cost, justifying their use in capturing the non-linear relationships within the 840-sample dataset.

This use of ensemble methods results in two additional models: the Stacking ensemble and the

Voting ensemble. Therefore, four ML models were ultimately considered: the individual CatBoost and XGBoost models, plus the Stacking and Voting ensembles derived from them. The performance of these models in predicting the shear strength of RCDB was subsequently evaluated using the testing set. This evaluation employed several statistical metrics: the coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and the A20 index. These selected metrics permit an appraisal of model accuracy concerning both goodness-of-fit and prediction error magnitude. Calculation formulas for these metrics were obtained from reference [26].

4. Results and Discussion

4.1. Hypeparameter tuning

The main objective of this subsection is to determine and establish the optimal set of hyperparameters for the individual ML models utilized in this investigation, specifically the CatBoost Regressor (CBR) and XGBoost Regressor (XGBR). The careful adjustment of these parameters is essential for maximizing the predictive accuracy and generalization capacity of each model when applied to the task of estimating the shear strength of RCDB using the dataset previously described. This optimization step ensures that the individual models perform optimally before their evaluation or potential integration into ensemble predictors. To achieve the best possible model configuration, a systematic hyperparameter tuning process was conducted for both the CBR and XGBR models. This involved defining a search space, shown in Table 2, that included key hyperparameters known to affect the performance of these gradient boosting algorithms (e.g., tree depth, learning rate, number of iterations/estimators, regularization terms). A search strategy, combined with cross-validation performed on the training dataset, was employed to examine different combinations of these hyperparameters. The performance of each combination was assessed using the RMSE metric, with the goal of identifying the parameter set that resulted in the lowest average RMSE during cross-validation.

Table 3 presents the specific results of the hyperparameter optimization procedure. It explicitly lists the hyperparameter sets that performed best for the CBR and XGBR models, based on the minimization of cross-validation error. The purpose of this table is to document the final, optimized configurations employed for these base models in subsequent analyses and comparisons. The table indicates that for the CBR model, the optimal configuration identified through the search process included a tree depth of 3, 500 iterations, an l2_leaf_reg (L2 regularization) value of 1, and a learning_rate of 0.09. This set of parameters resulted in a CV_Score_RMSE of approximately 196.96 kN. For the XGBR model, the optimal parameters were found to be a learning_rate of 0.1, a max_depth of 3, 500 n_estimators, and a reg_lambda (L2 regularization) value of 1, achieving a CV_Score_RMSE of about 242.62 kN. These results specify the particular model architectures and learning settings that yielded the best predictive performance during the cross-validation phase for each algorithm on this specific dataset. The selection of a relatively shallow max_depth (i.e., 3) for both models often suggests an approach to manage model complexity and potentially reduce overfitting. The achieved CV_Score_RMSE values serve as estimates of the models' generalization error; based on these scores, the tuned CBR model exhibited a slightly lower cross-validation error compared to the tuned XGBR model (196.96 kN vs 242.62 kN). These specific parameters fine-tune the learning process of each algorithm from the deep beam data.

Fig. 5 provides a visual summary of key performance results from the hyperparameter tuning phase for the CBR and XGBR models. Specifically, it illustrates the RMSE achieved on the training data, during cross-validation, and on the

test data. The bar chart clearly indicates that for both CBR and XGBR models, the RMSE on the training data is notably lower than the RMSE obtained during cross-validation and on the test data. When comparing the two models, the cross-validation RMSE for CBR appears lower than that for XGBR. The test set RMSE seems to be higher than the cross-validation RMSE for both models. Error bars are included on the cross-validation bars, representing the standard deviation or variability of the RMSE scores across the different cross-validation folds. Overall, Fig. 5 visually presents the RMSE performance attained by the CBR and XGBR models during and following the hyperparameter tuning phase across training, cross-validation, and test datasets. It emphasizes the difference in performance between training and validation/testing and provides a visual confirmation of the cross-validation results reported.

Table 2. Hyperparameter ranges explored during Grid Search optimization for CatBoost and XGBoost models

Hyperparameters	Value range
'learning_rate'	0.01, 0.03, 0.05, 0.07, 0.09, 0.1, 0.2
'iterations'	10, 20, 30, 40, 50, 60, 70 ,80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500
'n_estimators'	10, 20, 30, 40, 50, 60, 70 ,80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500
'reg_lambda'	1, 5, 10
'l2_leaf_reg'	1, 5, 10
'depth'	3, 4, 5, 6,
'max_depth'	3, 4, 5, 6

Table 3. Optimal hyperparameters and corresponding cross-validation RMSE scores identified for CatBoost (CBR) and XGBoost (XGBR) models

Model	Best_Params	CV_Score_RMSE
CBR	'depth': 3, 'iterations': 500, 'l2_leaf_reg': 1, 'learning_rate': 0.09	196.96
XGBR	'learning_rate': 0.1, 'max_depth': 3, ' n_estimators': 500, 'reg_lambda': 1	242.62

4.2. Assessment of Ensemble models

Fig. 6 offers a comprehensive visual comparison of the predictive performance of the four developed models: XGBR, CBR, Stacking ensemble (XGBR + CBR), and Voting ensemble (XGBR + CBR). It comprises five subfigures (a-e), each dedicated to illustrating a specific performance metric across the training, testing, and overall datasets. The combined purpose of these figures is to provide a clear graphical assessment of the models' accuracy, error magnitudes, relative errors, and reliability from multiple perspectives.

For all models and most metrics, performance on the training data is notably better (higher R^2/A20, lower errors) than on the test data, indicating a degree of overfitting, which is typical in ML. The Voting and CatBoost models exhibit the highest R^2 values on the test set (around 0.96), followed by XGBoost (0.94) and Stacking (0.92). In terms of MAE and RMSE, the bars representing test MAE and RMSE likely indicate that the Voting

ensemble (MAE ≈ 71.7 kN, RMSE ≈ 131.8 kN) and CatBoost (MAE ≈ 72.6 kN, RMSE ≈ 134.1 kN) yield smaller average prediction errors compared to XGBoost (MAE ≈ 78.6 kN, RMSE ≈ 160.5 kN) and Stacking (MAE ≈ 91.6 kN, RMSE ≈ 185.8 kN). Regarding MAPE, the test MAPE bars would likely show the Voting ensemble performing best (lowest relative error, ≈ 14.1%), followed closely by XGBoost (≈ 14.8%) and CatBoost (≈ 15.0%), with the Stacking model exhibiting the highest MAPE (≈ 18.8%). Concerning A20, the test A20 bars, representing the percentage of predictions within 20% of the actual value, would likely be highest for the Voting ensemble (≈ 78.2%), followed by XGBoost (≈ 77.4%) and CatBoost (≈ 76.2%), with Stacking being the lowest (≈ 67.5%).

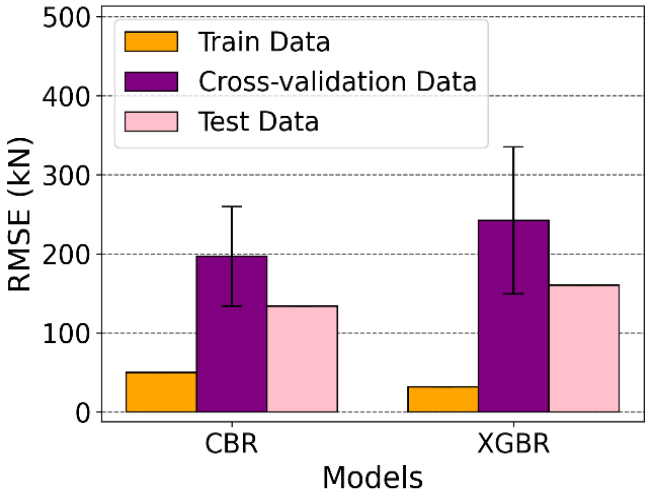

Fig. 5. Comparison of Root Mean Squared Error (RMSE) on training, cross-validation, and test datasets for the tuned CatBoost (CBR) and XGBoost (XGBR) models

The Voting and CatBoost models generally appear visually superior or highly competitive across most metrics on the test set, while the Stacking model consistently shows slightly lower performance. The collective evidence from Fig. 6 suggests that all developed models demonstrate good predictive capability for shear strength, as indicated by high R^2 and A20 values and manageable error metrics on the test set. The consistent performance difference between training and testing emphasizes the importance of evaluating models on unseen data. Comparing the models, the Voting ensemble generally exhibits the most favorable performance profile across the different metrics on the test set, suggesting that averaging predictions from XGBoost and CatBoost produces robust and accurate results. CatBoost also performs very strongly as an individual model. Recognition exists that the margin of improvement of the Voting ensemble over standalone CatBoost or XGBoost models appears modest in terms of metrics like R^2 (e.g., 0.961 vs. 0.959 for CatBoost). However, the primary advantage of the Voting ensemble lies in enhanced robustness and reduced prediction variance. Averaging outputs from two optimized models mitigates random errors and individual weaknesses, yielding a more stable predictor less prone to large errors on unseen data. The Stacking model, while still performing reasonably well, appears slightly less effective in generalization compared to the others based on these visualisations.

Fig. 6 provides a direct visual counterpart to the numerical data presented comprehensively in Table 4. Observing the performance across multiple metrics simultaneously allows for a more nuanced assessment than relying on a single metric. For instance, while XGBoost has slightly higher absolute errors (MAE/RMSE) than CatBoost on the test set, its relative error (MAPE) and A20 score are competitive, which might be relevant depending on the application's specific

requirements. The visual format makes these trade-offs and comparisons readily apparent.

The Voting model, which combines predictions from the optimized CatBoost and XGBoost models, consistently yielded the most favorable outcomes across the majority of key performance metrics on the critical test dataset. Consequently, considering the comprehensive results from both the tuning and final evaluation phases, the Voting ensemble model appears to be the most resilient and accurate predictor among the models examined in this investigation.

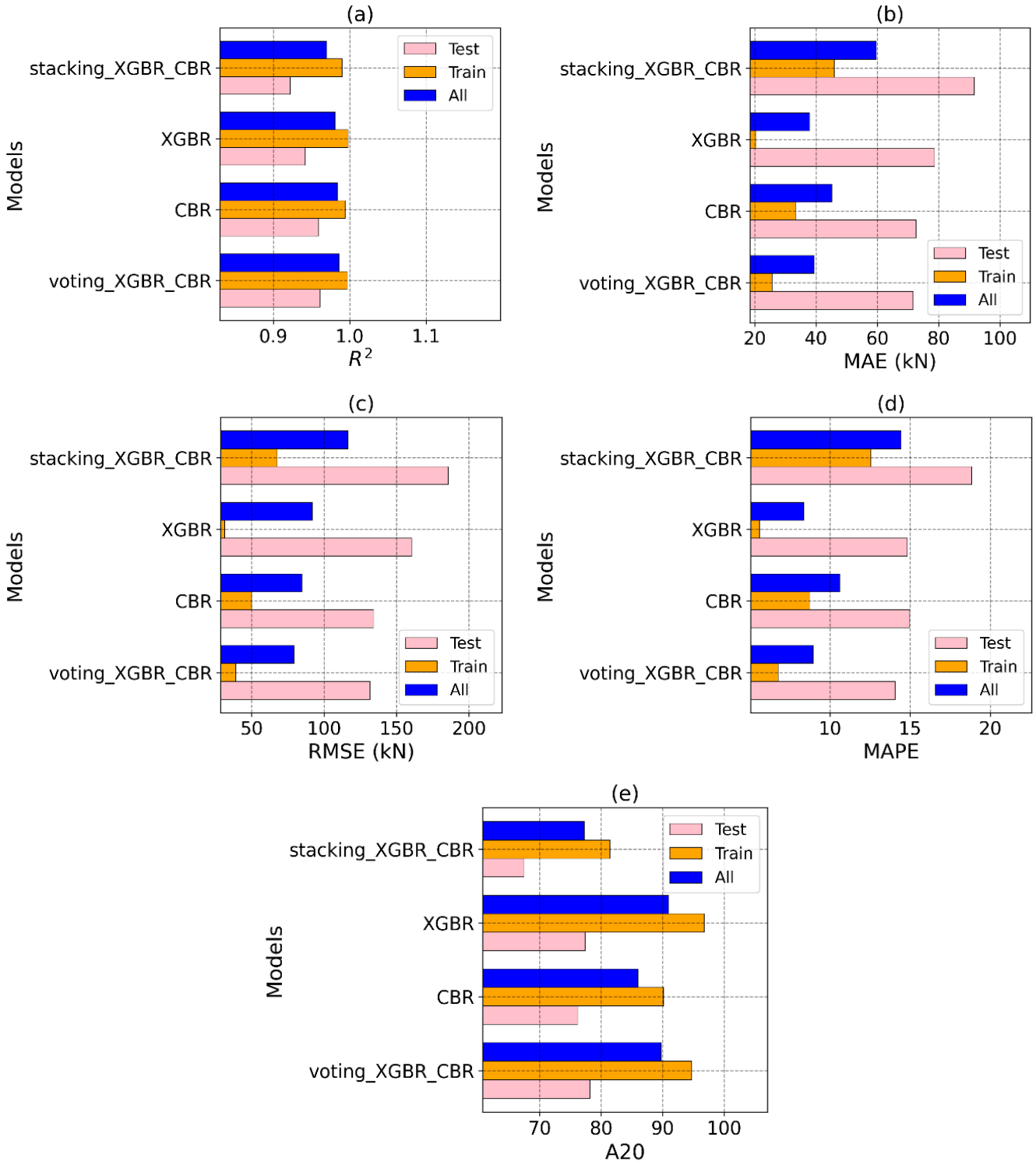

Fig. 6. Visual comparison of model performance metrics across training, testing, and all datasets for XGBoost, CatBoost, Stacking, and Voting models: (a) R², (b) MAE, (c) RMSE, (d) MAPE, (e) A20 index

Table 4. Performance metrics (MAE, RMSE, R^2, MAPE, A20) of individual (XGBoost, CatBoost) and ensemble (Stacking, Voting) models on training, testing, and all datasets

	Metrics \ Models	XGBoost	CatBoost	Stacking (XG+Cat)	Voting (XG+Cat)
Train	MAE (kN)	20.465	33.364	46.034	25.635
	RMSE (kN)	31.457	50.097	67.620	39.013
	R^2	0.998	0.994	0.990	0.997
	MAPE	5.604	8.743	12.559	6.768
	A20	96.769	90.136	81.463	94.728
Test	MAE (kN)	78.615	72.648	91.571	71.673
	RMSE (kN)	160.501	134.093	185.802	131.775
	R^2	0.942	0.959	0.922	0.961
	MAPE	14.817	14.97	18.827	14.081
	A20	77.381	76.19	67.46	78.175
All	MAE (kN)	37.910	45.149	59.695	39.446
	RMSE (kN)	91.765	84.564	116.437	79.213
	R^2	0.981	0.984	0.970	0.986
	MAPE	8.368	10.611	14.439	8.962
	A20	90.952	85.952	77.262	89.762

4.3. Representative prediction results

The main purpose of this section is to illustrate visually the predictive capability and accuracy of the developed ML models, with a likely focus on the Voting ensemble model identified earlier. Rather than relying solely on aggregated performance metrics, this section intends to provide a more qualitative understanding of the model's behavior by showing how closely its predictions align with the actual experimental shear strength values. This involves visualizing the distribution of prediction errors and the direct correlation between predicted and experimental outcomes.

Fig. 7 presents histograms illustrating the distribution of prediction errors for the developed model. Subfigures (a), (b), and (c) display these distributions for the train data, the test data, and the combined data, respectively. Each plot also includes a cumulative percentage curve (right y-axis). The primary purpose of this figure is to provide a visual assessment of the magnitude, spread, and central tendency of the errors (defined as Experimental Shear Strength minus Predicted Shear Strength), showing the model's accuracy and potential biases beyond single-point metrics.

Across all three plots, the error distributions are clearly centered around zero, with the highest frequency bin located at or very near the zero-error mark. This indicates that the model does not exhibit a substantial systematic tendency to over-predict or under-predict. A noticeable difference exists in the dispersion of errors between the training set and the test set. The errors in the training set are closely clustered, with the vast majority falling approximately within a ±100 kN range. In contrast, the errors on the test set are more dispersed, spanning a wider range (approximately -500 kN to +500 kN), although the majority of errors remain concentrated closer to zero (e.g., within ±200 kN). The distributions in all plots approximate a unimodal, somewhat bell-shaped curve, suggesting that large errors occur less frequently than small errors. The test set distribution appears to have slightly heavier tails compared to the training set distribution. The cumulative percentage curves rise steeply around zero error, particularly for the training set, indicating a high proportion of

predictions with small errors. For the test set, the curve rises less sharply, visually representing the greater spread of errors. For instance, it can be visually estimated that approximately 80-90% of the test errors fall within ±250 kN.

The near-zero centering of the error distributions across all datasets is a positive indicator, suggesting that the model is largely unbiased. The tighter error distribution on the training data compared to the test data visually confirms the expected difference between training performance and generalization performance (overfitting), consistent with the lower MAE/RMSE values for the training set reported. The wider spread on the test set reflects the model's generalization error when predicting unseen data. While generally well-behaved, the slightly heavier tails in the test set error distribution suggest that the model occasionally produces larger prediction errors on new data.

In summary, Fig. 7 effectively visualizes the distribution of prediction errors, confirming the model's generally unbiased nature and illustrating the difference in error magnitudes between the training and test sets. The analysis offers information regarding the reliability and typical error range of the predictions, reinforcing the findings derived from the aggregate metrics.

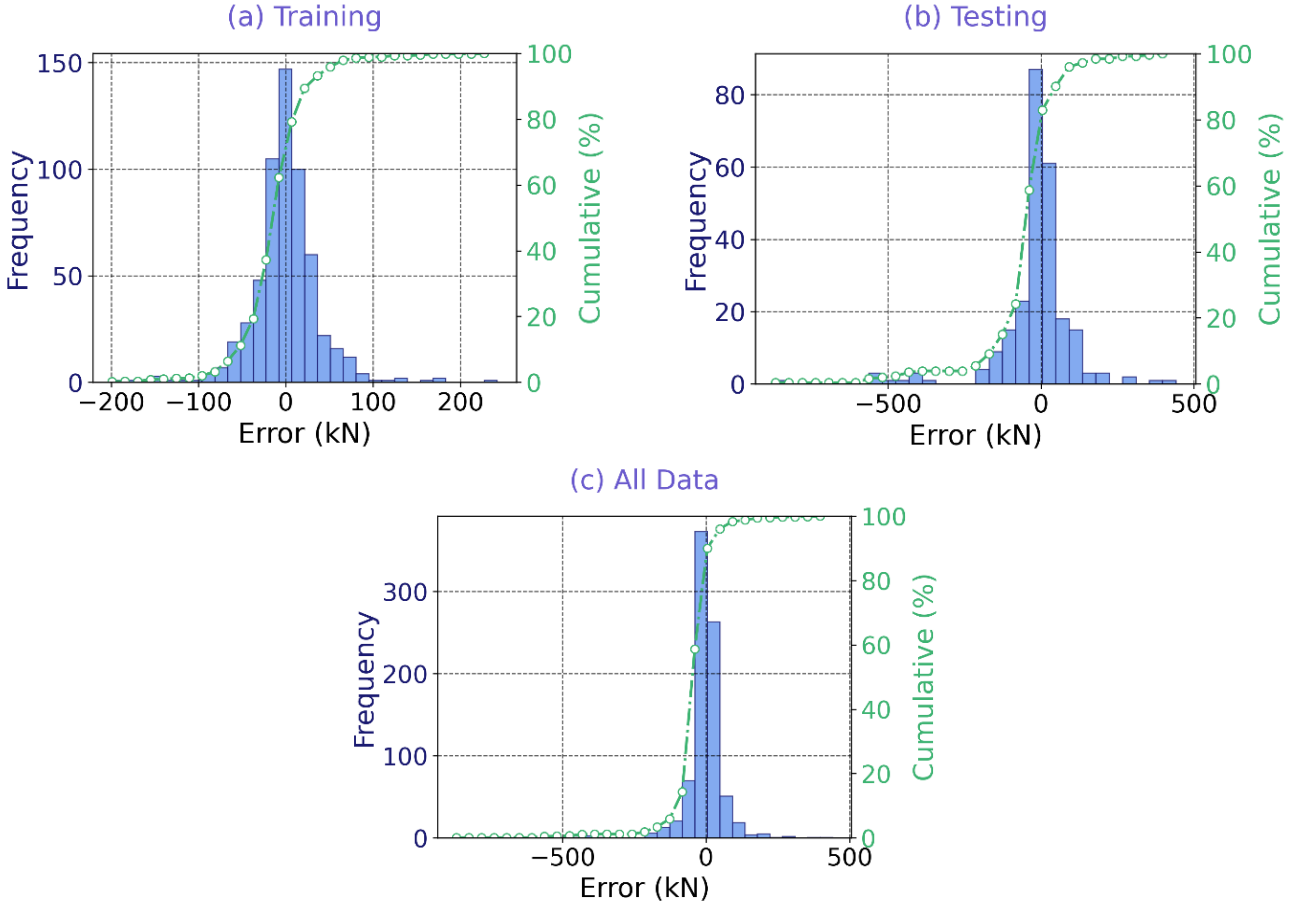

Fig. 7. Histograms of prediction errors (Experimental - Predicted Shear Strength) for the Voting ensemble model on: (a) Training data, (b) Testing data, and (c) All data

Fig. 8 presents regression charts that directly compare the shear strength values predicted by the Voting ensemble model against the actual experimental shear strength values for the training, testing, and combined datasets. Each plot displays the individual data points (predicted vs. actual), the line of perfect prediction (where Predicted Value equals Actual Value), and key performance metrics calculated for that specific dataset.

Across all three plots, the data points demonstrate a strong positive correlation and cluster closely around the "Perfect prediction" (y=x)

line. This indicates good agreement between the model's predictions and the experimental results. In the training dataset, the data points lie exceptionally close to the y=x line with minimal dispersion. The corresponding performance metrics displayed on the plot (R^2 = 0.997, MAE = 25.6 kN, RMSE = 39.0 kN) quantify this extremely high accuracy on the data the model was trained on. For the testing dataset, while still showing a strong linear trend along the y=x line, the data points exhibit noticeably more dispersion compared to the training set. This increased dispersion is reflected in the slightly lower, yet still very high, R^2 value of 0.961 and the higher error metrics (MAE = 71.7 kN, RMSE = 131.8 kN). The dispersion appears somewhat greater for higher shear strength values. The combined dataset integrates the trends from the training and test sets, showing overall strong performance (R^2 = 0.986) with the dispersion representing an average of the training and testing behavior.

The concentration of data points near the line of perfect agreement provides a visual representation of the Voting ensemble model's predictive accuracy for RCDB shear strength. This proximity suggests the model learned relevant patterns from the training data. While the scatter increases for the testing set, its limited extent implies generalization to unseen data, although performance on training and testing data typically differs. The calculated R^2 value of 0.961 for the test set corresponds to the model explaining approximately 96% of the observed variance in experimental shear strength for this data subset. Furthermore, the distribution of points around the line appears fairly uniform, consistent with the indication from Fig. 5 that substantial systematic prediction bias is not present.

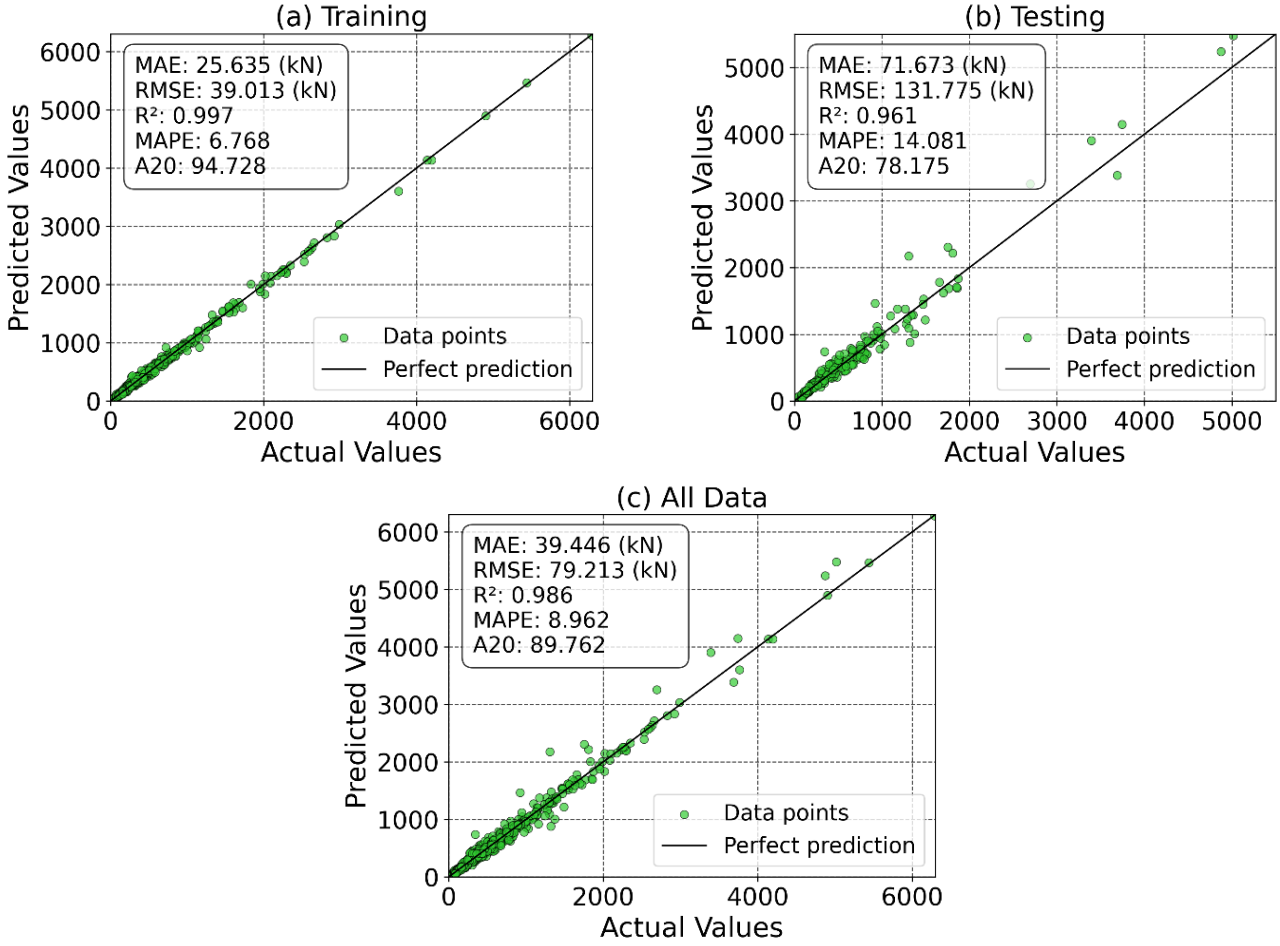

Fig. 8. Regression charts showing predicted versus actual shear strength values for the Voting ensemble model on: (a) Training data, (b) Testing data, and (c) All data, with the line of perfect prediction (y=x)

In conclusion, Fig. 8 provides strong visual validation of the Voting ensemble model's predictive power, showcasing the excellent correlation and agreement between predicted and experimental shear strengths, especially on the critical test dataset (R^2 = 0.961).

4.4. Comparison with other ML algorithms

To validate the proposed Voting ensemble model, its performance was benchmarked against three other machine learning algorithms: Support Vector Regression (SVR), a Multi-Layer Perceptron (MLP) as a representative artificial neural network, and Random Forest (RF). Hyperparameters for each model were optimized using grid search to ensure fair evaluation, with the search domains and best parameters summarized in Table 5.

Performance metrics for these models, alongside the proposed Voting (XGBoostR + CBR) model, were computed on both the train and test datasets (Fig. 9), and presented in Table 6. The results indicate that while all models achieved reasonable accuracy on the training data, their generalization capabilities on the test dataset varied significantly. The Voting model clearly outperformed the others, attaining the highest R^2 (0.961) and the lowest error metrics (RMSE = 131.775 kN, MAE = 71.673 kN, MAPE = 14.081%). In contrast, SVR showed the poorest performance on the test set (R^2 = 0.779, RMSE = 425.888 kN), indicating limitations in handling the dataset's complexity. Although MLP and RF delivered moderate results (R^2 of 0.939 and 0.885, respectively), they were still surpassed by the Voting model across all evaluation metrics.

4.5. Comparison with design standards

Table 5. Optimized Hyperparameters for SVR, MLP, and RF Models

Model	Hyperparameter	Search Domain	Best parameter
SVR	kernel	rbf - linear	linear
	C	0.1 - 100	1
	gamma	scale; auto; 0.001 - 10	scale
MLP	hidden_layer_sizes	(50), (100), (50,50), (100,50)	(100, 50)
	activation	relu ; tanh ; logistic	relu
	solver	adam ; sgd ; lbfgs	lbfgs
	alpha	0.0001 – 0.1	0.1
	batch_size	16 – 64 ; auto	16
RF	n_estimators	10 - 500	20
	max_depth	3 - 8	8
	max_features	0.3 – 1.0 ; sqrt ; log2	0.6

Table 6. Performance Metrics of ML Models on Train and Test Datasets

Criteria	RMSE (kN)	MAE (kN)	R^2	A20 (%)	MAPE (%)	RMSE (kN)	MAE (kN)	R^2	A20 (%)	MAPE (%)
Model	Train dataset					Test dataset				
MLP	144.741	102.640	0.940	51.935	30.240	222.725	132.105	0.939	51.786	35.605
RF	104.716	60.934	0.969	78.125	14.693	307.192	140.504	0.885	60.119	28.744
SVR	240.374	138.068	0.835	43.75	39.473	425.888	199.917	0.779	39.286	45.388
Voting (XGBR+CBR)	39.013	25.635	0.997	94.728	6.768	131.775	71.673	0.961	78.175	14.081

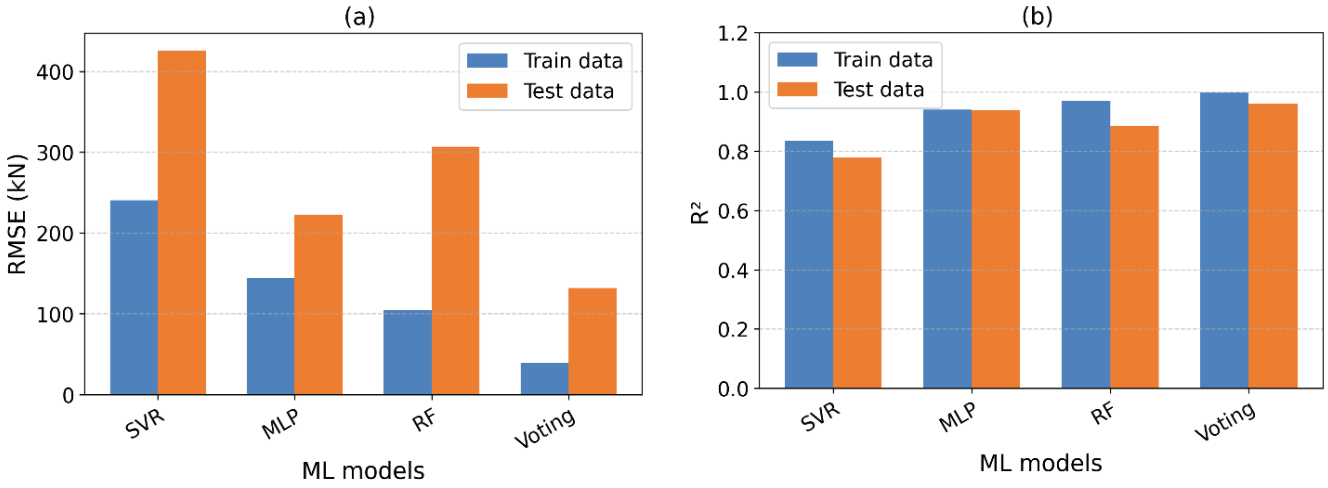

Fig. 9. Comparison of (a) RMSE and (b) R^2 across ML models for tand test datasets

To assess the practical applicability of the proposed Voting (XGBR + CBR) ensemble model, its predictive performance was compared against the ACI 318 and Eurocode 2 design standards. The evaluation was based on five performance metrics calculated for the alldataset, with the results presented in Table 7.

According to Table 7, the Voting model yielded an R^2 of 0.986, an RMSE of 65.542 kN, and a MAPE of 9.650%. In comparison, the metrics for Eurocode 2 were an R^2 of 0.927, an RMSE of 370.205 kN, and a MAPE of 35.981%. The ACI 318 standard resulted in an R^2 of 0.851, an RMSE of 432.998 kN, and a MAPE of 34.503%. Furthermore, the A20 index for the Voting model was 89.00%, while the values for Eurocode 2 and ACI 318 were 20.00% and 21.70%, respectively.

These metrics indicate that the Voting model's predictions have a stronger correlation and lower error magnitudes relative to the experimental data when compared to the calculations from the two design standards. The higher A20 value also suggests that a larger proportion of the model's predictions fall within a 20% margin of the experimental values. The Voting model therefore provides a more accurate alternative to the often oversimplified approaches of ACI 318 and Eurocode 2.

Table 7. Performance Comparison of Voting Model with ACI 318 and Eurocode 2 Standards

Criteria	RMSE (kN)	MAE (kN)	R^2	A20 (%)	MAPE (%)
ACI 318	432.998	204.138	0.851	21.70	34.503
Eurocode 2	370.205	202.502	0.927	20.00	35.981
Voting (XGBR + CBR)	65.542	39.105	0.986	89.00	9.650

4.6. SHAP value

The primary objective of this section is to improve the interpretability and transparency of the developed ML model, especially the best-performing one identified earlier (e.g., the Voting ensemble). Expanding beyond mere evaluation of predictive accuracy, this section seeks to elucidate the process by which the model generates its predictions for shear strength. This supports the reliability of the model and facilitates the derivation of engineering-relevant information from its learned relationships.

Fig. 10 presents a SHAP summary plot, which serves as a key tool for interpreting the predictions of the best-performing Voting ensemble model. This plot ranks the input features based on their overall importance to the model's predictions and simultaneously visualizes the direction and magnitude of each feature's impact across all samples in the dataset. The main purpose is to enhance model transparency by identifying the most influential factors driving the prediction of

shear strength and understanding how variations in these factors affect the output.

The features are arranged vertically in descending order of their global importance (based on the sum of absolute SHAP values). Beam width ('b') is shown as the most influential feature, followed by the shear span-to-depth ratio ('a/d'), beam height ('h'), concrete strength ('f_c'), and top plate width ('w_{tp}'). Features related to reinforcement, such as bottom reinforcement ratio ('rho'), vertical web reinforcement ratio ('ρ_v'), bottom plate width ('w_{bp}'), horizontal web reinforcement ratio (ρ_h'), and reinforcement strengths ('f_y', 'f_{yv}', 'f_{yh}'), are ranked lower, indicating comparatively less overall influence on the model predictions within this dataset context. It can be seen that the top four characteristics identified by SHAP - beam width ('b'), shear span-to-depth ratio ('a/d'), beam height ('h'), and concrete strength ('fc') - align closely with the primary variables in the shear strength equations of ACI 318 and Eurocode 2. These design codes incorporate beam dimensions ('b' and 'h') and concrete strength ('fc') as core factors in their semi-empirical formulas for deep beam shear capacity, reflecting their critical influence on stress distribution and load transfer mechanisms. Additionally, the significant negative impact of 'a/d' revealed by SHAP corroborates its pivotal role in defining deep beam behavior through arch action, which is a cornerstone of strut-and-tie models (STM). This alignment demonstrates that the model has autonomously learned physically meaningful relationships consistent with established design principles, enhancing confidence that it is not merely a "black box" but a tool that captures the fundamental mechanics of RCDB.

The SHAP summary plot reveals specific effects of individual features on the predicted shear strength. For beam width ('b'), higher values (indicated by red dots) are associated with a substantial increase in the predicted shear strength (large positive SHAP values), while lower values (blue dots) are associated with a decrease. Conversely, higher values of the shear span-to-depth ratio ('a/d', red dots) predominantly lead to a reduction in the predicted shear strength (negative SHAP values), whereas lower values (blue dots) tend to increase it. Increased beam height ('h', red dots) generally results in a higher predicted shear strength (positive SHAP values). Similarly, higher concrete strength ('f_c', red dots) tends to increase the predicted shear strength (positive SHAP values), although its impact is less pronounced compared to the top-ranked geometric features. An increase in the top plate width ('w_{tp}', red dots) generally correlates with a higher predicted shear strength. For web reinforcement parameters such as the vertical web reinforcement ratio ('ρ_v') and the yield strength of vertical web reinforcement ('f_{yv}'), higher values (red dots) typically correspond to positive SHAP values, suggesting they contribute to an increase in the predicted strength, although their overall influence is smaller than that of the top-ranked features.

This SHAP analysis reveals that the ML model has learned relationships consistent with established structural engineering principles for deep beams. The model identifies beam geometry (width 'b', height 'h', plate width 'w_{tp}') and the shear span-to-depth ratio ('a/d') as the dominant factors influencing shear strength predictions. The directions of these influences align with expectations: larger dimensions increase capacity, while a larger 'a/d' ratio (making the beam more slender-like) decreases the deep beam shear strength prediction. Higher concrete strength ('f_c') also correctly contributes positively to the predicted capacity. The model also attributes a positive, albeit smaller, influence to the presence and strength of reinforcement (both bottom longitudinal 'ρ', 'f_y' and web 'ρ_v', 'ρ_h', 'f_{yv}', 'f_{yh}'), acknowledging its role in shear resistance.

The feature importance ranking provided by SHAP offers more information than the simple linear correlation shown in Fig. 4. Notably, 'a/d',

which had a negligible linear correlation, is identified by SHAP as the second most important feature overall, highlighting the model's ability to capture its significant, potentially non-linear, impact on shear strength. The strong influence of geometric parameters and concrete strength aligns well with their prominence in design code equations [27, 28], and strut-and-tie models [29, 30] for deep beams. The findings support the physical understanding of shear transfer mechanisms in deep beams being highly sensitive to geometry and a/d ratio.

In essence, Fig. 10 provides substantial understanding of the developed ML model's behavior via the SHAP summary plot. The analysis corroborates that the model identifies key parameters consistent with structural engineering principles - beam geometry (b, h), shear span-to-depth ratio (a/d), and concrete strength (f_c) - as the main determinants influencing the prediction of shear strength, and that these factors operate as expected.

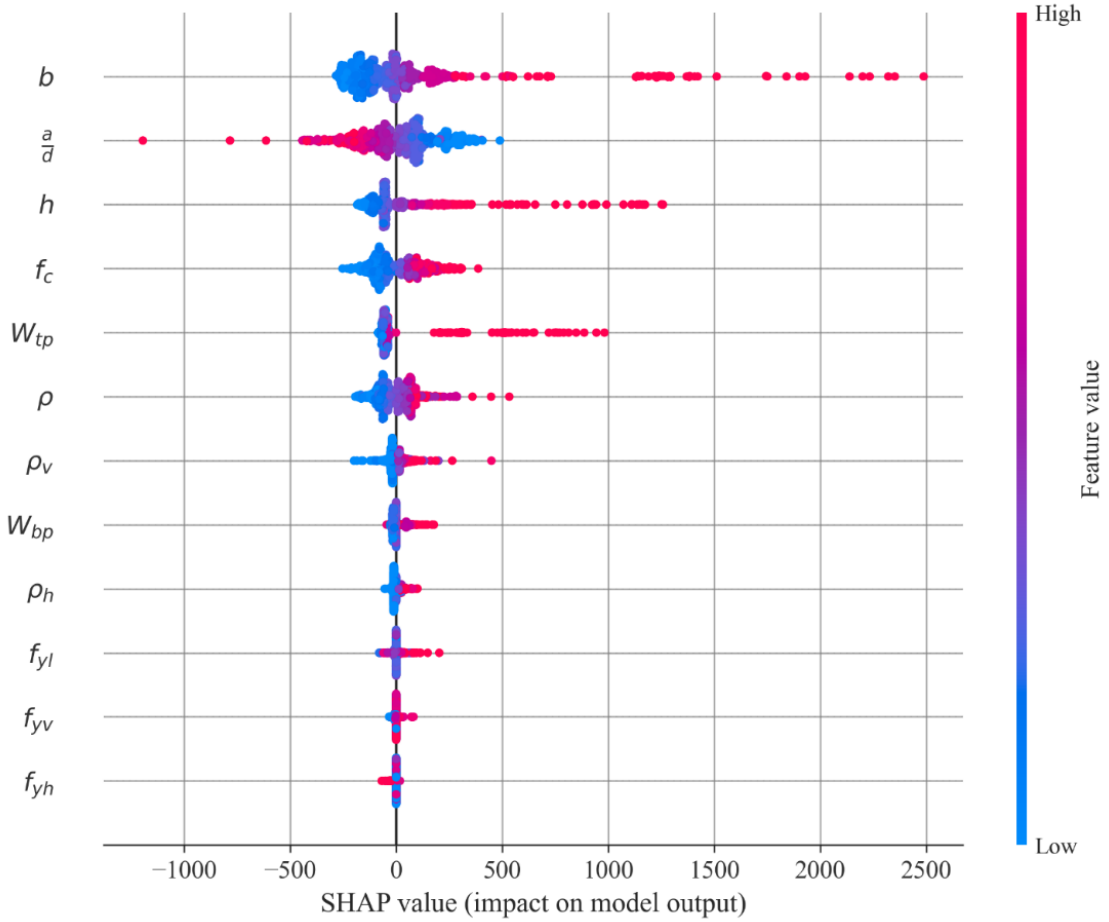

Fig. 10. SHAP summary plot illustrating global feature importance and the impact distribution of each input feature on the Voting ensemble model's output

A bar chart summarizing global feature importance, as determined by SHAP analysis, is shown in Fig. 11. The input features are ordered in the chart according to their mean absolute SHAP values, indicating the overall contribution of each input parameter to the model's shear strength prediction outputs.

The chart clearly ranks beam width ('b') as the most important feature (highest mean |SHAP| value ≈ 194), followed by shear span-to-depth ratio ('a/d' ≈ 135), beam height ('h' ≈ 119), concrete strength ('f_c' ≈ 94), top plate width ('w_{tp}' ≈ 85), and bottom reinforcement ratio ('ρ' ≈ 66). Other features related to reinforcement exhibit considerably lower mean absolute SHAP values. This visualization confirms that the model places significant

emphasis on the beam's primary geometry (b, h, w_{tp}), the a/d ratio, and concrete strength when predicting shear strength. These factors exert the largest average influence on the predictions. Reinforcement details, while still contributing, are interpreted by the model as having a less dominant role overall compared to the main geometric and material properties for this dataset. Fig. 11 presents a summary of the feature importance obtained from the SHAP analysis, reinforcing the detailed findings presented in Fig. 10. The ranking is consistent with established engineering principles, indicating that deep beam capacity is primarily governed by geometry and concrete strength.

In brief, Fig. 11 presents a clear ranking of feature importance based on mean absolute SHAP values, highlighting the dominance of geometric parameters, a/d ratio, and concrete strength in the model's predictions. This concludes the model interpretation component.

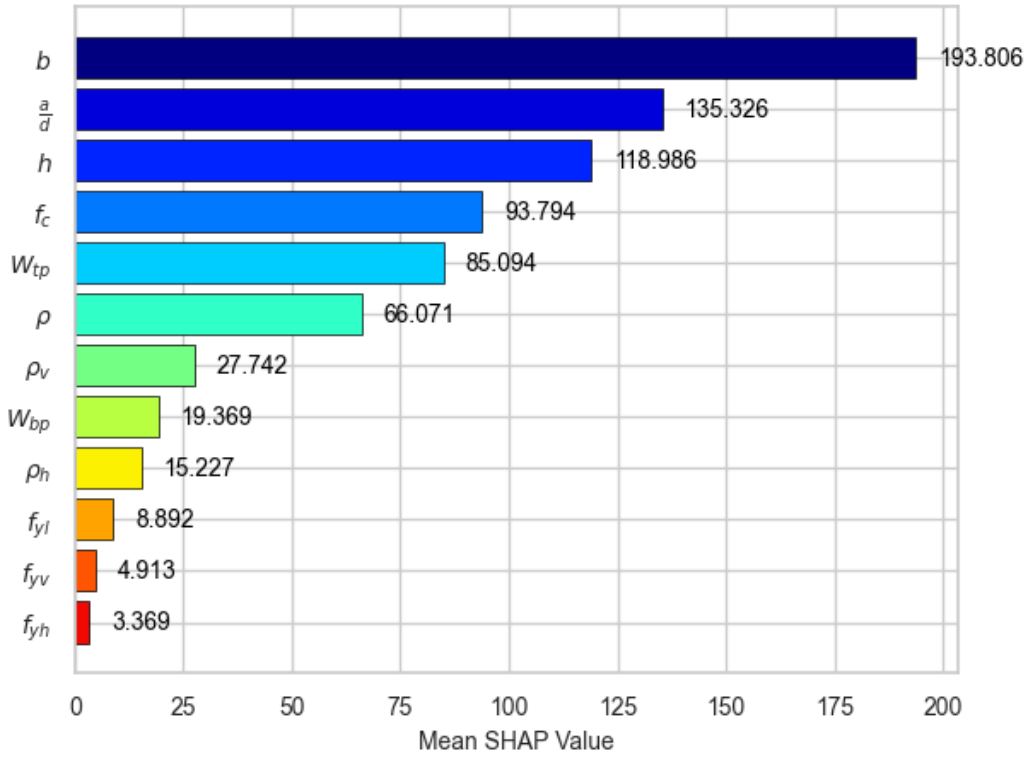

Fig. 11. Bar chart ranking input features based on their mean absolute SHAP values, indicating global feature importance for the Voting ensemble model

5. Conclusion

This study examined the application of ML techniques to predict the shear strength of RCDB. The work was motivated by the difficulties traditional empirical and theoretical models face in representing the non-linear behavior characteristic of these structural elements. Using a database of 840 experimental results, individual models (XGBR, CBR) along with ensemble approaches (Stacking, Voting) were developed and assessed. The main conclusions derived from this study are enumerated subsequently:

Both individual and ensemble ML models demonstrated significant potential for accurately predicting the shear strength of RCDB. Among the models assessed, the Voting ensemble, which integrated the strengths of optimized XGBoost and CatBoost regressors, exhibited the most resilient and accurate performance on the independent test dataset, achieving a high coefficient of determination (R^2 = 0.961) and low error metrics (RMSE = 131.8 kN, MAE = 71.7 kN). Comparative analysis revealed that the Voting model outperformed other ML algorithms, including Support Vector Regression (SVR), Multi-Layer Perceptron (MLP), and Random Forest (RF), as

well as traditional design standards such as ACI 318 and Eurocode 2, which exhibited higher error metrics and lower correlation.

Interpretability analysis using SHAP clarified the behavior of the best-performing model. The analysis corroborated that the model's predictions are primarily influenced by factors consistent with structural engineering principles. Beam width ('b'), shear span-to-depth ratio ('a/d'), beam height ('h'), and concrete strength ('f_c') were identified as the most influential parameters, aligning with their established impact on deep beam shear mechanisms such as arch action.

The SHAP analysis also validated the relationships learned by the model, illustrating, for example, that increased beam dimensions and concrete strength positively affect predicted shear strength, while an increase in the shear span-to-depth ratio has a negative effect. This consistency with domain knowledge enhances confidence in the model's predictive capabilities.

In conclusion, this research successfully demonstrates that a Voting ensemble ML approach can serve as a robust and reliable tool for predicting the shear strength of RCDB. The developed model not only achieves high accuracy but also offers interpretability through SHAP analysis, rendering it a promising alternative or supplement to conventional methods in structural engineering practice and research. However, the dataset, though comprehensive, primarily consists of controlled laboratory experiments, a common limitation in structural engineering research due to scarcity of systematically measured in-situ data. This controlled setting facilitates establishment of a robust foundational model by minimizing environmental uncertainties, serving as a critical initial step. Incorporation of field-scale or in-situ structural performance data represents a promising direction for future research to enhance real-world applicability. Besides, extension to serviceability aspects, such as crack width, deflection, and ductility, as well as construction-related parameters, including curing conditions and reinforcement detailing, offers a valuable pathway for future research to create a comprehensive design tool. Future work could expand evaluation of model performance by incorporating datasets with diverse concrete types (e.g., high-performance or fiber-reinforced concrete), varied failure modes, field-scale or in-situ structural performance data, serviceability aspects (e.g., crack width, deflection, ductility), and construction-related parameters (e.g., curing conditions, reinforcement detailing). Such extensions, building on the robust foundational model established in this study, aim to enhance real-world applicability and develop a comprehensive design tool for RCDB.

References

[1] K.-H. Tan, F.-K. Kong, S. Teng, L. Guan. (1995). High-strength concrete deep beams with effective span and shear span variations. *Structural Journal*, 92(4), 395-405. DOI: 10.14359/991

[2] C. Ma, S. Wang, J. Zhao, X. Xiao, C. Xie, X. Feng. (2023). Prediction of shear strength of RC deep beams based on interpretable machine learning. *Construction and Building Materials*, 387, 131640. https://doi.org/10.1016/j.conbuildmat.2023.131640

[3] A.H. Gandomi, A.H. Alavi, D.M. Shadmehri, M.G. Sahab. (2013). An empirical model for shear capacity of RC deep beams using genetic-simulated annealing. *Archives of Civil and Mechanical Engineering*, 13(3), 354-369. https://doi.org/10.1016/j.acme.2013.02.007

[4] M. Pal, S. Deswal. (2011). Support vector regression based shear strength modelling of deep beams. *Computers & Structures*, 89(13-14), 1430-1439. https://doi.org/10.1016/j.compstruc.2011.03.005

[5] R.S. Londhe. (2011). Shear strength analysis and prediction of reinforced concrete transfer

beams in high-rise buildings. *Structural Engineering & Mechanics*, 37(1), 39-59. DOI:10.12989/SEM.2011.37.1.039

[6] G. Russo, R. Venir, M. Pauletta. (2005). Reinforced concrete deep beams-shear strength model and design formula. *Structural Journal*, 102(3), 429-437. DOI: 10.14359/14414

[7] G. Aguilar, A.B. Matamoros, G.J. Parra-Montesinos, J.A. Ramírez, J.K. Wight. (2002). Experimental evaluation of design procedures for shear strength of deep reinforced concrete beams. *ACI Structural Journal*, 99(4), 539-548. American Concrete Institute (ACI). DOI:10.14359/12123

[8] A.C.I. Committee. (2005). Building code requirements for structural concrete (ACI 318-05) and commentary (ACI 318R-05). American Concrete Institute.

[9] B.S. Institution. (2004). Eurocode 2: Design of concrete structures: Part 1-1: General rules and rules for buildings. British Standards Institution.

[10] F.J. Vecchio, M.P. Collins. (1986). The modified compression-field theory for reinforced concrete elements subjected to shear. *ACI Journal Proceedings*, 83(2), 219-231. DOI: 10.14359/10416

[11] W.-Y. Lu, I.-J. Lin, H.-W. Yu. (2013). Shear strength of reinforced concrete deep beams. *ACI Structural Journal*, 110(4), 671-680. DOI: 10.14359/51685752

[12] K.-H. Yang, A.F. Ashour. (2011). Strut-and-Tie Model Based on Crack Band Theory for Deep Beams. *Journal of Structural Engineering*, 137(10), 1030-1038. https://doi.org/10.1061/(ASCE)ST.1943-541X.0000351

[13] S.-J. Hwang, H.-J. Lee. (2002). Strength Prediction for Discontinuity Regions by Softened Strut-and-Tie Model. *Journal of Structural Engineering*, 128(12), 1519-1526. https://doi.org/10.1061/(ASCE)0733-9445(2002)128:12(1519)

[14] K. Megahed. (2024). Prediction and reliability analysis of shear strength of RC deep beams. *Scientific Reports*, 14, 14590. https://doi.org/10.1038/s41598-024-64386-w

[15] D.-C. Feng, W.-J. Wang, S. Mangalathu, G. Hu, T. Wu. (2021). Implementing ensemble learning methods to predict the shear strength of RC deep beams with/without web reinforcements. *Engineering Structures*, 235, 111979.https://doi.org/10.1016/j.engstruct.2021.111979

[16] T.-A. Nguyen, H.-B. Ly, H.-V.T. Mai, V.Q. Tran. (2021). On the Training Algorithms for Artificial Neural Network in Predicting the Shear Strength of Deep Beams. *Complexity*, 2021, 5548988. https://doi.org/10.1155/2021/5548988

[17] A.F. Ashour, L.F. Alvarez, V.V. Toropov. (2003). Empirical modelling of shear strength of RC deep beams by genetic programming. *Computers & Structures*, 81(5), 331-338. https://doi.org/10.1016/S0045-7949(02)00437-6

[18] M.-Y. Cheng, M.-T. Cao. (2014). Evolutionary multivariate adaptive regression splines for estimating shear strength in reinforced-concrete deep beams. *Engineering Applications of Artificial Intelligence*, 28, 86-96. https://doi.org/10.1016/j.engappai.2013.11.001

[19] K. Le Nguyen, H.T. Trinh, T.T. Nguyen, H.D. Nguyen. (2023). Comparative study on the performance of different machine learning techniques to predict the shear strength of RC deep beams: Model selection and industry implications. *Expert Systems with Applications*, 230, 120649. https://doi.org/10.1016/j.eswa.2023.120649

[20] A. Tiwari, A.K. Gupta, T. Gupta. (2024). A robust approach to shear strength prediction of reinforced concrete deep beams using ensemble learning with SHAP interpretability. *Soft Computing*, 28, 6343-6365. https://doi.org/10.1007/s00500-023-09495-w

[21] P. Chetchotisak, J. Teerawong, S. Yindeesuk. (2022). Modified interactive strut-and-tie modeling of reinforced concrete deep beams and corbels. *Structures*, 45, 284-298. https://doi.org/10.1016/j.istruc.2022.08.116

[22] K.H. Tan, G.H. Cheng. (2006). Size Effect on Shear Strength of Deep Beams: Investigating with Strut-and-Tie Model. *Journal of Structural Engineering*, 132(5), 673-685. https://doi.org/10.1061/(ASCE)0733-9445(2006)132:5(673)

[23] N. Zhang, K.-H. Tan. (2007). Size effect in RC deep beams: Experimental investigation and STM verification. *Engineering Structures*, 29(12), 3241-3254. https://doi.org/10.1016/j.engstruct.2007.10.005

[24] A.V. Dorogush, V. Ershov, A. Gulin. (2018). CatBoost: gradient boosting with categorical features support. *ArXiv*. https://doi.org/10.48550/arXiv.1810.11363

[25] T. Chen, C. Guestrin. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794.* https://doi.org/10.1145/2939672.2939785

[26] T.-A. Nguyen, M.H. Nguyen, H.-B. Ly. (2024). Unified machine learning approach for predicting CFST column axial load capacity. *Innovative Infrastructure Solutions*, 9, 295. https://doi.org/10.1007/s41062-024-01593-4

[27] K.S. Ismail, B.O. Mawlood, A.H. Mohammad, M. Guadagnini. (2025). Numerical investigation of the shear capacity of reinforced concrete continuous deep beams. Structural Concrete, 27(2), 1489-1506. https://doi.org/10.1002/suco.202401131

[28] L. Todisco, O. Bayrak, K.-H. Reineck. (2018). ACI-DAFSTB database for tests on deep beams and comparisons with code provisions. *Structural Concrete*, 19(1), 296-304. https://doi.org/10.1002/suco.201700061

[29] P. Chetchotisak, J. Teerawong, S. Yindeesuk, J. Song. (2014). New strut-and-tie-models for shear strength prediction and design of RC deep beams. *Computers and Concrete*, 14(1), 19-40.

[30] F.B.A. Beshara, I.G. Shaaban, T.S. Mustafa. (2015). Strut-and-Tie modelling of RC continuous deep beams. *THE 13TH ARAB STRUCTURAL ENGINEERING CONFERENCE ASEC 2015.*