

**Article info****Type of article:**

Original research paper

DOI:

<https://doi.org/10.58845/jstt.utt.2026.en.6.5.127-149>

***Corresponding author:**

Email address:

trinhnq@utt.edu.vn**Received:** 09/11/2025**Received in Revised Form:** 03/08/2026**Accepted:** 03/09/2026

Interpretable hybrid machine learning frameworks for predicting the axial load capacity of concrete filled double skin steel tube columns

Dang Khoa Do¹, Thuy-Anh Nguyen², Trinh Quoc Ngo^{3*}, Viet-Hung Phan⁴¹The University of Sydney, NSW 2006, Australia; dado0559@uni.sydney.edu.au²Research group on Industry 4.0 in Transportation (I4T group), University of Transport Technology, Hanoi 100000, Vietnam; anhnt@utt.edu.vn³University of Transport Technology, Hanoi 100000, Vietnam; trinhnq@utt.edu.vn⁴University of Transport and Communications, Hanoi 100000, Vietnam; Phanviethung@utc.edu.vn

Abstract: Predicting the axial load capacity of concrete-filled double-skin steel tube (CFDST) columns is difficult because of the complex mechanical interactions between the concrete core and the dual steel tubes. Although machine learning (ML) methods provide advantages in prediction, current models frequently lack sufficient hyperparameter optimization and mechanistic interpretability. To resolve these limitations, novel hybrid ML frameworks are proposed integrating eXtreme Gradient Boosting (XGBoost) with advanced metaheuristic algorithms, specifically the Bald Eagle Search (BES) and Aquila Optimizer (AO). A database containing 153 experimental samples defined by eight geometric and material parameters was used to develop and validate the models. Systematic optimization comparisons demonstrated that the XGB-BES framework provides high generalization and accuracy, producing a coefficient of determination (R^2) of 0.998, a Root Mean Square Error (RMSE) of 62.054 kN, and an A20 index of 0.957 when tested on validation data. Furthermore, the inherent opacity of complex ML models was addressed using post-hoc interpretability methods such as SHAP (Shapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations). Analyses of global and local feature importance demonstrated that the diameter and thickness of the outer tube most strongly affect the structural capacity, which connects the computational predictions to fundamental structural mechanics principles. These findings indicate that the developed XGB-BES model functions as an accurate and mechanically transparent instrument for the evaluation and design of CFDST columns.

Keywords: Machine Learning, Concrete-filled double skin steel tubes, axial load capacity, SHAP, LIME.

1. Introduction

Concrete-filled double-skin steel tubes

(CFDST) are advanced composite structural members comprising two concentric steel tubes

with a concrete-filled annular space. The inherent mechanical synergy between these constituent materials, whereby the outer steel tube affords high tensile strength and confinement [1,2], while the concrete core provides substantial compressive resistance, yields exceptional load-bearing capacity, ductility, and favorable constructability [3]. Beyond fundamental mechanical superiority, CFDST columns exhibit significant energy dissipation and large deformation capacities, making them highly suitable for seismic-resistant designs [4,5]. Furthermore, this composite configuration enhances environmental and thermal robustness; the outer steel shell mitigates corrosion in harsh offshore or industrial environments, while the concrete core provides vital thermal insulation to maintain structural stability at elevated temperatures [6]. Consequently, CFDST elements are increasingly specified in the design of high-rise buildings, long-span bridges, and complex infrastructural projects [7].

Accurately quantifying the ultimate axial load capacity (P_u) of CFDST columns presents a profound analytical challenge, driven by highly non-linear structural responses and complex multi-axial interfacial mechanics. The load-sharing behavior between the dual steel shells and the confined concrete core is dependent on interacting factors such as material constitutive properties, multi-axial stress states, geometric slenderness, and specific boundary conditions [8,9]. Although conventional experimental studies generate basic datasets for different cross-sectional typologies, these empirical methods require significant physical resources and are limited in their parametric scope [9–11]. Alternatively, detailed numerical simulations, particularly the finite element method (FEM), supply reliable predictive capacity but necessitate substantial computational overhead, extensive mesh sensitivity analyses, and complicated calibration of interfacial contact mechanics and material plastic damage [5,12,13].

Therefore, a need exists within structural engineering for a computationally economical, generalized, and accurate analytical framework to assess CFDST load capacities.

To mitigate the computational demands associated with conventional methodologies, machine learning (ML) is frequently applied as a data-driven approach for estimating CFDST load-bearing capacities. Existing literature contains various applications of predictive algorithms, including Artificial Neural Networks (ANN), Gene Expression Programming (GEP), and Adaptive Neuro-Fuzzy Inference Systems, to model structural responses [14–16]. Rather than relying on isolated models, recent investigations commonly employ algorithmic hybridization to account for complex mechanical interactions. For example, combining GEP and ANN with Response Surface Methodology has been utilized to determine structural capacity from specific experimental datasets [14,16]. Furthermore, evolutionary optimization techniques, including the integration of ANNs with Genetic Algorithm, have been implemented to evaluate the axial capacity of columns across varying cross-sectional morphologies [15].

Although foundational predictive algorithms exhibit known efficacy, current ML applications within structural engineering face specific methodological limitations. The requisite hyperparameter optimization typically depends upon computationally intensive grid searches or manual adjustments, which are prone to converging on local instead of global optima. Additionally, the opaque nature of these nonlinear models prevents mechanistic transparency. While these models attain high predictive accuracy, they yield limited physical understanding of the underlying structural mechanisms, restricting their adoption in practical design. Furthermore, existing research lacks investigations that combine robust ensemble methods, such as XGBoost, with contemporary metaheuristic optimization

algorithms, including BES and AO. Applying these global search techniques to systematically optimize high-dimensional models for predicting CF DST capacity constitutes an unaddressed area of study.

Consequently, this research addresses these methodological limitations by presenting a computational framework that integrates the XGBoost algorithm with two recent metaheuristic optimizers, namely BES and AO. XGBoost was selected based on its established predictive accuracy, computational efficiency, and capability to model complex, high-dimensional data [17]. Simultaneously, BES and AO are incorporated to autonomously search the hyperparameter space, thereby reducing the probability of converging on local optima and assisting in the identification of a globally optimal model architecture [18,19].

The primary focus of this study lies in combining these search algorithms with XGBoost to generate a predictive tool for CF DST column axial capacity, which is trained and evaluated using an experimental database consisting of 153 samples. To clarify the traditionally opaque nature of ensemble learning methods, post-hoc interpretability frameworks, specifically SHAP and LIME, are applied to quantitatively assess the structural influence of each input parameter and provide transparency regarding the underlying predictive mechanisms. The subsequent sections of the manuscript are organized such that Section 2 details the theoretical basis of the ML algorithms and the dataset, Section 3 provides the performance evaluation and optimization discussion alongside an interpretability analysis, and Section 4 summarizes the findings and limitations of the study.

2. Computational methodology and empirical database

2.1. Predictive architectures and optimization algorithms

2.1.1. XGBoost

The eXtreme Gradient Boosting (XGBoost)

architecture was selected as the primary predictive engine due to its exceptional capacity to map high-dimensional, non-linear structural data [17]. By iteratively minimizing a regularized objective function, this ensemble framework systematically corrects the residual errors of preceding weak learners, thereby suppressing the overfitting tendencies common in structural capacity prediction models. This process is guided by an optimized objective function, which includes two main components: a loss function that measures the difference between the predicted and actual values, and a regularization component to control model complexity. The integration of this regularization term helps to mitigate overfitting, a common issue in ML models.

The performance of an XGBoost model is significantly dependent on the tuning of a set of hyperparameters. Some of the important hyperparameters include `n_estimators`, which specifies the total number of decision trees to be built; `max_depth`, which limits the maximum depth of each tree to control complexity; and `learning_rate`, which determines the contribution of each tree to the final result, thereby affecting the algorithm's convergence speed [20].

2.1.2. Metaheuristic search frameworks (BES and AO)

To circumvent the computational demands and the tendency to converge on local optima inherent to conventional grid searches, the XGBoost hyperparameter space was optimized employing two metaheuristic algorithms. Specifically, the BES and AO were chosen based on their multi-stage stochastic search mechanisms, which establish a necessary mathematical equilibrium between broad global exploration and focused localized exploitation.

BES algorithm

The BES algorithm is a nature-inspired metaheuristic optimization algorithm that simulates the hunting strategy of the bald eagle [19]. The BES optimization process is divided into three main

stages. The first stage is the selection phase, where the algorithm identifies a search space based on existing information, similar to how an eagle selects an area to hunt. The second stage is the search phase, in which the eagle moves in a spiral trajectory within the selected space to efficiently search for prey. The final stage is the swooping phase, which simulates the eagle's rapid dive from the best-found position toward the prey, corresponding to the convergence toward the optimal solution in the search space.

AO algorithm

The AO is a population-based optimization algorithm developed by simulating the hunting behaviors of the Aquila eagle in nature [18]. The operational principle of AO is based on a flexible balance between exploration and exploitation phases through four distinct hunting methods. Expanded exploration is achieved by simulating the eagle's high-altitude soaring to identify prey areas. Following this, a narrowed exploration simulates the behavior of contour flying over a selected target area. For the exploitation phase, expanded exploitation describes a low-altitude attack with a steep trajectory, while narrowed exploitation simulates the action of swooping and capturing the prey, converging on the solution with the highest fitness.

2.1.3. Post-hoc interpretability frameworks (SHAP and LIME)

SHAP

SHAP [21] provides a game-theoretic methodology designed to interpret the outputs of machine learning models. This approach bridges optimal credit allocation with local explanations through the application of classical Shapley values derived from game theory. The primary objective is to calculate the precise contribution of individual features to the predictive output of a specific instance. By assigning a discrete importance value to each variable, SHAP guarantees both consistency and local accuracy. Specifically, the SHAP value indicates the alteration in the expected

model prediction when a particular feature is introduced into the conditioning set. Consequently, this framework facilitates both global interpretability, achieved by aggregating Shapley values across the entire dataset, and localized interpretation, enabled by examining feature contributions for individual predictions.

LIME

LIME [22] constitutes a model-agnostic methodology formulated to interpret the predictions of arbitrary classifiers or regressors by generating faithful local approximations using an interpretable surrogate model. The fundamental principle relies on perturbing the input features of a specific data sample and monitoring the corresponding fluctuations in the predictive output. Subsequently, a linear surrogate model is trained on this generated dataset of perturbed samples, with individual instances weighted according to their spatial proximity to the original data point within the high-dimensional feature space. This localized procedure isolates the specific feature interactions driving a given prediction, thereby establishing mechanistic transparency for complex, black-box architectures within the immediate vicinity of the evaluated instance.

2.2. Empirical CFDST database and statistical distribution

The ML models in this study were developed using a database comprising 153 experimental samples of CFDST columns. This dataset was systematically compiled from 12 published articles [5,9,12,23–31]. To ensure data consistency and model relevance, all selected samples were experimentally tested under axial compression loading conditions. The database is structured with eight input parameters that define the key geometric and material properties of the columns; these inputs include: column length (V_1), outer diameter (V_2), outer thickness (V_3), inner diameter (V_4), inner thickness (V_5), yield strength of the outer steel tube (V_6), yield strength of the inner steel tube (V_7), and concrete compressive strength (V_8). The

sole output parameter for prediction is the ultimate P_u of the column. For the purpose of model development and evaluation, the dataset was randomly partitioned into two subsets: a training set, accounting for 70% of the data, and a testing set, comprising the remaining 30%. This division strategy is widely adopted in data-driven structural engineering studies to establish a balanced trade-off between model training and validation [32,33]. Allocating 70% of the data to the training set ensures that the ML algorithm has access to sufficient information to learn the underlying patterns and complex relationships between the input variables and the P_u . The remaining 30% serves as an independent testing set, which is

large enough to provide a statistically significant evaluation of the model's generalization performance on unseen data, thereby minimizing the risk of overfitting. To further ensure the robustness of the predictive model, the random partitioning was conducted to preserve the statistical distribution of key parameters across both training and testing sets. While K-fold cross-validation is a powerful technique, a single robust 70/30 split was prioritized in this study to facilitate the detailed, instance-specific local interpretability analysis (LIME and SHAP) on a fixed, representative testing set. This approach ensures that the black box explanations are consistent and directly traceable to specific validation samples.

Table 1. Statistical summary of the geometric parameters, material properties, and experimental axial capacities for the 153 CFDST column specimens

Variables	Unit	Min	Mean	Media	Max	Std
Input						
Column length (V_1)	mm	230.00	907.85	660.00	3502.00	703.01
Outer diameter (V_2)	mm	74.70	212.77	160.00	603.40	139.38
Outer thickness (V_3)	mm	0.59	2.84	3.00	8.00	1.28
Inner diameter (V_4)	mm	38.00	120.02	76.00	402.10	92.23
Inner thickness (V_5)	mm	0.55	2.33	2.00	8.00	1.02
Yield strength of the outer steel tube (V_6)	MPa	205.00	375.07	366.00	549.00	86.76
Yield strength of the inner steel tube (V_7)	MPa	205.00	381.95	374.50	549.00	83.38
Concrete compressive strength (V_8)	MPa	18.70	37.43	45.64	63.40	14.91
Output						
Axial load capacity (P_u)	kN	293.00	1836.75	1195.60	8839.80	1769.43

The statistical characteristics and distribution of the dataset are summarized in Table 1 and visualized in Fig. 1 and Fig. 2. The statistical distributions aggregated in Table 1 reflect a highly diverse empirical parameter space. Notably, the ultimate axial capacities (P_u) span a massive operational envelope, from 293.00 kN to nearly 8840 kN, ensuring the predictive framework is trained across both small-scale structural members and massive, high-capacity columns. The large

standard deviation (1769.43) and the notable difference between the mean (1836.75 kN) and the median (1195.60 kN) indicate a right-skewed distribution, which can also be observed in the box plots in Fig. 2. The geometric parameters (V_1 - V_5) also show wide variability; for instance, column length (V_1) spans from 230 mm to 3502 mm, and outer diameter (V_2) ranges from 74.7 mm to 603.4 mm. Similar to the output, the mean values for V_1 , V_2 , V_4 , and V_5 are larger than their respective

medians, suggesting a right-skew in their distributions as well. The material properties (V_6 - V_8) cover a practical range of steel grades and concrete strengths. Concrete compressive strength (V_8) is the only parameter to show a left-

skewed distribution, with a mean (37.43 MPa) lower than its median (45.64 MPa). Fig. 1 provides a visual representation of the data distribution for each input variable. It confirms the skewness identified in Table 1.

Distribution of the input variables

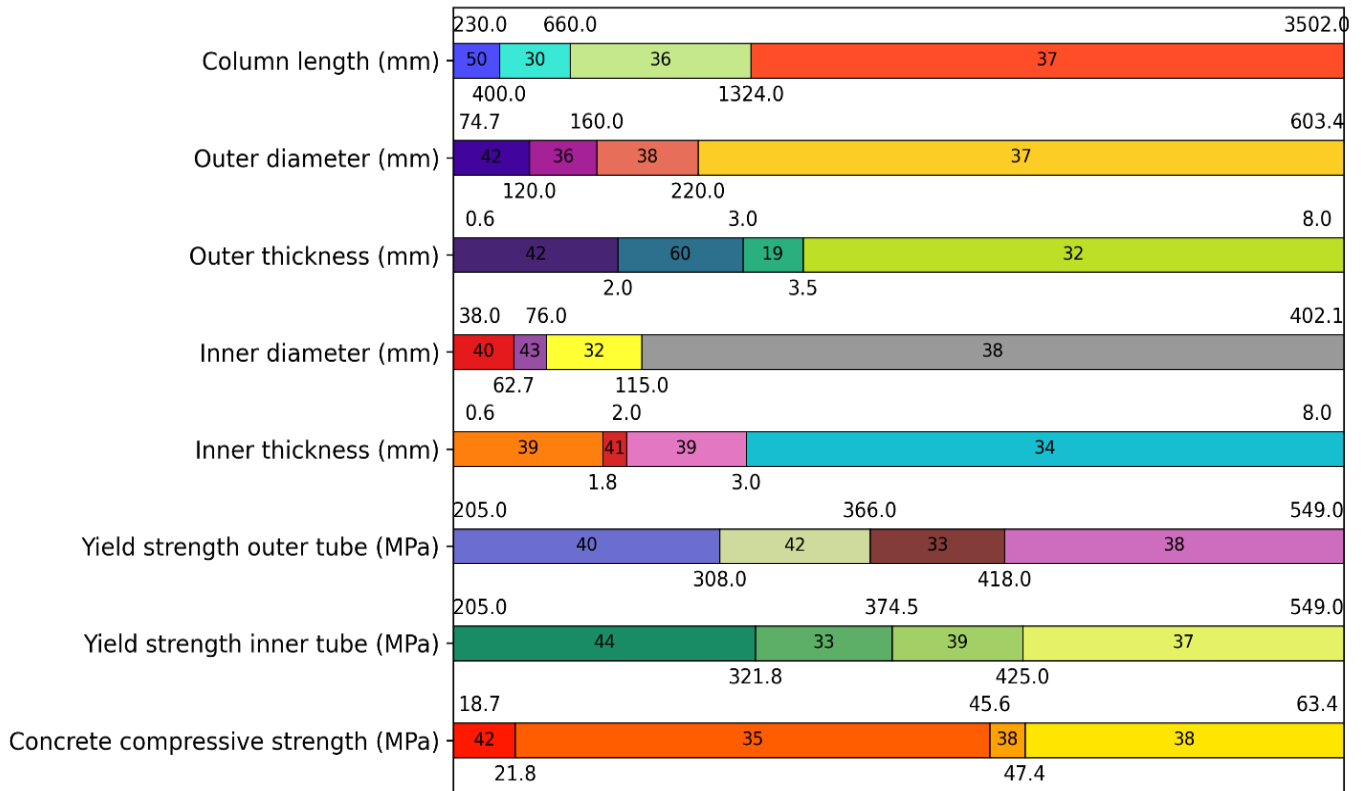


Fig. 1. Global distribution and parametric ranges of the eight input variables utilized in the experimental database

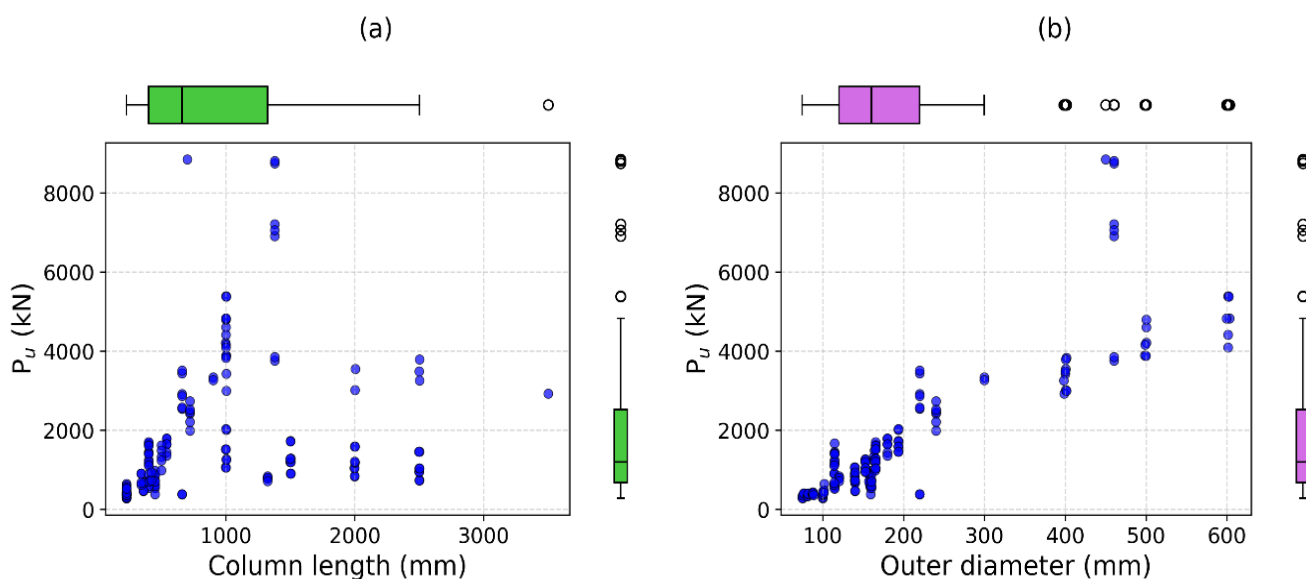


Fig. 2. Bivariate scatter relationships and marginal distributions pairing individual structural input parameters against the experimental P_u .

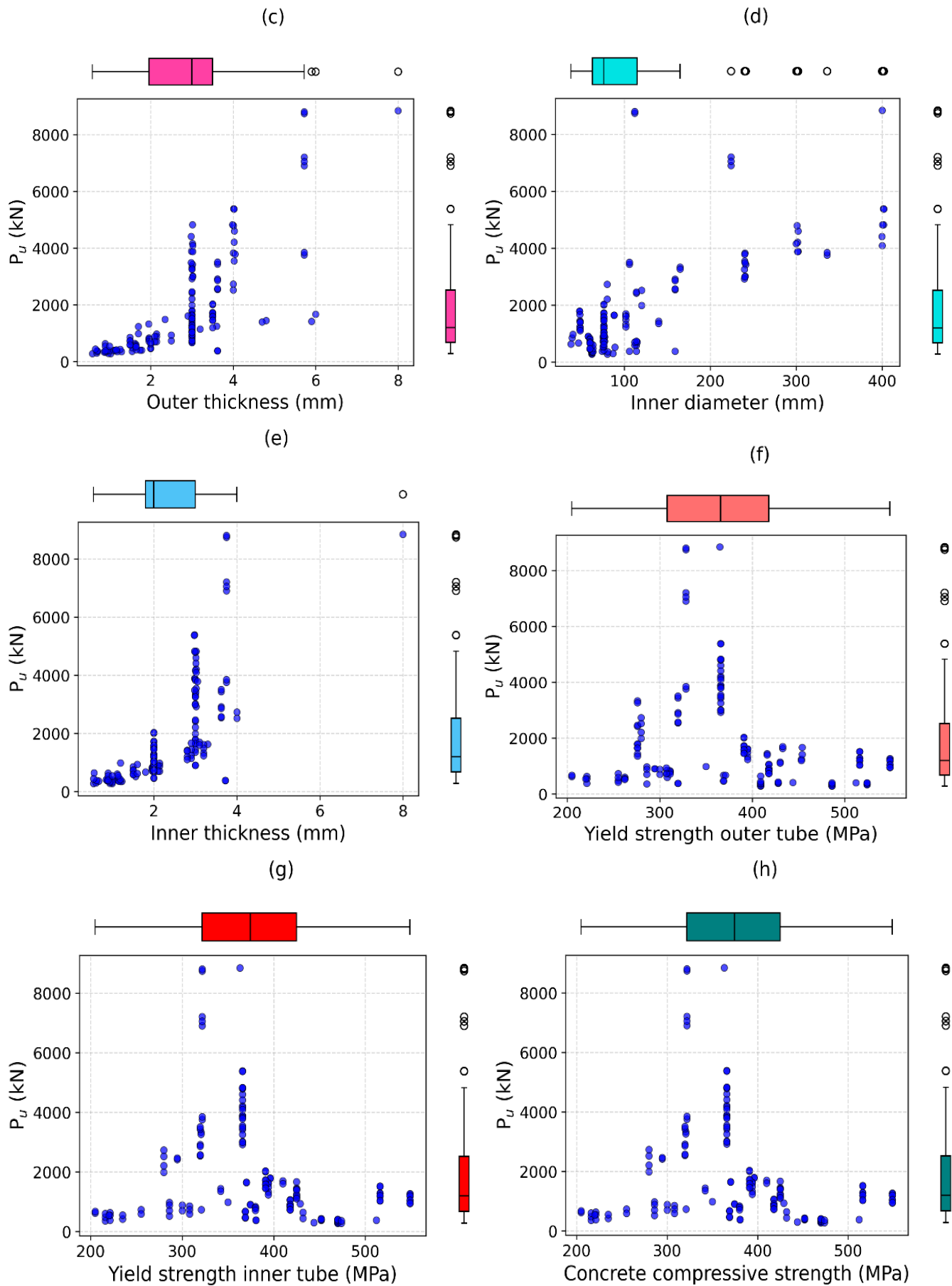


Fig. 2. (continued)

Fig. 2 provides a detailed visual analysis of each input variable against the P_u using a combination of scatter plots and marginal box plots. This combined visualization reveals several complex, multivariate relationships. A strong positive correlation is evident for outer diameter (V_2), outer thickness (V_3), and the steel yield strengths (V_6 , V_7). This is not observed as a simple linear trend, but rather as a clear upward shift in the central cluster of P_u values as the input variable increases. For instance, in plot (b), P_u values are clustered below 2000 kN for outer diameters near 100-200 mm, but the entire cluster shifts upward to the 4000-6000 kN range for diameters of 500-600 mm. Similarly, for yield strengths (V_6 , V_7), the main data cluster jumps from a low P_u range (at 250-300 MPa) to a significantly higher range (at 350-400 MPa). However, these plots also highlight the impact of confounding variables. For V_6 and V_7 , some samples with very high yield strength (>450 MPa) exhibit low P_u . This suggests that for these specific samples, unfavorable geometric properties (e.g., high slenderness from V_1 or low diameter from V_2) created a failure mode, such as buckling, that occurred at a lower load and prevented the high material strength from being fully utilized. The influence of other variables is more nuanced. Column length (V_1) shows a complex, non-linear relationship. The highest capacity columns are all concentrated at shorter lengths (< 1500 mm), which is consistent with the mitigation of slenderness-induced buckling effects. Conversely, inner diameter (V_4) shows a non-positive relationship; the highest P_u values are found at the smallest inner diameters, and the data variance narrows as inner diameter increases. Concrete compressive strength (V_8) displays a generally positive trend, but with significant dispersion (high variance) at any given strength value, indicating it is an influential but not singularly dominant factor. A key statistical feature, heteroscedasticity, is prominent in the plots for V_2 and V_3 . This is seen as a fanning or funnel shape: at low values of these

inputs, the P_u is tightly clustered in a narrow, low-capacity band. At high values, the P_u variance becomes much wider. This has a clear physical interpretation: a small diameter or thickness acts as a limiting factor, ensuring low capacity. In contrast, a large diameter or thickness enables high capacity but does not guarantee it, as the final P_u will still depend on other factors (like length or concrete strength). Finally, the marginal box plots for P_u consistently show a large number of high-side outliers, representing these high-performance columns where a favorable combination of multiple input variables was achieved.

2.3. Statistical evaluation metrics

To evaluate the performance of the ML models in this study, five common statistical criteria were utilized, including: the R^2 , Root Mean Square Error (RMSE), the A20 index, Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). These criteria are represented by Equations (1–5) [34–36]. R^2 assesses the degree of model fit to the data by comparing the variation of the actual data with the predictions; a value closer to 1 indicates a better model. RMSE measures the average deviation of the predictions from the actual values. MAE is similar to RMSE but calculates the average of the absolute differences between predicted and actual data. MAPE represents the average percentage difference between the predicted and actual values. The A20 index measures the proportion of cases where the absolute difference between the predicted and actual value is within or equal to 20% of the actual value. A higher A20 value, particularly one approaching 1, suggests high prediction accuracy.

$$R^2 = 1 - \frac{\sum_{i=1}^N (P_{pre_i} - P_{exp_i})^2}{\sum_{i=1}^N (P_{pre_i} - \bar{P}_{pre})^2} \quad (1)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_{pre_i} - P_{exp_i})^2} \quad (2)$$

$$MAE = \frac{1}{N} |(P_{pre_i} - P_{exp_i})| \quad (3)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|(P_{pre_i} - P_{exp_i})|}{P_{pre_i}} \times 100\% \quad (4)$$

$$A20(P_{pre_i}, P_{exp_i}) = \frac{1}{N} \sum_{i=1}^N \begin{cases} 1 & \text{if } \frac{|(P_{pre_i} - P_{exp_i})|}{P_{pre_i}} \leq 0.2 \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

where N is the number of samples, P_{pre} is the predicted value, P_{exp} is the actual value, and $\overline{P_{exp}}$ is the average of the actual value.

3. Results and discussion

3.1. Metaheuristic hyperparameter optimization and convergence analysis

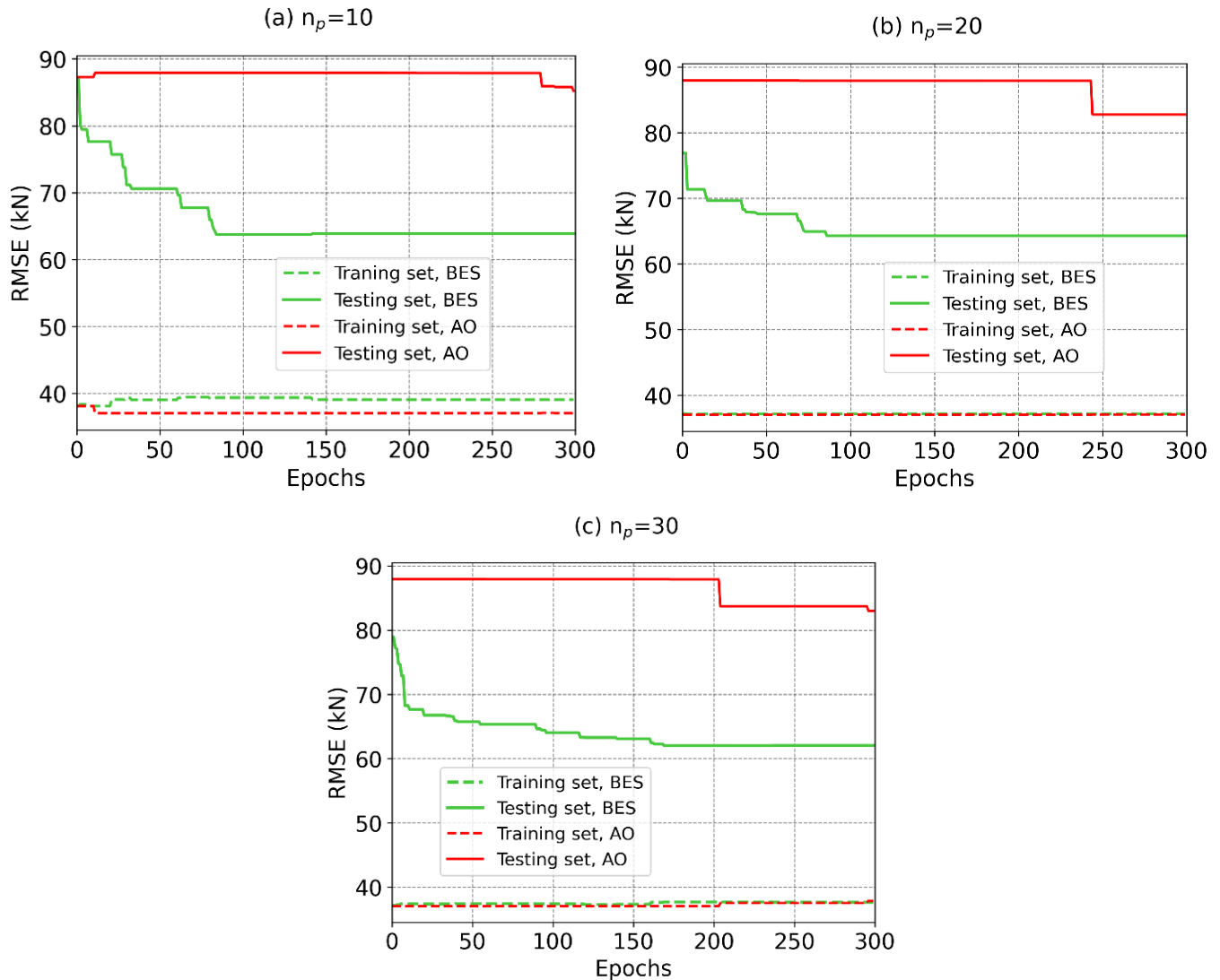


Fig. 3. Algorithmic convergence trajectories of the XGBoost objective function optimized via the BES and AO frameworks across three distinct population scales ($n_p = 10, 20, 30$).

To ensure predictive accuracy within the XGBoost architecture, hyperparameter optimization is necessary to balance model bias and variance. Rather than utilizing traditional approaches such as Grid or Random search, two metaheuristic algorithms, BES and AO, were

applied to determine the optimal model parameters. The optimization procedure focused on five hyperparameters that dictate the learning capacity and complexity of the ensemble. Both continuous and discrete search spaces were constrained to maintain computational stability and

limit overfitting. Specifically, the total number of sequential decision trees ($n_estimators$) was restricted to a range of 1 to 500, the step size shrinkage ($learning_rate$) was defined from 0.01 to 0.3, and the maximum tree depth (max_depth) was set between 1 and 8. Additionally, the regularization parameters controlling the subsampling ratio of the training instances ($subsample$) and the fraction of features evaluated per tree ($colsample_bytree$) were limited to the $[0.5, 1.0]$ interval.

The definition of these hyperparameter boundaries was determined through a combination of standard operating ranges derived from recent XGBoost literature [37,38] and preliminary sensitivity analyses. These initial evaluations were conducted to isolate parameter domains that affect model performance while maintaining computational stability. Bounding the search space, which included limiting the estimator count to control computational overhead and restricting the learning rate to maintain stable algorithmic convergence, guides the BES and AO metaheuristics toward feasible parameter regions. This approach prevents the evaluation of computationally impractical solutions. All remaining XGBoost hyperparameters were kept at their standard default values.

The optimization framework functions through an iterative feedback process that connects the XGBoost regressor with the selected metaheuristic algorithms. Following the stochastic initialization of candidate hyperparameter populations within the specified boundaries, each iteration evaluates the respective solutions by training the XGBoost model and calculating a defined fitness function. While various statistical metrics were utilized for the final model assessment, the RMSE was chosen as the sole objective function for optimization. This decision is based on the capacity of RMSE to provide an absolute measurement of error magnitude in dimensional units consistent with the target

variable (kN). Consequently, this measurement corresponds with the minimization procedures of the BES and AO algorithms more effectively than relative correlational metrics such as R^2 . Based on the computed RMSE fitness scores, the BES and AO algorithms update the spatial coordinates of their population agents, utilizing distinct exploration and exploitation mechanisms to converge on a hyperparameter configuration that produces the minimum prediction error.

The performance of the applied metaheuristic algorithms depends on their internal control parameters, primarily the population size (n_p) and the maximum number of iterations, or epochs. To establish a balance between computational cost and global convergence capabilities, a limit of 300 epochs was applied uniformly across all simulations. Simultaneously, a parametric study was executed using three population scales ($n_p = 10, 20$, and 30) to assess the effect of agent density on search space exploration. This experimental matrix, which incorporates the two optimization algorithms across three population configurations, resulted in six distinct hybrid XGBoost architectures that were subsequently evaluated using the RMSE objective function on both the training and independent testing datasets.

The convergence trajectories for the six hybrid architectures are presented in Fig. 3, which plots the RMSE objective function over the 300 optimization epochs. Quantitative results for the training ($RMSE_{train}$) and independent testing ($RMSE_{test}$) datasets are summarized in Table 2. Observation of the convergence curves shows an initial rapid decrease in the validation error during the early exploratory phase, which is followed by asymptotic stabilization as the algorithms shift toward localized exploitation. This downward trend indicates the capability of both metaheuristic strategies to minimize prediction error and refine the XGBoost hyperparameter space.

An evaluation of the statistical metrics in

Table 2 indicates a distinct difference in performance between the two metaheuristic optimization methods. Across all examined population configurations, the AO selected hyperparameter subsets that produced slightly lower training errors ($RMSE_{train}$ near 37.10 kN), which were consistently accompanied by higher testing errors ($RMSE_{test}$ exceeding 82.00 kN). This disparity between training and validation accuracy suggests a tendency toward overfitting, wherein the AO method optimized for the training data while compromising out-of-sample generalization. In

contrast, the BES algorithm displayed a greater capacity to manage the bias-variance tradeoff. Adjusting the population density from $n_p = 10$ to 20 resulted in stable performance metrics, and increasing the search capacity to $n_p = 30$ permitted the BES algorithm to identify the optimal hyperparameter configuration. This BES-optimized architecture generated the lowest validation error ($RMSE_{test} = 62.054$ kN) while retaining a comparable training error ($RMSE_{train} = 37.650$ kN), verifying its generalization performance on independent structural datasets.

Table 2. Comparative hyperparameter configurations and resulting RMSE metrics for the XGBoost architectures optimized via the BES and AO algorithms.

	$n_p = 10$		$n_p = 20$		$n_p = 30$	
	BES	AO	BES	AO	BES	AO
$RMSE_{train}$ (kN)	39.083	37.097	37.201	37.090	37.650	37.881
$RMSE_{test}$ (kN)	63.894	85.222	64.287	82.786	62.054	82.981
$n_{estimators}$	271	187	319	169	239	166
learning_rate	0.093	0.287	0.104	0.3	0.190	0.3
max_depth	4	4	7	4	4	3
subsample	0.686	0.973	0.696	0.988	0.753	0.989
colsample_bytree	0.845	0.973	0.806	0.988	0.511	0.980

Therefore, the hybrid architecture optimized utilizing the BES algorithm with a population scale of $n_p = 30$ was selected as the final predictive model. This configuration minimized the absolute validation error ($RMSE_{test} = 62.054$ kN) and produced the smallest difference between the training and testing subsets ($\Delta_{RMSE} = 24.404$ kN). This low divergence functions as an indicator of structural reliability, establishing that the model retains the generalization capacity necessary to evaluate independent CFDST column geometries while avoiding overfitting to the training data. The specific hyperparameter vector that produced this optimal performance consists of an $n_{estimators}$ value of 239, a learning_rate of 0.190, a max_depth of 4, a subsample of 0.753, and a colsample_bytree of 0.511.

3.2. Predictive performance of the globally optimal XGB-BES framework

Following the optimization procedure, the predictive performance of the finalized XGB-BES framework ($n_p = 30$) was evaluated utilizing the empirical database of CFDST axial load capacities. The accuracy and error distributions of the resulting architecture are illustrated through regression and error density analyses (Fig. 4). Furthermore, the corresponding statistical performance metrics are summarized in Table 3.

The regression analysis presented in Fig. 4a demonstrates the correlation between the experimental and predicted ultimate axial capacities. Data points for both the training and independent testing sets display homoscedasticity, grouping along the $y = x$ equivalence line without apparent systematic bias. The majority of predictive outputs fall within the $\pm 20\%$ error margins, and a large portion remains within the $\pm 10\%$ boundaries. This agreement across the

evaluated load range, encompassing both low-load and high-load columns, indicates that the optimized XGB-BES framework generalizes its nonlinear structural mappings to independent empirical data. To analyze the predictive variance, the absolute residual magnitudes and their density distribution are presented using a polar spatial plot (Fig. 4b) and a Kernel Density Estimation (KDE) representation (Fig. 4c). The polar plot indicates that the residual magnitudes for the training and

testing sets are concentrated within the inner radial limits, demonstrating a low frequency of large predictive outliers. Furthermore, the KDE analysis displays a leptokurtic error distribution centered at a mean absolute residual of 0 kN. The symmetric decrease in probability density extending outward from zero indicates that the XGB-BES framework functions with minimal systematic bias, yielding consistent capacity predictions throughout the dataset.

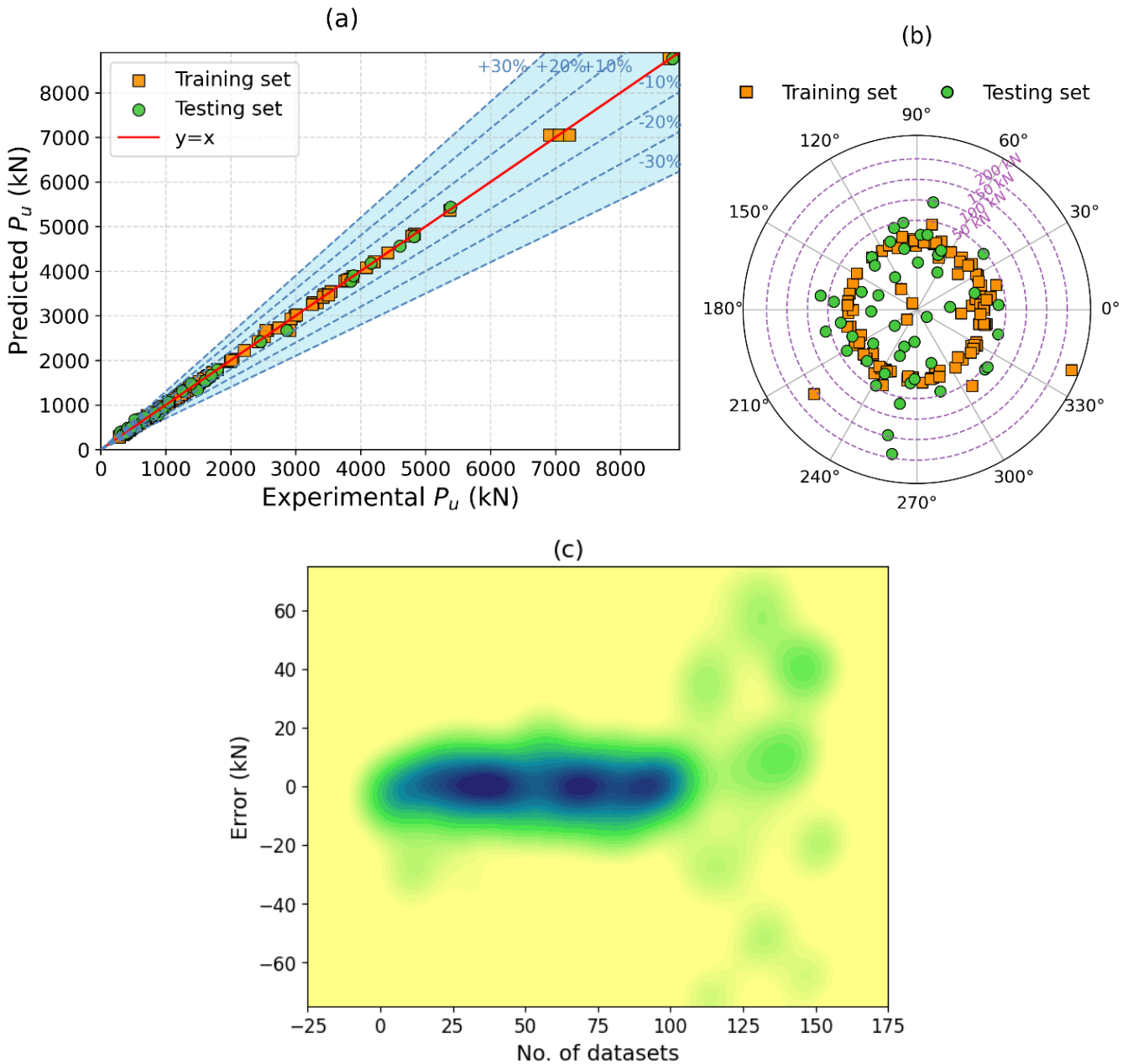


Fig. 4. Predictive performance of the optimal XGB-BES framework: (a) regression analysis of predicted versus experimental capacity, (b) polar spatial distribution of residual magnitudes, and (c) kernel density estimation of signed prediction residuals.

Table 3. Comprehensive statistical evaluation metrics demonstrating the global predictive fidelity of the optimal XGB-BES framework.

	R^2	RMSE (kN)	MAE (kN)	MAPE	A20
Training set	0.999	37.650	14.952	0.009	1.000
Testing set	0.998	62.054	48.211	0.058	0.957
All dataset	0.999	46.358	24.951	0.023	0.987

The statistical evaluations summarized in Table 3 support the prior graphical observations, indicating high predictive accuracy across all data partitions. The XGB-BES framework attained $R^2 > 0.998$ on both the training and independent testing cohorts, establishing a stable relationship between the computational outputs and empirical capacities. The application of the model to the unseen validation subset resulted in a slight increase in the absolute error metrics, including a change in RMSE from 37.650 kN to 62.054 kN. This minor reduction in performance indicates that the metaheuristic optimization effectively mitigated algorithmic overfitting. Furthermore, the model produced an A20 reliability index of 0.957 on the testing set. Within a structural engineering context, characterized by material uncertainties and geometric imperfections that dictate design safety margins, this metric is significant. It demonstrates that approximately 96% of the out-of-sample predictions occur within a 20% tolerance, providing a reliable analytical tool for practical CFDST column assessment.

From a structural engineering standpoint, the accuracy of the XGB-BES framework meets the requirements for composite design. In applications where material heterogeneities and geometric imperfections dictate safety margins, estimating ultimate loads with a 95.7% probability of falling within a 20% error threshold ($A20 = 0.957$) constitutes a sufficient analytical standard. Furthermore, evaluating the validation RMSE (62.054 kN) relative to the capacity range of the dataset, which includes ultimate axial loads up to 8840 kN, indicates a low error-to-capacity ratio. This low predictive variance and the established

mitigation of overfitting confirm the proposed hybrid model as an accurate, scale-invariant, and reliable computational method for assessing CFDST columns.

3.3. Stochastic stability analysis

Considering the high-capacity architecture of XGBoost and the limited size of the empirical database ($N = 153$), evaluating the model on a single data partition introduces a risk of stochastic sampling bias. To mitigate the potential for algorithmic overfitting, namely, the tendency to memorize the training data rather than capture the underlying structural mechanics, a comprehensive stochastic stability analysis was performed. This validation was conducted using a repeated random sub-sampling validation protocol, in which the optimized XGB-BES framework underwent 300 independent training and testing iterations.

In each successive iteration, the empirical database was randomly shuffled and re-partitioned into distinct subsets, allocating 70% of the data for training and 30% for testing. Following this data division, the XGB-BES model was re-trained from its initial state and assessed using the newly generated, unseen validation cohort.

The statistical dispersion of predictive performance across the 300 independent iterations demonstrated high consistency (Fig. 5). The framework produced a mean testing R^2 of 0.993, constrained by a narrow standard deviation R^2 of ± 0.0033 . Similarly, the absolute error metrics displayed a mean testing RMSE of 124.57 kN, accompanied by a standard deviation of ± 12.36 kN.

The minimal width of these standard deviation bands confirms that the predictive

accuracy of the XGB-BES framework is scale-invariant and highly robust against random data partitioning. This rigorous statistical cross-validation demonstrates that the hybrid model possesses strong out-of-sample generalization capabilities, effectively modeling the complex, nonlinear physical mechanics that govern CFDST column capacity rather than relying on localized statistical noise or advantageous dataset splits.

3.4. Comparative analysis with established structural design codes

To advance the validation of the XGB-BES framework beyond algorithmic optimization and demonstrate its practical utility in civil engineering,

a rigorous benchmarking analysis was performed against established structural design standards. Specifically, the experimental axial capacities of the 153 CFDST specimens were compared to the predictive equations outlined by ACI [39], Eurocode 4 (EC4) [40], or AISC 360-16 [41]. Given that conventional design codes are predominantly calibrated for single-skin composite members, the theoretical code-based capacities were calculated by superimposing the individual yield capacities of the inner and outer steel tubes alongside the compressive resistance of the annular concrete core, in accordance with standard principles of composite mechanics.

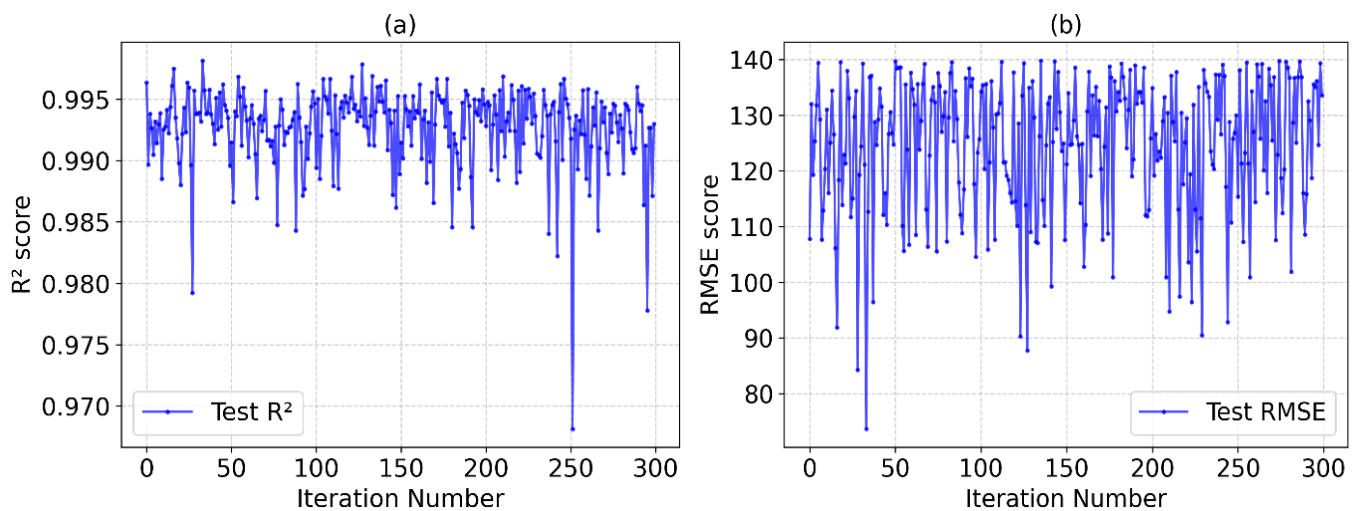


Fig. 5. Statistical dispersion of predictive performance across the 300 independent iterations on the test set: (a) R^2 , (b) RMSE.

Table 4. Statistical criteria of design codes for predicting the P_u of CFDST columns.

	R^2	RMSE (kN)	MAE (kN)	MAPE	A20
ACI [39]	0.956	754.035	431.242	0.267	0.556
EC4 [40]	0.950	1104.922	606.592	0.324	0.536
AISC 360-16 [41]	0.956	861.687	464.633	0.262	0.582

A comparative evaluation of the predictive error metrics reveals a pronounced performance differential. The design code calculations produced are shown in Table 4. As depicted in the comparative scatter distribution (Fig. 6), the code-derived predictions display considerable variance alongside a strict conservative bias, consistently underestimating the true ultimate capacity. This dispersion is fundamentally attributable to the

limitations of simplified empirical formulations, which inadequately capture the nonlinear, multi-axial stress states and the synergistic confinement mechanisms interacting between the dual steel tubes and the internal concrete core.

In contrast, the XGB-BES framework effectively captures these complex, high-dimensional structural interactions, resulting in a substantially reduced global RMSE of 46.358 kN

and an R^2 of 0.999. The tight convergence of the ML predictions along the ideal $y = x$ equivalence line demonstrates that the optimized gradient boosting architecture significantly outperforms conventional design codes. Ultimately, this

comparative validation confirms that the XGB-BES model functions not merely as a theoretical statistical exercise, but as a highly accurate and structurally viable alternative to traditional codified capacity estimations.

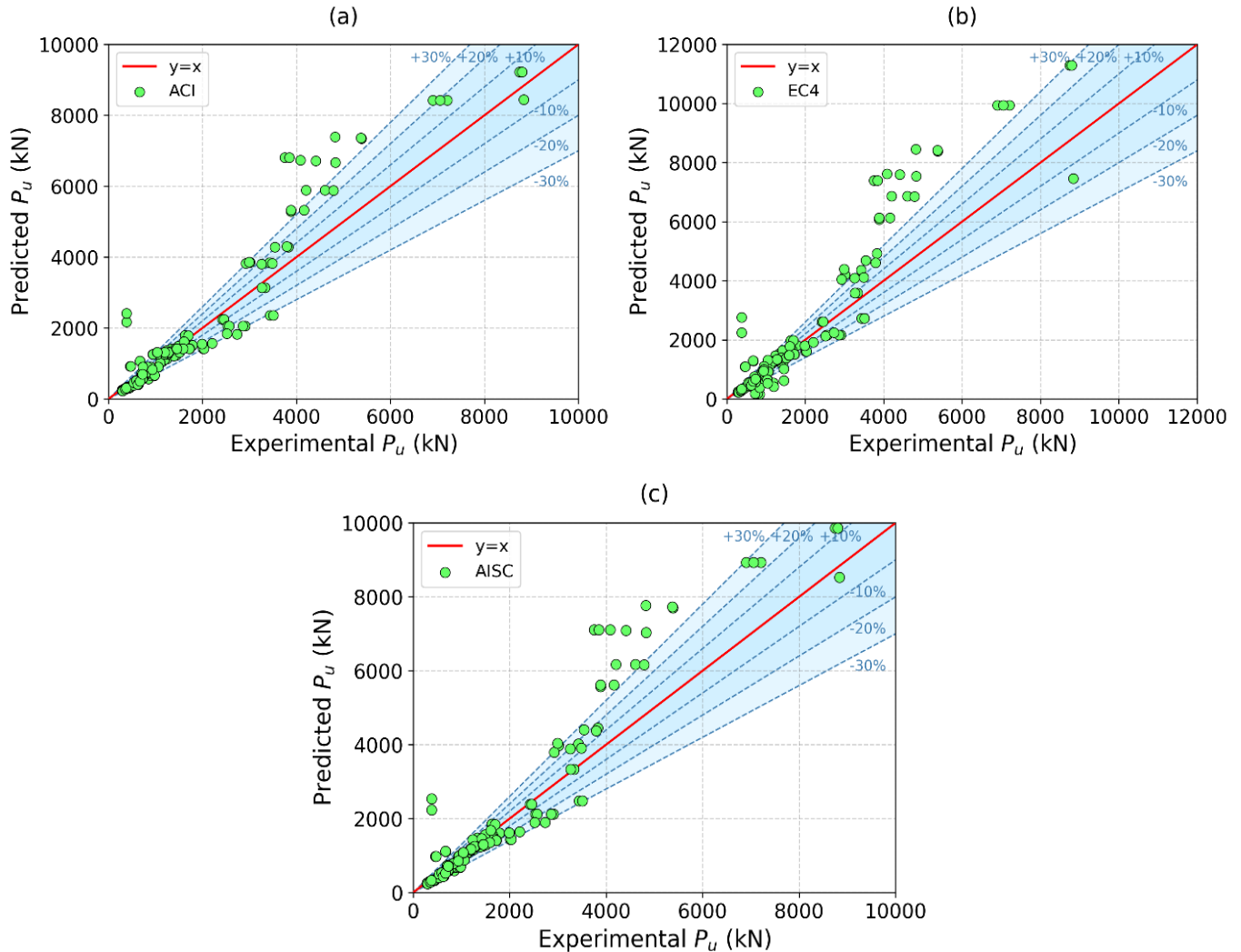


Fig. 6. Regression analysis of the design codes over all dataset: (a) ACI, (b) EC4, and (c) AISC.

3.5. Mechanistic interpretability and feature attribution

Although the optimized XGB-BES framework demonstrates high predictive accuracy, the mathematical complexity of ensemble gradient boosting algorithms results in opaque systems lacking mechanistic transparency. To address the discrepancy between empirical predictive performance and physical structural principles, this study applies post-hoc interpretability frameworks, namely the game-theoretic SHAP and LIME methods. This interpretability analysis is conducted

across two complementary domains. Initially, a global SHAP-based evaluation is utilized to quantify macroscopic feature dominance across the dataset. Subsequently, a detailed, instance-specific local analysis cross-validates the SHAP and LIME outputs to clarify the decision logic governing individual column capacity predictions.

3.5.1. Global feature sensitivity and macroscopic structural influence

The global sensitivity of the predictive framework to the input variables is evaluated using the SHAP summary distribution (Fig. 7). To ensure

out-of-sample validation, this distribution is calculated solely on the independent testing cohort. Advancing beyond standard linear correlation metrics, this game-theoretic method orders the geometric and material parameters based on their mean absolute Shapley values to account for nonlinear structural interactions. Furthermore, the spread of individual data points along the SHAP magnitude axis, combined with their parametric variance, provides a continuous representation of how variations in structural inputs alter the predicted ultimate capacity relative to the baseline expectation of the dataset.

An analysis of the global SHAP attributions indicates that the structural capacity is primarily

determined by the cross-sectional geometry rather than the material yield limits. Specifically, the outer diameter (V_2) and outer tube thickness (V_3) operate as the most influential predictive variables. The positive SHAP values associated with increased V_2 and V_3 dimensions correspond to their respective quadratic and linear proportionality to the cross-sectional area and moment of inertia. Additionally, these geometric parameters establish the hoop stress required for effective concrete confinement. Furthermore, the internal boundary conditions established by the inner tube geometry (V_4 , V_5) exert a secondary but notable effect on the model predictions by restricting inward concrete crushing and reducing local buckling [42,43].

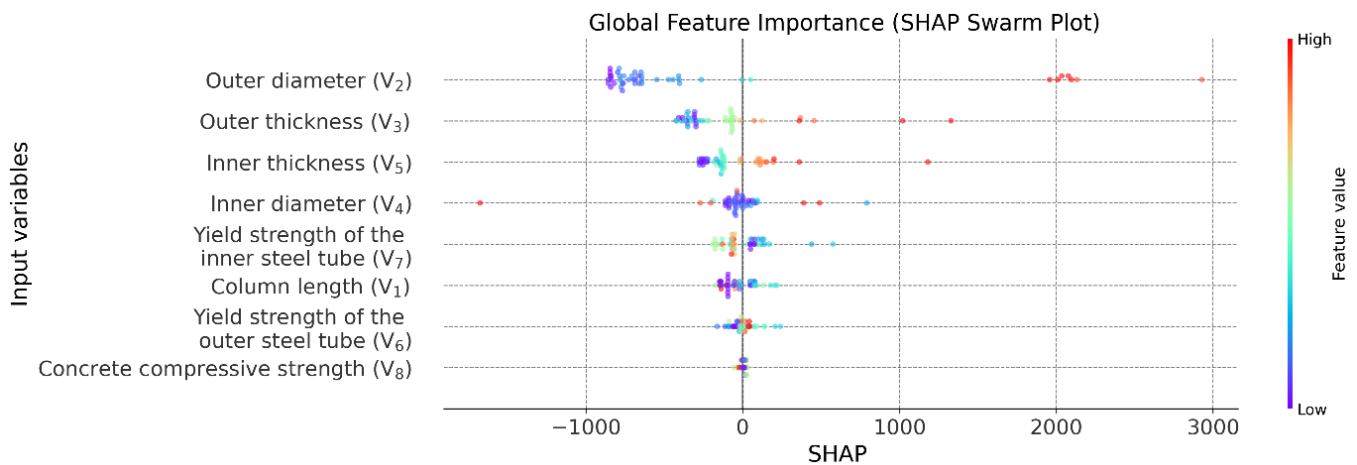


Fig. 7. Global SHAP summary distribution ranking the hierarchical influence and parametric variance of input features on the predicted axial capacity.

The algorithm autonomously modeled the nonlinear structural mechanics associated with column slenderness. Specifically, the SHAP profile for column length (V_1) displays a distinct bifurcation where low length values produce positive contributions that indicate full cross-sectional yielding in short columns, while increased lengths result in substantial negative SHAP values that correspond to the capacity reduction caused by global Euler buckling susceptibility [44,45]. In contrast, the primary material parameters, encompassing the steel yield strengths (V_6 , V_7) and concrete compressive strength (V_8), demonstrated relatively lower global impact magnitudes. This statistical observation does not negate the

structural necessity of these materials but suggests that, within standard construction grade limits, variations in ultimate load-bearing capacity depend primarily upon the dimensional scaling and geometric confinement provided by the steel tubes rather than localized material strengths.

3.5.2. Instance-specific local interpretability and algorithmic cross-validation

Although the global SHAP analysis identifies broad parametric influence, instance-specific local interpretability remains necessary to confirm the physical validity of the model across various structural conditions. To evaluate the explanatory accuracy of both the SHAP and LIME frameworks, three distinct scenarios representing boundary and

central-tendency conditions were selected from the independent testing dataset. These instances comprise a low-capacity configuration where P_u equals 299.0 kN, a high-capacity configuration with P_u at 8800.0 kN, and a representative median-capacity state where P_u is 879.0 kN. The local feature attributions that drive the algorithmic predictions for these specific structural states are compared in Fig. 8 and Fig. 9.

Regarding the low-capacity specimen ($P_u = 299.0$ kN), the SHAP force plot (Fig. 8a) illustrates the negative contributions that reduce the predicted capacity below the baseline expectation of the dataset. The attribution model identifies specific geometric parameters, namely an outer thickness of $V_3 = 0.69$ mm and an inner thickness of $V_5 = 0.61$ mm, as the primary factors contributing to this calculated reduction. The LIME surrogate approximation (Fig. 9a) offers localized validation by assigning the highest negative correlational weights to the threshold conditions $V_3 \leq 2.07$ mm and $V_5 \leq 2.00$ mm. Agreement between the two algorithms indicates that the ML framework associates insufficient wall thickness, a geometric condition that causes premature local buckling and limited concrete core confinement, with the governing failure mechanism for this low-capacity prediction.

In contrast, the local attribution profile for the high-capacity specimen ($P_u = 8800.0$ kN) demonstrates a significant positive deviation from the dataset mean. The SHAP force plot (Fig. 8b) indicates that a large outer diameter ($V_2 = 460.0$ mm) functions as the primary variable for this increased capacity, supported by elevated outer ($V_3 = 5.72$ mm) and inner ($V_5 = 3.75$ mm) steel wall thicknesses. Supporting this physical interpretation, the LIME linear surrogate (Fig. 9b) allocates the highest positive correlational weights to the condition thresholds of $V_2 > 240.00$ mm and $V_3 > 3.50$ mm. This consistent result across both interpretability frameworks verifies that the XGB-BES algorithm correlates large cross-sectional

geometries with high axial stiffness, increased cross-sectional area, and enhanced concrete confinement, aligning with established structural engineering principles.

The median-capacity specimen ($P_u = 879.0$ kN) represents a multivariate structural equilibrium. As the final prediction approaches the global baseline of the dataset, the local attribution profile reflects the interaction of competing physical parameters. The SHAP decomposition (Fig. 8c) indicates that the outer diameter ($V_2 = 160.0$ mm) and inner yield strength ($V_7 = 300.0$ MPa) decrease the predicted capacity, an effect that is subsequently offset by multiple minor positive geometric factors. Furthermore, the LIME surrogate model (Fig. 9c) accounts for this variance by presenting competing correlational weights, such as assigning negative values to the $120.00 < V_2 \leq 165.10$ mm range while identifying positive contributions from particular V_7 limits. This evaluation verifies that the XGB-BES framework avoids direct extrapolation from boundary conditions and calculates the competing mechanical parameters present within standard structural configurations.

The systematic local analyses reveal a diagnostic consensus between the SHAP and LIME frameworks, as both methodologies consistently identify the cross-sectional geometry (V_2 , V_3 , and V_5) as the primary mechanical factors determining ultimate capacity boundaries. This validation across two algorithms is relevant given that the techniques rely on distinct mathematical foundations. Specifically, SHAP applies game-theoretic credit allocation to produce globally consistent, additive feature contributions, while LIME generates a localized, model-agnostic linear surrogate by perturbing the high-dimensional space adjacent to a specific empirical data point. Although LIME establishes interpretable domain threshold rules, such as instances where $V_2 > 240.00$ mm, SHAP supplies an attribution map that connects local predictions to the broader global

sensitivity hierarchy. The agreement between an exact additive explanation method and a local perturbation surrogate indicates that the optimized XGB-BES framework has modeled the nonlinear

principles of structural mechanics rather than merely fitting statistical noise, thereby addressing common interpretability concerns associated with complex ensemble models.

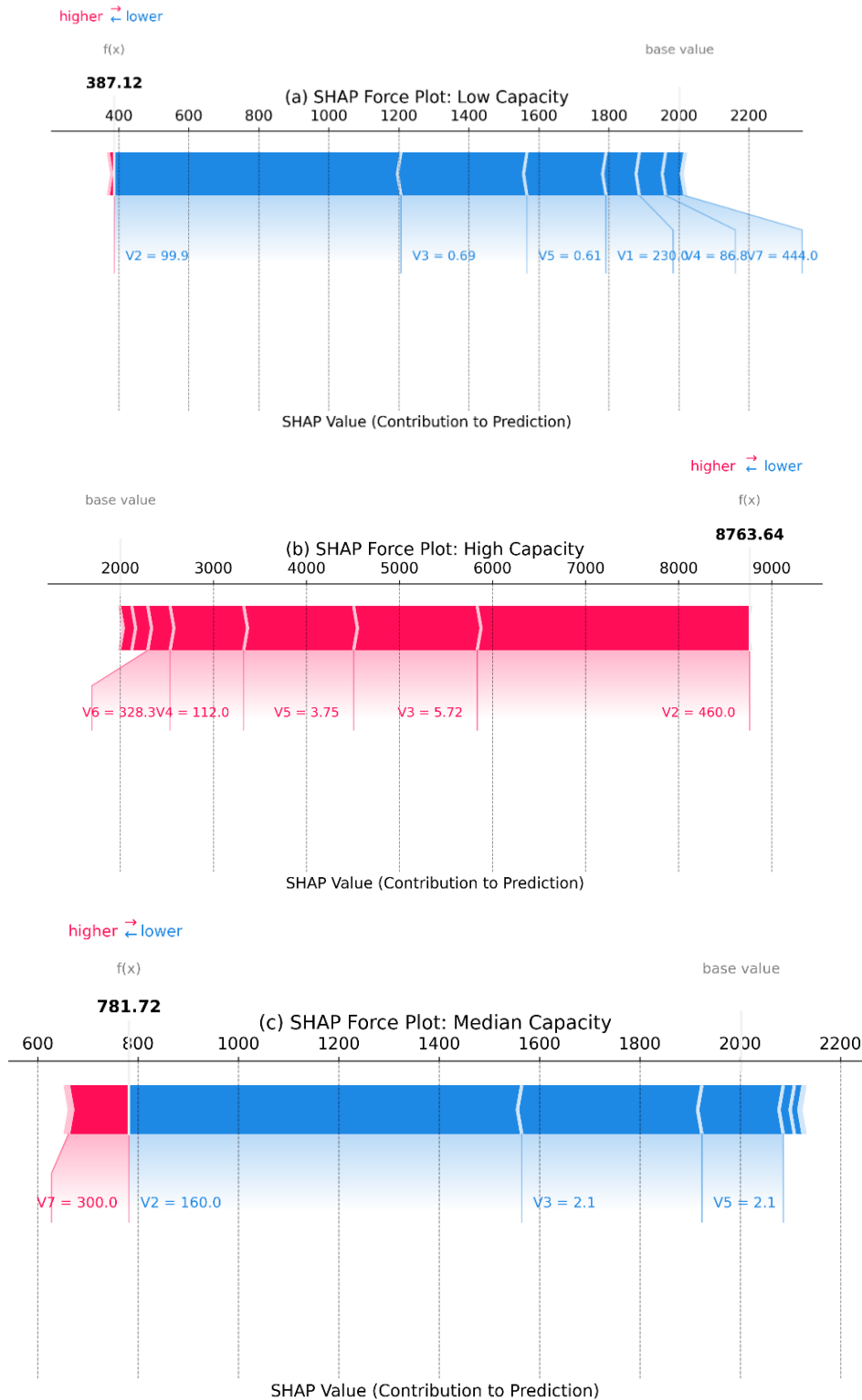


Fig. 8. Instance-specific SHAP force plots detailing localized feature attributions for representative CF DST specimens: (a) severe low-capacity limit, (b) extreme high-capacity boundary, and (c) multivariate median-capacity state.

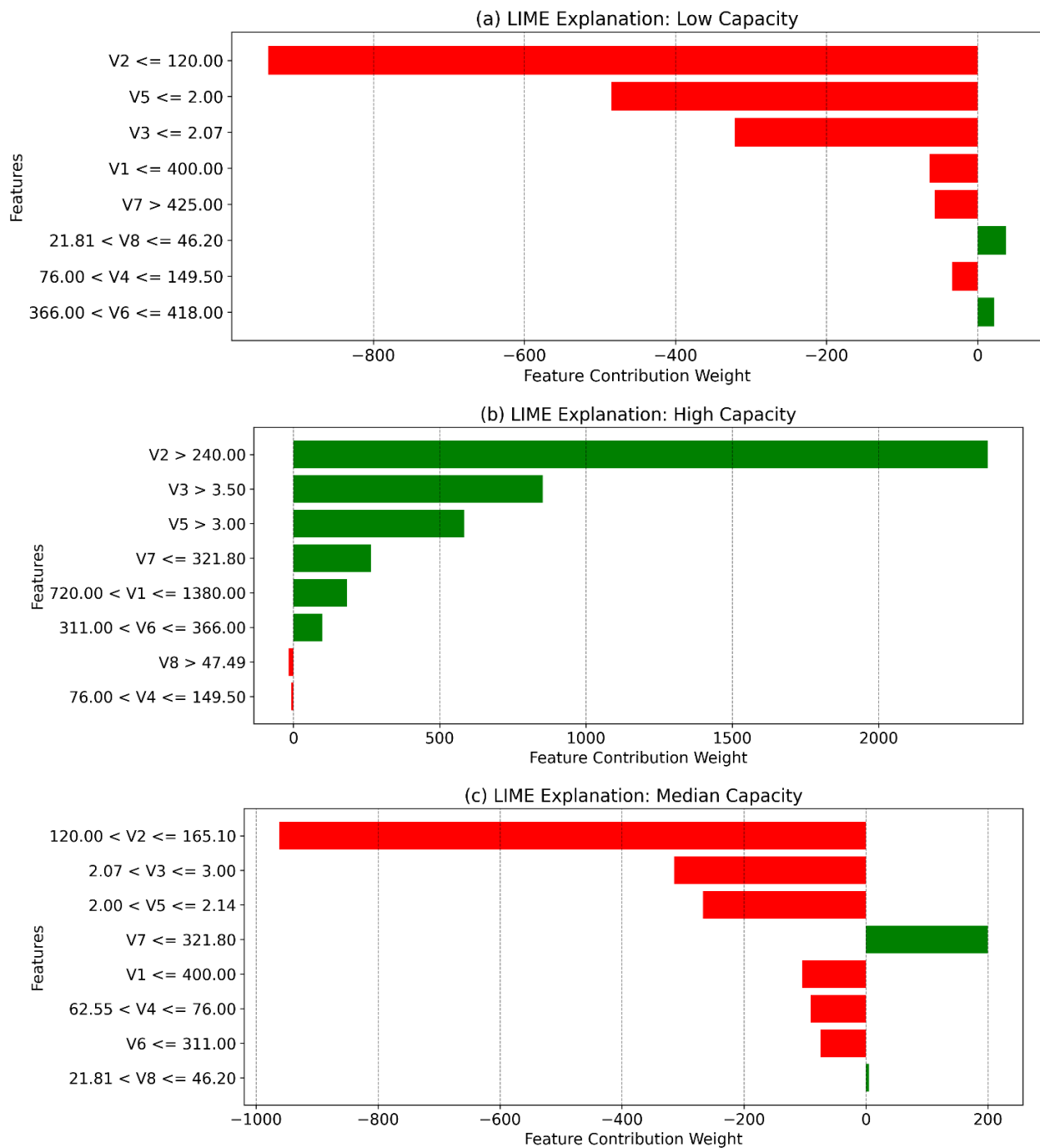


Fig. 9. LIME linear surrogates cross-validating feature contribution weights for the identical representative specimens: (a) severe low-capacity limit, (b) extreme high-capacity boundary, and (c) multivariate median-capacity state.

4. Conclusions

This study develops and evaluates a computational framework designed to estimate the ultimate P_u of CFDST columns. By combining the XGBoost algorithm with metaheuristic optimization techniques, including BES and AO, the research addresses the constraints associated with conventional hyperparameter tuning methods.

Assessment utilizing an empirical database of 153 column specimens indicated the performance capabilities of the XGB-BES framework. The BES-optimized model demonstrated a balance between training data adherence and out-of-sample generalization, resulting in high predictive accuracy on the independent validation subset, characterized by an R^2 of 0.998 and an RMSE of

62.054 kN.

In addition to empirical accuracy, a primary contribution of this work is the application of the BES-driven optimization framework within structural engineering, combined with an evaluation of the algorithm's complex architecture. By implementing a dual-interpretability approach that utilizes SHAP and LIME, this study connects ML outputs with applied physical mechanics. The global SHAP assessment determined the outer tube diameter (V_2) and thickness (V_3) as the principal geometric variables controlling ultimate axial capacity. Simultaneously, localized instance analyses confirmed that the XGB-BES framework aligns with established structural engineering principles by isolating specific confinement parameters (V_2 , V_3 , and V_5) as the main physical factors influencing load-bearing performance. Consequently, this mechanistic verification establishes the proposed hybrid framework as an interpretable and physically consistent computational tool applicable to the design and evaluation of CFDST structures.

Although the proposed XGB-BES framework demonstrates predictive accuracy, its applicability is currently limited to the parametric boundaries of the 153-specimen empirical database used for training and validation. To progress this computational approach, future research should focus on expanding the dataset to include a wider range of material properties and geometric slenderness ratios. Furthermore, subsequent studies could extend the predictive architecture of the model to account for complex structural boundary conditions, such as varying degrees of load eccentricity, and investigate different cross-sectional shapes encompassing elliptical, square, and polygonal profiles. Finally, continually evaluating the XGBoost framework against emerging metaheuristic optimization algorithms provides a logical pathway for the ongoing development of computational structural design tools.

Declaration of Generative AI

During the preparation of this work, the authors used ChatGPT (OpenAI) to assist with language editing and improving the readability of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

- [1] K. Sakino, H. Nakahara, S. Morino, I. Nishiyama. (2004). Behavior of Centrally Loaded Concrete-Filled Steel-Tube Short Columns. *Journal of Structural Engineering*, 130(2), 180–188. [https://doi.org/10.1061/\(ASCE\)0733-9445\(2004\)130:2\(180\)](https://doi.org/10.1061/(ASCE)0733-9445(2004)130:2(180))
- [2] A. Kuranovas, D. Goode, A.K. Kvedaras, S. Zhong. (2009). Load-bearing capacity of concrete-filled steel columns. *Journal of Civil Engineering and Management*, 15(1), 21–33. <https://doi.org/10.3846/1392-3730.2009.15.21-33>
- [3] Y.-H. Wang, G.-B. Lu, X.-H. Zhou. (2018). Experimental study of the cyclic behavior of concrete-filled double skin steel tube columns subjected to pure torsion. *Thin-Walled Structures*, 122, 425–438. <https://doi.org/10.1016/j.tws.2017.10.034>
- [4] L.-H. Han, H. Huang, X.-L. Zhao. (2009). Analytical behaviour of concrete-filled double skin steel tubular (CFDST) beam-columns under cyclic loading. *Thin-Walled Structures*, 47(6–7), 668–680. <https://doi.org/10.1016/j.tws.2008.11.008>
- [5] K. Uenaka, H. Kitoh, K. Sonoda. (2010). Concrete filled double skin circular stub columns under compression. *Thin-Walled Structures*, 48(1), 19–24. <https://doi.org/10.1016/j.tws.2009.08.001>
- [6] H. Lu, L.-H. Han, X.-L. Zhao. (2010). Fire performance of self-consolidating concrete filled double skin steel tubular columns: Experiments. *Fire Safety Journal*, 45(2), 106–

115.
<https://doi.org/10.1016/j.firesaf.2009.12.001>
- [7] M. Pagoulatou, T. Sheehan, X.H. Dai, D. Lam. (2014). Finite element analysis on the capacity of circular concrete-filled double-skin steel tubular (CFDST) stub columns. *Engineering Structures*, 72, 102–112. <https://doi.org/10.1016/j.engstruct.2014.04.039>
- [8] M. Elchalakani, X.-L. Zhao, R. Grzebieta. (2002). Tests on concrete filled double-skin (CHS outer and SHS inner) composite short columns under axial compression. *Thin-Walled Structures*, 40(5), 415–441. [https://doi.org/10.1016/S0263-8231\(02\)00009-5](https://doi.org/10.1016/S0263-8231(02)00009-5)
- [9] Z. Tao, L.-H. Han, X.-L. Zhao. (2004). Behaviour of concrete-filled double skin (CHS inner and CHS outer) steel tubular stub columns and beam-columns. *Journal of Constructional Steel Research*, 60(8), 1129–1158. <https://doi.org/10.1016/j.jcsr.2003.11.008>
- [10] W. Li, L.-H. Han, Q.-X. Ren, X.-L. Zhao. (2013). Behavior and calculation of tapered CFDST columns under eccentric compression. *Journal of Constructional Steel Research*, 83, 127–136. <https://doi.org/10.1016/j.jcsr.2013.01.010>
- [11] W. Li, Q.-X. Ren, L.-H. Han, X.-L. Zhao. (2012). Behaviour of tapered concrete-filled double skin steel tubular (CFDST) stub columns. *Thin-Walled Structures*, 57, 37–48. <https://doi.org/10.1016/j.tws.2012.03.019>
- [12] L.-H. Han, Q.-X. Ren, W. Li. (2011). Tests on stub stainless steel–concrete–carbon steel double-skin tubular (DST) columns. *Journal of Constructional Steel Research*, 67(3), 437–452. <https://doi.org/10.1016/j.jcsr.2010.09.010>
- [13] J. Wang, W. Liu, D. Zhou, L. Zhu, H. Fang. (2014). Mechanical behaviour of concrete filled double skin steel tubular stub columns confined by FRP under axial compression. *Steel and Composite Structures*, 17(4), 431–452. <https://doi.org/10.12989/scs.2014.17.4.431>
- [14] S. İpek, E.M. Güneyisi. (2019). Ultimate Axial Strength of Concrete-Filled Double Skin Steel Tubular Column Sections. *Advances in Civil Engineering*, 2019, 6493037. <https://doi.org/10.1155/2019/6493037>
- [15] F.A. Sofi, H. Wani, M. Zakir, M.A. Tantray. (2022). Estimating Axial Load Capacity of Concrete-Filled Double-Skin Steel Tubular Columns of Multiple Shapes Using Genetic Algorithm-Optimized Artificial Neural Networks. in: V. Kushvaha, M.R. Sanjay, P. Madhushri, S. Siengchin (Eds.), *Machine Learning Applied to Composite Materials*, Springer, Singapore, 2022: pp. 167–198. https://doi.org/10.1007/978-981-19-6278-3_9
- [16] S. İpek, E.M. Güneyisi, K. Mermerdaş, Z. Algin. (2021). Optimization and modeling of axial strength of concrete-filled double skin steel tubular columns using response surface and neural-network methods. *Journal of Building Engineering*, 43, 103128. <https://doi.org/10.1016/j.jobe.2021.103128>
- [17] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, K. Chen, R. Mitchell, I. Cano, T. Zhou. (2015). Xgboost: extreme gradient boosting, R Package Version 0.4-2 1 (2015) 1–4.
- [18] L. Abualigah, D. Yousri, M. Abd Elaziz, A.A. Ewees, M.A.A. Al-Qaness, A.H. Gandomi. (2021). Aquila optimizer: a novel meta-heuristic optimization algorithm. *Computers & Industrial Engineering*, 157, 107250. <https://doi.org/10.1016/j.cie.2021.107250>
- [19] H.A. Alsattar, A.A. Zaidan, B.B. Zaidan. (2020). Novel meta-heuristic bald eagle search optimisation algorithm. *Artificial Intelligence Review*, 53, 2237–2264. <https://doi.org/10.1007/s10462-019-09732-5>
- [20] Deepika, G. Pandove. (2025). Prediction of traffic time using XGBoost model with hyperparameter optimization. *Multimedia Tools and Applications*, 84, 37045–37090. <https://doi.org/10.1007/s11042-025-20646-z>

- [21] A.E. Roth. (1988). The Shapley Value: Essays in Honor of Lloyd S. Shapley, edited by A.E. Roth, 1988, pp. 1–28. *Cambridge University Press*.
- [22] P. Biecek, T. Burzykowski. (2021). Local interpretable model-agnostic explanations (LIME). Explanatory Model Analysis, 1st Edition, pp. 107–123. *Chapman and Hall/CRC*. <https://doi.org/10.1201/9780429027192-11>
- [23] Y. Essopjee, M. Dundu. (2015). Performance of concrete-filled double-skin circular tubes in compression. *Composite Structures*, 133, 1276–1283. <https://doi.org/10.1016/j.compstruct.2015.08.033>
- [24] J.Y. Chen, Y. Hai. (2010). Research on bearing capacity of short concrete filled double skin steel tubes columns under axial compression. *Advanced Materials Research*, 168-170, 2154–2157. <https://doi.org/10.4028/www.scientific.net/AMR.168-170.2154>
- [25] Z.K. Duan, R. Wang. (2014). Research of the stainless steel-concrete-carbon steel circular concrete-filled double skin steel tubes under axial compression. *Advanced Materials Research*, 1065–1069, 1349–1353. <https://doi.org/10.4028/www.scientific.net/AMR.1065-1069.1349>
- [26] J.S. Fan, M.N. Baig, J.G. Nie. (2008). Test and analysis on double-skin concrete filled tubular columns. *Proceedings of Tubular Structures XII, 2008*, 407–411. 2009 Taylor & Francis Group, London, ISBN 978-0-415-46853-4.
- [27] L.-H. Han, Y.-J. Li, F.-Y. Liao. (2011). Concrete-filled double skin steel tubular (CFDST) columns subjected to long-term sustained loading. *Thin-Walled Structures*, 49(12), 1534–1543. <https://doi.org/10.1016/j.tws.2011.08.001>
- [28] W. Li, D. Wang, L.-H. Han. (2017). Behaviour of grout-filled double skin steel tubes under compression and bending: Experiments. *Thin-Walled Structures*, 116, 307–319. <https://doi.org/10.1016/j.tws.2017.02.029>
- [29] S. Wei, S.T. Mau, C. Vipulanandan, S.K. Mantrala. (1995). Performance of New Sandwich Tube under Axial Loading: Experiment. *Journal of Structural Engineering*, 121(12), 1806–1814. [https://doi.org/10.1061/\(ASCE\)0733-9445\(1995\)121:12\(1806\)](https://doi.org/10.1061/(ASCE)0733-9445(1995)121:12(1806))
- [30] Y.-F. Yang, L.-H. Han, B.-H. Sun. (2012). Experimental behaviour of partially loaded concrete filled double-skin steel tube (CFDST) sections. *Journal of Constructional Steel Research*, 71, 63–73. <https://doi.org/10.1016/j.jcsr.2011.11.005>
- [31] X.L. Zhao, R. Grzebieta, M. Elchalakani. (2002). Tests of concrete-filled double skin CHS composite stub columns. *Steel and Composite Structures*, 2(2), 129–146. <https://doi.org/10.12989/scs.2002.2.2.129>
- [32] I. Faridmehr, M.L. Nehdi, Predicting axial load capacity of CFST columns using machine learning, *Structural Concrete* 23(3) (2022) 1642–1658. <https://doi.org/10.1002/suco.202100641>.
- [33] Y. Ren, H.F. Isleem, W.J.K. Almoghaye, A.K. Hamed, P. Jangir, Arpita, G.G. Tejani, A.E. Ezugwu, A.A. Soliman. (2025). Machine learning-based prediction of elliptical double steel columns under compression loading. *Journal of Big Data*, 12, 50. <https://doi.org/10.1186/s40537-025-01081-1>
- [34] T.-A. Nguyen, K. Le Nguyen, H.-B. Ly. (2024). Universal boosting ML approaches to predict the ultimate load capacity of CFST columns. *The Structural Design of Tall and Special Buildings*, 33(2), e2071. <https://doi.org/10.1002/tal.2071>
- [35] H.-B. Ly. (2025). Enhanced Predictive Modeling and Insights into Geopolymer Concrete Compressive Strength Prediction. *Journal of Testing and Evaluation*, 53(4), 938–

959. <https://doi.org/10.1520/JTE20230808>
- [36] H.-B. Ly. (2026). An interpretable machine learning-based measurement system for predicting the compressive strength of concrete at elevated temperatures. *Measurement*, 264, 120273. <https://doi.org/10.1016/j.measurement.2025.120273>
- [37] E.H. Houssein, M.H.A. Gafar, N. Fawzy, A.Y. Sayed. (2025). Recent metaheuristic algorithms for solving some civil engineering optimization problems. *Scientific Reports*, 15, 7929. <https://doi.org/10.1038/s41598-025-90000-8>
- [38] M. Varshney, P. Kumar, M. Ali, Y. Gulzar. (2024). Using the grey wolf aquila synergistic algorithm for design problems in structural engineering. *Biomimetics*, 9(1), 54. <https://doi.org/10.3390/biomimetics9010054>
- [39] ACI. (2011). ACI 318M-11, Building code requirements for structural concrete. *American Concrete Institute*.
- [40] European Standard. (1994). EN 1994-Eurocode 4: Design of composite steel and concrete structures.
- [41] American Institute of Steel Construction (AISC). (2016). Specification for Structural Steel Buildings, ANSI/AISC 360-16. An American National Standard.
- [42] Q. Qiyun, P. Jia, C. Wanlin, W. Haipeng. (2024). Behavior of concrete-filled double steel tube stub columns under partial compression. *Journal of Constructional Steel Research*, 214, 108451. <https://doi.org/10.1016/j.jcsr.2024.108451>
- [43] J. Ci, H. Jia, S. Chen, W. Yan, T. Song, K.-S. Kim. (2020). Performance analysis and bearing capacity calculation on circular concrete-filled double steel tubular stub columns under axial compression. *Structures*, 25, 554–565. <https://doi.org/10.1016/j.istruc.2020.03.038>
- [44] J.-L. Guo, S. Pan, X. Guo, Z.-Y. Wu. (2023). Numerical simulation of the dynamic responses and impact-bearing capacity of CFDST columns under lateral impact. *Buildings*, 13(3), 805. <https://doi.org/10.3390/buildings13030805>
- [45] X. Liu, S. Liu, Y. Hui, B. Wang, (2022). Mechanical properties of CFDST column under eccentric compression after exposed to fire. *Structures*, 43, 1200–1215. <https://doi.org/10.1016/j.istruc.2022.07.036>