Sustainable Structures
ISSN: 2789-3111 (Print); ISSN: 2789-312X (Online)
http://www.sustain-dpl.com/journal/sust
DOI: 10.54113/j.sust.2026.000112

ORIGINAL ARTICLE

Bayesian optimization-enhanced machine learning axial load prediction across diverse CFST columns, strengths, geometries, and slenderness

Khalid Saqer Alotaibi *

Department of Civil and Construction Engineering, College of Engineering, Imam Abdulrahman Bin Faisal University, Dammam, 31451, Saudi Arabia
*Corresponding Author: Dr. Khalid Saqer Alotaibi. Email: kalotaibi@iau.edu.sa

Abstract: Concrete-filled steel tubes (CFST) are recognized for their superior structural performance, offering enhanced strength, ductility, and construction efficiency. This enhanced performance is primarily due to the synergistic relationship between the steel casing and the concrete core; the steel imparts confinement to the concrete, thereby augmenting its compressive strength and ductility, while the concrete infill simultaneously restrains the steel tube against local buckling. This study addresses the critical need for accurate prediction of axial compression capacity in diverse CFST columns, spanning both circular and rectangular cross-sections, and encompassing short to slender configurations. A key focus was accommodating a wide range of material strengths, from conventional normal concrete and steel to contemporary ultra-high-performance concrete and high-strength steel. To achieve this, advanced machine learning (ML) algorithms specifically, ExtraTrees, XGBoost, and GradientBoosting, were employed. The hyperparameters of these models were optimized using Bayesian Optimization to maximize predictive efficacy. The developed models were rigorously validated through 10-fold cross-validation, demonstrating high accuracy, with coefficients of determination (R^2) achieving up to 0.99 in predicting CFST column failure loads. Comprehensive error and sensitivity analyses were conducted to thoroughly assess their predictive capabilities and robustness. Furthermore, SHapley Additive exPlanations (SHAP) analysis was utilized to elucidate model decision-making processes and identify the most influential input parameters, enhancing interpretability and confidence. This research significantly advances the efficient, reliable, and data-informed design of CFST structures, expanding their applicability and fostering greater industry acceptance, particularly for innovative designs utilizing high-performance materials.

Keywords: concrete-filled steel tubes, axial compression capacity, machine learning, bayesian optimization, explainable AI, slenderness

1 Introduction

Concrete-filled steel tubes (CFST) are efficient composite structural systems, leveraging a synergistic mechanical interaction between the steel tube and concrete core [1-3]. This synergy offers superior structural performance, including enhanced strength, ductility, and energy absorption compared to conventional steel or reinforced concrete members [4]. The outer steel tube confines the concrete core, boosting its compressive strength and ductility, while the concrete infill restrains the steel

Received: 24 December 2025; Received in revised form: 4 July 2026; Accepted: 8 July 2026

tube against inward local buckling, enhancing stability [5]. A practical advantage is the steel tube's role as permanent formwork, accelerating construction and reducing project costs [6]. Foundational CFST research, such as that early conducted by Klöppel and Goder (1957) [7], and comprehensive programs, such as summarized by Nishiyama et al. (2002) [8], have solidified this understanding. Due to benefits like enhanced fire resistance [9], CFST columns are widely used in high-rise buildings, long-span bridges [10], and other critical infrastructure.

Accurate prediction of CFST column axial load-carrying capacity is vital for safe, reliable, and optimized design [11]. This capacity is governed by numerous parameters, including column geometry (cross-sectional shape, principal dimensions, steel tube slenderness via width-to-thickness (b/t) or diameter-to-thickness (D/t) ratios, material properties like concrete compressive strength (f_c') and steel yield strength (f_y), and overall column slenderness (length-to-depth (L/D or L/B) ratios), which distinguishes short from slender column behavior [6, 12-16]. The trend towards advanced materials like high-strength steel (HSS) [17] and high-performance/ultra-high-performance concrete (HPC/UHPC) [2, 3, 17-20] complicates mechanical response prediction, as these enable slender, efficient members but may alter failure modes, a possibility suggested by early work [21].

Historically, CFST axial capacity prediction relied on several approaches. Experimental investigations have been fundamental [13, 22-24], but are costly, time-consuming, and often limited in parametric scope. Finite Element Analysis (FEA) allows parametric studies beyond experimental limits [18, 25-27], yet its accuracy depends on sophisticated material models [28], steel-concrete interface modeling, and it requires significant computational resources. Consequently, design codes like AISC 360 [29], Eurocode 4 [30], and AS/NZS 2327 [31] provide empirical or semi-empirical equations, which, while standardized, can be conservative with discrepancies between codes [32].

A critical limitation of many codes is restricted applicability or reduced accuracy for CFSTs with high-strength materials or slender sections [3]. For instance, a previous study [15] noted shortcomings of earlier AISC versions [33] for noncompact/slender rectangular CFSTs, and Li et al. (2024) [3] state current codes do not formally endorse ultra-high strength CFSTs (defined as $f_y \geq 690$ MPa and $f_c' \geq 120$ MPa). In response, machine learning (ML) techniques have gained traction [34-39]. Data-driven ML models can learn complex, non-linear relationships from experimental datasets, offering more accurate and versatile predictive tools [40], addressing experimental costs, the computational intensity of FEA, and code conservatism. Recent ML applications to CFSTs include Artificial Neural Networks (ANNs) for FRP-confined CFSTs [41], Gaussian Process Regression (GPR), and Extreme Gradient Boosting (XGBoost) for rectangular CFSTs [11], Support Vector Regression (SVR) with Particle Swarm Optimization (PSO) [40], and early ANNs [16, 42].

Despite these advancements, more robust and comprehensive ML models are needed for accurately predicting axial capacity across diverse column characteristics: geometric configurations, full slenderness spectrum, and a wide range of material strengths including ultra-high-performance concrete and high-strength steel, where codes often lack validation. While advanced ML algorithms like XGBoost [11] are effective, systematic hyperparameter optimization is crucial; this study employs Bayesian Optimization [43]. Furthermore, high accuracy alone is insufficient for engineering acceptance; model interpretability via Explainable AI (XAI) techniques like SHapley Additive exPlanations (SHAP) [44] and robustness assessment via sensitivity analysis are essential. Existing literature has not comprehensively addressed this broad material/geometric scope integrated with systematic hyperparameter optimization and in-depth model interpretation for CFST axial load prediction.

This research, therefore, aims to develop highly accurate and interpretable ML-based predictive models for the axial compression capacity of diverse CFST columns, including circular and rectangular, short and slender configurations, and a wide range of material strengths from conventional to ultra-high-performance. The key innovations of this study are threefold: (1) the utilization of a comprehensive database that includes contemporary ultra-high-strength materials often excluded from older models; (2) the systematic application and comparison of three distinct, advanced ensemble models (ExtraTrees, XGBoost, GradientBoosting), all optimized via a rigorous Bayesian Optimization framework to ensure peak performance; and (3) a dual-pronged interpretability analysis using both SHAP for local and global feature attribution and Sobol' sensitivity analysis to explicitly quantify parameter interactions, providing a deeper and more robust understanding of the underlying mechanics than previously achieved.

2 Experimental Database

A comprehensive, high-quality experimental database is an indispensable prerequisite for constructing robust and accurate ML models capable of predicting the axial compression capacity of CFST columns. This section details the assembly and characteristics of the database utilized in the present study, which forms the empirical foundation for training and validating the predictive models.

2.1 Source and Scope of the Compiled Database

The experimental database for this research is primarily built upon the extensive collection of test results for CFST members meticulously compiled and reported by Li et al. (2024) [3]. Their work provides a wide-ranging dataset covering various CFST specimen types and loading conditions. For the specific focus of this study on axial compression behavior, data pertaining to CFST columns subjected to axial loading were extracted.

The comprehensive database from Li et al. (2024) [3] on axially loaded CFST members includes a total of 532 test specimens. These specimens are systematically categorized in their supplementary materials, detailing 318 tests on circular CFST short columns, 139 tests on rectangular CFST short columns, 24 tests on circular CFST long columns, and 51 tests on rectangular CFST long columns, all subjected to axial compression.

A key feature of the source database is its emphasis on ultra-high strength materials, with steel yield strengths (f_y) reaching up to 1233 MPa and concrete compressive strengths (f_c') up to 193.3 MPa, alongside conventional strength materials. The database encompasses critical parameters for each specimen, including geometric properties such as member length (L), cross-sectional dimensions (diameter D for circular sections; width B and depth H for rectangular sections), and steel tube thickness (t). Additionally, material properties like steel yield strength (f_y) and concrete compressive cylinder strength (f_c') are provided. Finally, the experimental results for each specimen include the ultimate axial load capacity ($P_{\exp}$).

The curation process by Li et al. (2024) [3] involved specific exclusion criteria to ensure data quality and relevance, such as the omission of specimens using stainless steel tubes, those with greased inner steel surfaces, members loaded only on the steel tube or concrete core, and tests involving preloading or dynamic/long-term loading effects. Furthermore, for consistency in concrete strength reporting, f_c' values derived from non-standard cylinder sizes or cube/prism tests were converted to the standard 150 mm × 300 mm cylinder strength using established conversion factors.

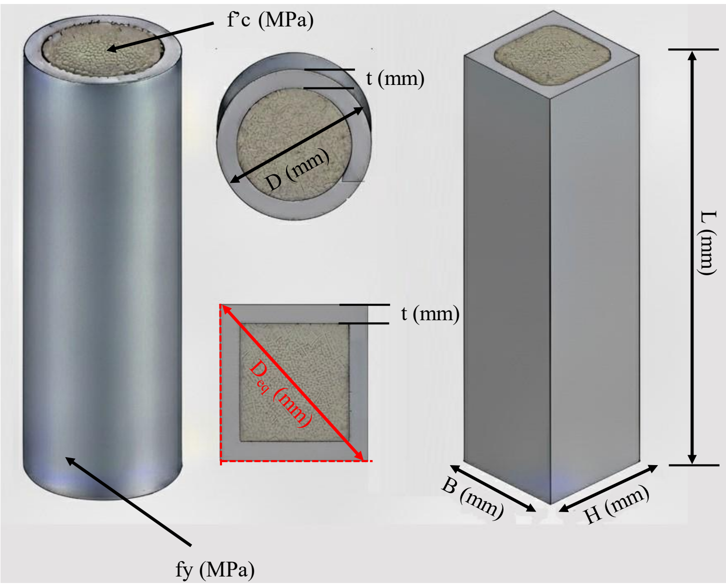

Fig. 1. Typical cross-sections and defining geometric parameters for the CFST columns.

2.2 Processed Database for Current Study

The current machine learning investigation utilized a carefully compiled dataset derived from the previously mentioned sources, which consists of 532 CFST column experiments conducted under axial compression. This processed dataset was specifically selected to ensure the completeness of all input features required for the ML models and to represent a diverse range of CFST column configurations. The dataset includes both circular and rectangular cross-sections, encompassing short columns, where failure is primarily governed by material strength, and long (slender) columns, where stability effects are significant. A schematic diagram illustrating these column types and defining their key geometric parameters is provided in **Fig. 1.** Specimens are sub-categorized as either "short columns" or "long columns." Short columns are defined as circular CFST columns with a length-to-diameter ratio (L/D) ≤ 4 or rectangular CFST columns with a length-to-depth ratio (L/H) ≤ 4. Conversely, long columns are identified by L/D > 4 for circular sections or L/H > 4 for rectangular sections. Employing a systematic classification founded on slenderness ratios enables a more detailed performance assessment, recognizing that cross-sectional strength is the principal determinant for short columns, whereas stability and buckling phenomena govern the behavior of long columns.

To standardize geometric input for rectangular sections, an equivalent diameter (D_{eq}) was calculated based on the principle of equal cross-sectional area. For the purpose of determining column slenderness, the methodology previously described by Li et al. (2024) [3] is maintained. For subsequent calculations involving D/t and L/D, the equivalent diameter (D_{eq}) is utilized for rectangular sections, effectively making it interchangeable with the actual diameter (D) for circular sections in these specific contexts, allowing for a unified geometric feature, and this new calculation is not utilized in the initial classification of slenderness. This approach allows for a unified representation of both circular and rectangular specimens within the model.

The total cross-sectional area of a rectangular CFST member is determined by multiplying its width (B) by its depth (H). For an equivalent circular CFST member, its total cross-sectional area is calculated using the formula for the area of a circle, incorporating its equivalent diameter. By setting these two area calculations equal to each other, this establishes a relationship between the dimensions of the rectangular section and the equivalent diameter of the circular section. Consequently, by rearranging this relationship, the equivalent diameter can be directly determined from the width and depth of the rectangular section.

The equivalent diameter (D_{eq}) can be derived by solving the following equation:

$$D_{\mathrm{eq}} = \sqrt{4 \times B \times H / \pi} \tag{1}$$

A statistical summary of the key input parameters for these 532 specimens is presented in **Table 1**. This table highlights the wide variation in material strengths and geometric properties, ensuring the developed ML models are trained on a dataset reflective of diverse practical scenarios, including normal-strength to ultra-high-performance concrete and high-strength steel.

Table 1 clearly demonstrates a wide range of values observed for each parameter, underscoring the comprehensive nature of the compiled experimental database. This broad spectrum of input features is crucial for developing ML models that are generalizable and robust across diverse CFST column configurations and material properties encountered in practical engineering applications. For instance, the length (L) varies from 210 mm to 4195 mm, encompassing both short columns where material strength primarily governs failure, and slender columns where buckling phenomena become dominant. Similarly, the diameter (D) or equivalent diameter (D_{eq}) ranges from 75.8 mm to 359 mm, and the steel tube thickness (t) from 1 mm to 18.5 mm. These geometric variations lead to significant differences in the diameter-to-thickness (D/t) ratios (6.1 to 130) and length-to-diameter (L/D) ratios (2.2 to 37.08), which are critical parameters influencing the confinement effect and overall column stability. Furthermore, the material properties exhibit an impressive range, with steel yield strengths (f_y) spanning from 236 MPa to 1233 MPa and concrete compressive strengths (f_c') from 20 MPa to 193.3 MPa. This includes conventional normal-strength materials as well as high-strength steel and ultra-high-performance concrete, which are increasingly adopted in modern construction for their enhanced structural efficiency. The considerable variation in the experimental ultimate axial load capacity, from

410 kN to 20,900 kN, directly reflects the influence of these diverse input parameters. Training ML models on such a rich and varied dataset is essential to avoid overfitting to a narrow range of conditions and to ensure that the developed predictive tools can reliably estimate the axial capacity of CFST columns across their full practical spectrum, addressing a key limitation of many existing empirical design codes that often have restricted applicability to high-strength materials or slender sections.

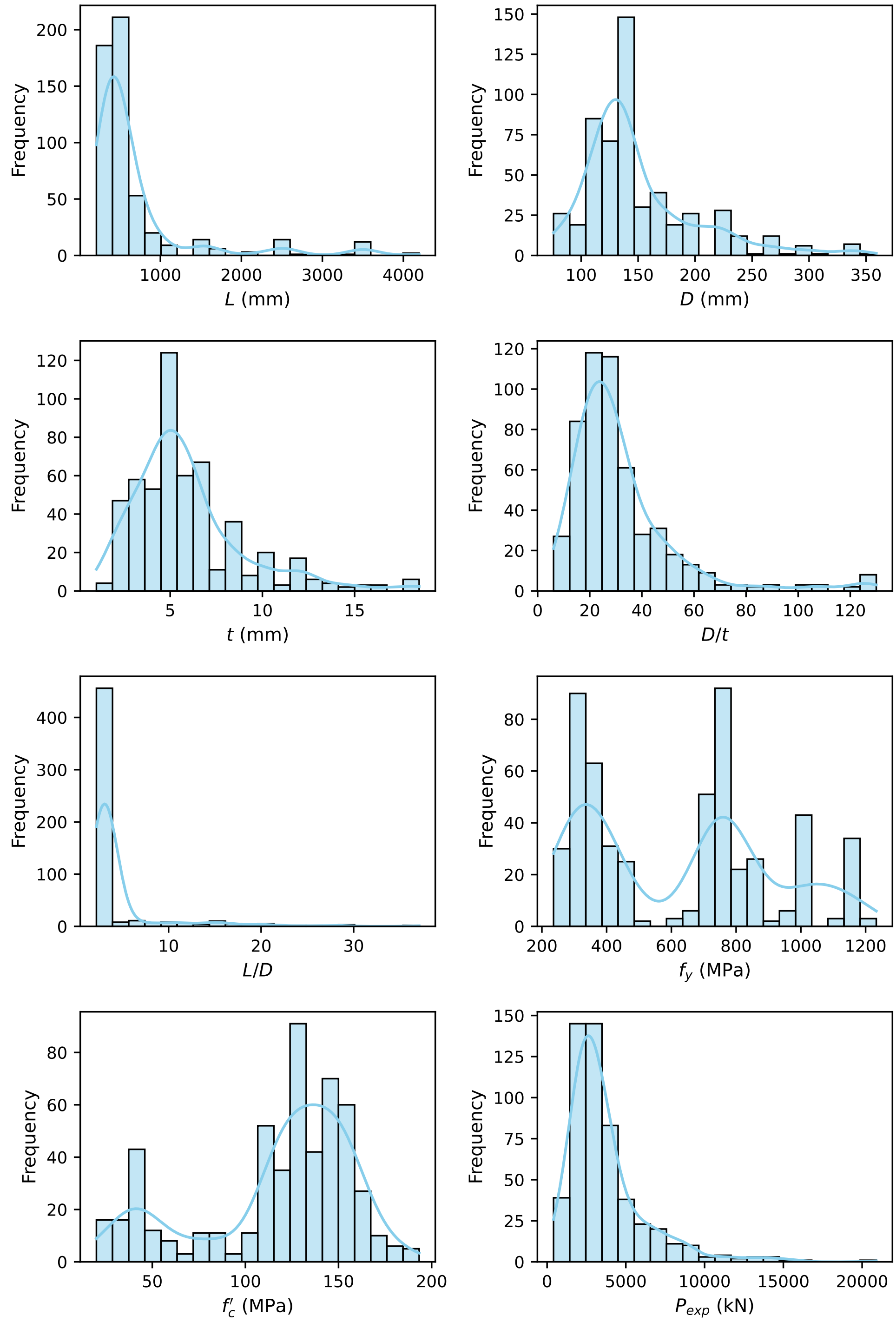

Fig. 2. Histogram plots of input and targeted output parameter distributions.

The histogram plots presented in **Fig. 2** provide valuable insights into the distribution characteristics of the input and output parameters within the processed dataset. These visualizations are critical for understanding the underlying data structure and identifying potential biases or concentrations within specific ranges, which can influence the model learning behavior and convergence patterns of ML models. For instance, the distribution of L and L/D shows a notable concentration towards shorter values, indicating a larger proportion of short columns in the dataset. Conversely, the distribution of steel yield strength (f_y) and concrete compressive strength (f_c') appears to span across both lower and higher strength ranges, with some distinct peaks, particularly for f_y, suggesting the inclusion of both conventional and high-performance materials. The distribution of the target variable, P_{exp}, provides an understanding of the range and frequency of the ultimate loads observed in the experiments.

Table 1. Statistical summary of key parameters in the compiled experimental database

Parameters	L (mm)	D (mm)	t (mm)	D/t	L/D	f_y (MPa)	f_c' (MPa)	P_{exp} (kN)
count	532	532	532	532	532	532	532	532
Mean	669.32	149.91	5.91	32.00	4.53	624.37	117.58	3706.53
Std.	666.67	50.37	3.08	20.92	4.59	284.22	41.55	2540.54
Min	210	75.8	1	6.1	2.2	236	20	410
25th Percentile	375	120	3.9775	19.7	2.7	343	106.3	2168.95
50th Percentile	447.5	136.08	5	26.7	3	704	129.1	3004.95
75th Percentile	609	169.26	6.5	36.335	3.5	793	147.3	4302
Max	4195	359	18.5	130	37.08	1233	193.3	20900

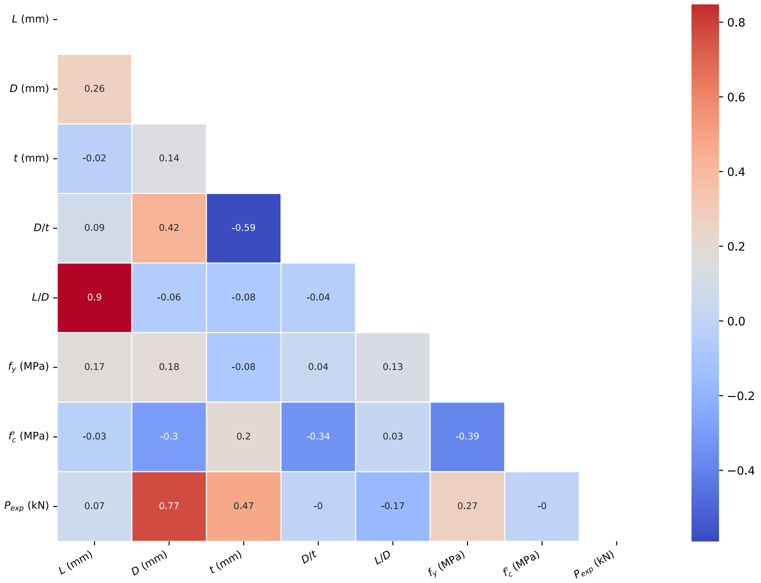

Fig. 3. Pearson Correlation Coefficients among features and target variable in the dataset.

To further understand the interdependencies among the features and their relationship with the target variable (P_{exp}), a correlation analysis was performed. **Fig. 3** presents a heatmap illustrating the Pearson correlation coefficients between features and the target. This visualization aids in identifying potentially strong linear relationships that might influence the ML model's learning process. In statistical analysis, the linear relationship between multiple variables is quantified using the correlation coefficient. This coefficient ranges from -1 to 1, with a value of 0 indicating no correlation between the two characteristics under examination. Pearson's method provides a means to calculate this correlation coefficient using the following formula:

$$Pearson's\ coefficient = \frac{\sigma_{xy}}{\sigma_x \times \sigma_x} = \frac{\sum_i^n (X_i - \bar{X}) \times (Y_i - \bar{Y})}{\sqrt{\sum_i^n (X_i - \bar{X})^2} \times \sqrt{\sum_i^n (X_i - \bar{Y})^2}} \tag{2}$$

In the given equation, y_i denotes the i-th experimental value and x_i denotes the i-th value of a second feature. The symbols $\bar{y}$ and $\bar{x}$ represent the mean values of their respective features over all n observations.

From **Fig. 3**, it is evident that D has a very strong positive correlation with P_{exp} (0.77), indicating that larger column diameters generally correspond to higher axial load capacities. This aligns with fundamental structural mechanical principles. Similarly, the thickness also exhibits a substantial positive correlation with P_{exp} (0.47). The L/D shows a negative correlation with P_{exp} (-0.17), confirming that increasing slenderness tends to reduce axial capacity due to stability effects. While f_y shows a positive correlation (0.27), f_c' shows a negligible correlation (approximately zero). These linear correlation coefficients only capture linear relationships. Given the complex synergistic behavior of CFST columns, it is anticipated that non-linear interactions among parameters, which ML models are adept at learning, will play a more significant role in determining the ultimate axial capacity.

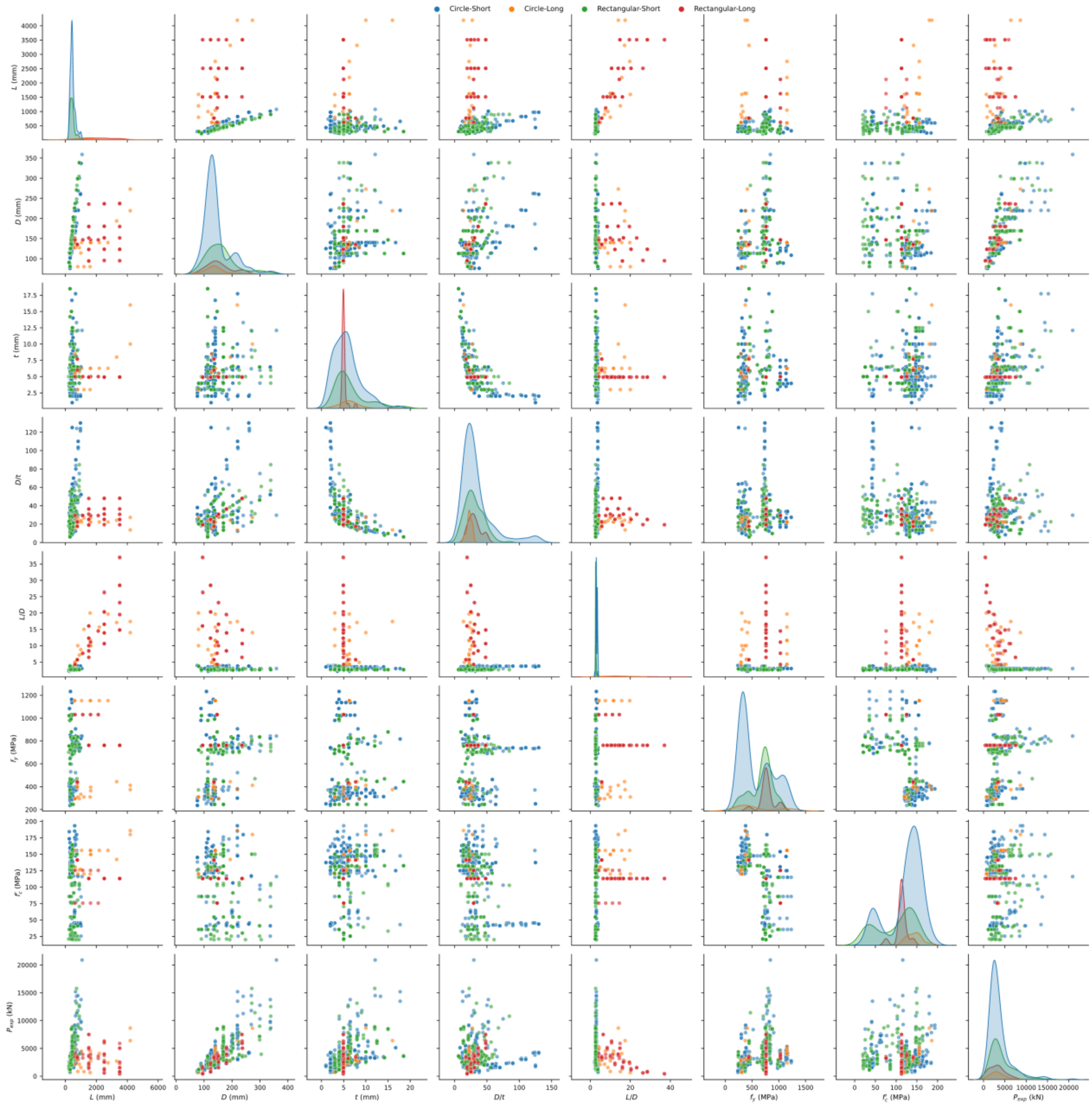

Fig. 4. Scatter matrix plot illustrates relationships between input features and the target variable.

Fig. 4 presents the scatter matrix plot for the filtered dataset, offering a comprehensive visual representation of the relationships between all pairs of input features and the target variable, P_{exp}. For example, observing the scatter plots involving P_{exp} on the bottom row, one can visually corroborate the strong positive relationship between D and P_{exp}, as the data points tend to cluster along an upward trend. Similarly, the relationship between t and P_{exp} also shows an increasing trend.

3 Artificial Intelligence Techniques

The axial capacity of CFST columns presents a complex challenge due to the nonlinear interactions of geometric and material properties. To effectively predict the axial capacity, advanced ML algorithms were applied. These include ExtraTrees, XGBoost, and GradientBoosting, selected for their robust performance in regression. Their efficacy is enhanced through Bayesian Optimization for hyperparameter tuning and validated through K-Fold cross-validation to ensure generalization. The theoretical background underpinning these methods will be explored in the following subsections.

3.1 ExtraTrees

The Extremely Randomized Trees, or ExtraTrees, Regressor is an ensemble learning method [45] that belongs to the family of tree-based algorithms. It operates by constructing a multitude of decision trees from the training dataset and outputting the average of their predictions for regression tasks. The fundamental principle of ExtraTrees is to introduce a higher degree of randomization in the tree-building process compared to standard Random Forests.

The ExtraTrees algorithm incorporates two primary sources of randomness. First, for each split in a decision tree, only a random subset of features is considered as candidates for the split. Second, regarding cut-point selection, instead of exhaustively searching for the most discriminative cut-point for a selected feature, ExtraTrees randomly selects a cut-point for each candidate feature. The best among these randomly generated splits is then chosen to split the node.

This increased randomness serves to reduce the variance of the model, often at the cost of a slight increase in bias. By averaging the predictions of many decorrelated trees, ExtraTrees can achieve strong predictive performance and is generally robust to overfitting, especially when the number of trees in the ensemble is sufficiently large.

Let the ensemble consist of M trees. For a given input vector x, each tree $T_m(\mathrm{x})$ in the ensemble produces a prediction. The final prediction $\hat{y}(\mathrm{x})$ of the ExtraTrees regressor is the average of these individual tree predictions:

$$\hat{y}(X) = \frac{1}{M}\sum_{m=1}^{M} T_m(X) \tag{3}$$

The construction of each tree T_m involves recursively splitting nodes. At each node, a split, defined by a feature j and a corresponding cut-point s, is selected to partition the data into two child nodes. In ExtraTrees, a random subset of K features is chosen from the total p features, and for each of these K features, a random split point is drawn from a uniform distribution within the feature's range in the current node's data. The split $\left(j^*, s_{j^*}^*\right)$ that maximizes the reduction in the chosen impurity criterion (e.g., variance), or equivalently, minimizes the impurity (e.g., variance) of the resulting child nodes, is then selected. This process continues until a stopping criterion is met, such as reaching a maximum tree depth or a minimum number of samples per leaf.

3.2 XGBoost

Extreme Gradient Boosting (XGBoost) is a highly efficient and scalable implementation of the gradient boosting framework [46]. It has gained considerable popularity due to its exceptional performance in many ML competitions and real-world applications. XGBoost builds upon the principles of gradient boosting by incorporating several algorithmic enhancements and system optimizations.

The core idea of gradient boosting is to add new models (typically decision trees) sequentially, with each new model attempting to correct the errors made by the previous ensemble of models. XGBoost refines this process through a regularized learning objective. The objective function at iteration t for predicting $\hat{y}_i$ for an instance x_i can be expressed as:

$$Obj^{(\mathrm{t})} = \sum_{i=1}^{n} l\left(y_{i,} y_l^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) + C \tag{4}$$

where n is the number of training instances, y_i is the true target value, $l(y_i, \hat{y}_i)$ is a differentiable convex loss function that measures the discrepancy between the prediction $\hat{y}_i$ and the true value y_i. $\hat{y}_i^{(t-1)}$ is the prediction for instance i from the ensemble of t − 1 trees, and $f_t(x_i)$ is the new tree being added at iteration t. The term $\Omega(f_t)$ is a regularization term that penalizes the complexity of the model f_t, and C represents constants from previous steps.

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{5}$$

Here, T is the number of leaves in the tree f, w_j is the score (weight) of the j-th leaf, and γ and λ are regularization parameters controlling the penalty for the number of leaves and the magnitude of leaf weights, respectively.

To optimize the objective function, XGBoost employs a second-order Taylor expansion of the loss function around the current prediction $\hat{y}_i^{(t-1)}$:

$$Obj^{(t)} \approx \sum_{i=1}^{n} \left[l\left(y_{i,} \hat{y}_l^{(t-1)}\right) + g_i f_t(x_i) + \frac{1}{2} g_i f_t(x_i)^2 \right] \Omega(f_t) + C \tag{6}$$

where $g_i = \dfrac{\partial l\left(y_{i,} \hat{y}_i^{(t-1)}\right)}{\partial \hat{y}_i^{(t-1)}}$ and $h_i = \dfrac{\partial^2 l\left(y_{i,} \hat{y}_i^{(t-1)}\right)}{\left(\partial \hat{y}_i^{(t-1)}\right)^2}$ are the first and second-order gradient statistics of the loss function. This formulation allows for an efficient tree construction process, where the optimal leaf weights and the corresponding objective function value for a given tree structure can be derived analytically.

3.3 GradientBoosting

The Gradient Boosting Regressor is another powerful ensemble learning technique based on the boosting paradigm [47]. Like XGBoost, it builds an additive model in a stagewise fashion, where each new tree (weak learner) is trained to correct the errors (pseudo-residuals) of the previous ensemble.

Let $F_0(\mathrm{x})$ be an initial constant model, often the mean of the target values. The model is then iteratively updated:

$$F_m(x) = F_{m-1}(x) + \nu h_m(x),\ m = 1, \ldots, \mathrm{M} \tag{7}$$

where M is the total number of trees, $h_m(\mathrm{x})$ is the m-th base learner (a decision tree), and ν is the learning rate (or shrinkage parameter, $0 < \nu \leq 1$), which scales the contribution of each tree to reduce overfitting. At each stage m, the algorithm first computes the pseudo-residuals r_{im} for each training instance i:

$$\gamma_{im} = -\left[\frac{\partial L\left(y_{i,} F(X_i)\right)}{\partial F(X_i)} \right]_{F(X) = F_{m-1}(X)} \tag{8}$$

where $L(y_i, F(\mathrm{x}_i))$ is the loss function. For regression with squared error loss, $L(y, F(\mathrm{x})) = \frac{1}{2}\left(y - F(\mathrm{x})\right)^2$, the pseudo-residuals are simply the true residuals $y_i - F_{m-1}(\mathrm{x}_i)$.

A new base learner $h_m(\mathrm{x})$ is then fitted to these pseudo-residuals $(\mathrm{x_i}, \mathrm{r_{im}})_{\mathrm{i}=1}^{\mathrm{n}}$. If the base learners are decision trees, specific terminal node (leaf) values γ_{jm} are determined for each region R_{jm} of the m-th tree:

$$\gamma_{im} = \arg\min_{\gamma} \sum_{Xi \in R_{jm}} L\left(y_i, \mathrm{F}_{m-1}(X_i) + \gamma\right) \tag{9}$$

The model update is then $F_m(x) = F_{m-1}(x) + \nu \sum_{j=1}^{J_m} \gamma_{j,m} I\left(x \in R_{j,m}\right)$, where J_m is the number of terminal nodes in tree h_m, ν is the learning rate, and $I\left(x \in R_{j,m}\right)$ is an indicator function that is 1 if x falls into the j-th terminal region and 0 otherwise. Gradient Boosting Regressors are effective at capturing complex non-linearities and interactions in the data. The choice of loss function, the number of trees, tree depth, and the learning rate are critical hyperparameters influencing its performance.

3.4 K-Fold Cross-Validation

The widely adopted resampling technique known as K-Fold Cross-Validation (CV) serves the dual purpose of assessing the generalization performance of ML models and streamlining the process of hyperparameter tuning. It helps in obtaining a more robust estimate of model performance than a single train-test split, especially when the dataset size is limited.

The K-Fold CV process involves the following steps: first, the entire dataset D is randomly partitioned into K mutually exclusive and roughly equally sized subsets (or "folds"), denoted as $D_1, D_2, \ldots, D_K$. Secondly, the model is trained K times. In each iteration $k \in 1, \ldots, K$: The fold D_k is held out as the validation set (test set). The remaining $K-1$ folds ($D \backslash D_k$) are used as the training set. The model is trained on $D \backslash D_k$ and its performance (e.g., R^2, MSE) is evaluated on D_k. Let this performance be P_k. Third, the overall performance estimate is typically the average of the performance obtained in each of the K iterations:

$$\bar{P} = \frac{1}{K} \sum\nolimits_{k=1}^{K} P_k \tag{10}$$

The standard deviation of these K performance scores can also be computed to assess the stability of the model's performance. A common choice for K is 5 or 10. K-Fold CV provides a less biased estimate of the model's performance on unseen data. During hyperparameter optimization, K-Fold CV is often used within each iteration of the optimization process to evaluate a given set of hyperparameters.

3.5 Bayesian Optimization

Bayesian Optimization [43] is a sequential model-based optimization (SMBO) strategy designed for global optimization of black-box functions $f(\mathrm{x})$ that are computationally expensive to evaluate. In the context of ML, $f(\mathrm{x})$ typically represents the performance of a model (e.g., cross-validation R^2 score) as a function of its hyperparameters x. The goal is to find the set of hyperparameters x^* that maximizes (or minimizes) $f(\mathrm{x})$:

$$X^* = \arg\max_{X \in \chi} f(X) \tag{11}$$

where $\mathcal{X}$ is the search space of hyperparameters. Bayesian Optimization works by iteratively building a probabilistic surrogate model of the objective function $f(\mathrm{x})$ and then using this model to decide where next to sample x in the search space. The two main components are:

Surrogate Model: A probabilistic model that approximates $f(\mathrm{x})$ and quantifies the uncertainty in that approximation. Gaussian Processes (GPs) are commonly used as surrogate models. A GP a defines prior over functions, and as data points $(\mathrm{x}_i, f(\mathrm{x}_i))$ are observed, it is updated to a posterior distribution over functions. This posterior provides a mean prediction $\mu(\mathrm{x})$ and a variance $\sigma^2(\mathrm{x})$ for any point x in the search space.

$$f(\mathrm{X}) \sim \varsigma\left(m(x), k\left(x, x'\right)\right) \tag{12}$$

where $m(\mathrm{x})$ is the mean function and $k(\mathrm{x}, \mathrm{x}')$ is the kernel (or covariance) function. Acquisition Function determines the utility of sampling at a candidate point x, balancing exploration (sampling in regions of high uncertainty) and exploitation (sampling in regions is likely to yield high objective values). Common acquisition functions include Expected Improvement (EI), Probability of Improvement (PI), and Upper Confidence Bound (UCB). For instance, the Expected Improvement is defined as:

$$EI(x) = IE\left[\max\left(0, f(x) - f\left(x^*\right)\right)\right] \tag{13}$$

where $f(\mathrm{x}^*)$ is the best value of the objective function observed so far. Using the GP posterior, $EI(\mathrm{x})$ can be computed analytically:

$$EI(x) = \left(\mu(x) - f\left(x^*\right) - \xi\right)\Phi(Z) + \sigma(x)\phi(Z) \quad if\, \sigma(x) > 0 \tag{14}$$

$$EI(x) = 0 \quad if\, \sigma(x) > 0 \tag{15}$$

where $Z = \frac{\mu(\mathrm{x}) - f(\mathrm{x}^*) - \xi}{\sigma(\mathrm{x})}$, $\Phi(\mathrm{Z})$ and $\phi(\mathrm{Z})$ are the CDF and PDF of the standard normal distribution, respectively, and ξ is a trade-off parameter (often set to 0). The iterative process involves: First, initialize with a few randomly chosen hyperparameter configurations and evaluate $f(\mathrm{x})$ for them. Second, fit the surrogate model (e.g., GP) to all observed data. Third, maximize the acquisition function to select the next hyperparameter configuration x_{next} to evaluate. Fourth, evaluate $f(\mathrm{x}_{next})$ (i.e., train and validate the ML model with x_{next}). Fifth, augment the observed data with $(\mathrm{x}_{next}, f(\mathrm{x}_{next}))$ and repeat from step 2 until a budget (e.g., number of iterations) is exhausted.

For complex models that require computationally expensive evaluations, Bayesian Optimization is an especially suitable method for hyperparameter tuning because it is designed to identify the optimal parameter set with a smaller number of evaluations compared to grid or random search strategies.

The analysis in this study was conducted in Python (3.9) using key libraries including scikit-learn (1.3.0), scikit-optimize (0.9.0), xgboost (1.7.5), shap (0.45.0), and SALib (1.4.5). Input features were used directly without scaling, as is appropriate for tree-based models. Hyperparameter tuning was performed with the BayesSearchCV function for 32 iterations, guided by an Expected Improvement acquisition function. Model explainability utilized the shap. TreeExplainer, and Sobol' sensitivity was assessed with a sample size of 1024. A fixed random seed (123) was used for all stochastic processes to ensure deterministic results.

4 Results and Discussion

In the preceding section, the utilization of ML methods was discussed. This section presents a comprehensive analysis of the predictive performance of the developed models: ExtraTrees, XGBoost, and GradientBoosting Regressors. The evaluation is based on rigorous statistical metrics, K-Fold cross-validation, and detailed error analysis, ensuring a thorough understanding of their capabilities in predicting the axial load capacity of CFST columns across diverse strengths, geometries, and slenderness.

The predictive accuracy and reliability of the models were quantified using several standard statistical metrics: the Coefficient of Determination (R^2), Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE).

$$R^2 = 1 - \frac{\sum_{i=1}^{N} \left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{N} \left(y_i - \bar{y}_i\right)^2} \tag{16}$$

$$MSE = \frac{1}{N} \sum_{i=1}^{N} \left(y_i - \hat{y}_i\right)^2 \tag{17}$$

$$MAE = \frac{1}{N} \sum_{i=1}^{N} \left|y_i - \hat{y}_i\right| \tag{18}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(y_i - \hat{y}_i\right)^2} = \sqrt{MSE} \tag{19}$$

To assess the generalization capability and robustness of the models, a 10-fold cross-validation (CV) procedure was employed. The dataset was partitioned into ten equally sized folds. In each iteration, one fold served as the test set, and the remaining nine folds were used for training.

$$\bar{R}^2_{CV} = \frac{1}{K} \sum_{k=1}^{N} R_k^2 \tag{20}$$

$$SD_{\bar{R}^2_{CV}} = \sqrt{\frac{1}{k-1}\sum_{k=1}^{k}\left(R_k^2 - \bar{R}^2_{CV}\right)^2} \tag{21}$$

The 10-fold cross-validation results, further detailed in **Table 2**, are particularly crucial for assessing model generalization and robustness against overfitting. For instance, the XGBoost model achieved a mean CV R^2 of 0.9719 with a remarkably low standard deviation of 0.0149, while GradientBoosting yielded a mean CV R^2 of 0.9775 (SD = 0.0167) and ExtraTrees a mean CV R^2 of 0.9411 (SD = 0.0448). These low standard deviations across the folds underscore the stability and consistent predictive power of the models, indicating that their performance is not unduly sensitive to specific subsets of the training data. The high mean R^2 values obtained during cross-validation (e.g., >0.97 for XGBoost and GradientBoosting) affirm the models' capability to generalize well to unseen data.

Table 2. Cross-validation results for the models

Fold	ExtraTrees	XGBoost	GradientBoosting
1	0.9908	0.9939	0.9960
2	0.8782	0.9575	0.9626
3	0.9782	0.9870	0.9929
4	0.9040	0.9666	0.9799
5	0.9822	0.9764	0.9870
6	0.9874	0.9882	0.9874
7	0.8849	0.9483	0.9356
8	0.8857	0.9565	0.9749
9	0.9676	0.9824	0.9828
10	0.9519	0.9621	0.9761
Mean	0.9411	0.9719	0.9775
SD	0.0448	0.0149	0.0167

Table 3 presents MAE, MSE, and RMSE values for all the models, illustrating the contrasting results between the predicted outcomes of the ML regression model and the actual observed data. When tested on the entire dataset, the models exhibited even higher R^2 values (e.g., ExtraTrees: 0.9995, XGBoost: 0.9977, GradientBoosting: 0.9985), approaching near-perfect correlation. The corresponding MAE and RMSE values were also notably low. For example, the ExtraTrees model demonstrated an MAE of 26.13 kN and an RMSE of 54.65 kN, indicating that, on average, the predictions were very close to the experimental values, considering the wide range of experimental capacities (410 kN to 20900 kN). Such exceptional performance underscores the effectiveness of both the selected algorithms and the Bayesian Optimization approach in meticulously adjusting hyperparameters, which enables the models to capture the inherent non-linear complexities of CFST column behavior across diverse strengths, geometries, and slenderness ratios.

Table 3. Performance metrics for ML models

	ExtraTrees		XGBoost		GradientBoosting	
	CV	Entire Dataset	CV	Entire Dataset	CV	Entire Dataset
R^2	0.9411	0.9995	0.9719	0.9977	0.9775	0.9985
MSE	-	2.99×10^3	-	1.47×10^4	-	9.47×10^3
MAE	-	26.13	-	92.09	-	70.81
RMSE	-	54.65	-	121.10	-	97.31

Visual confirmation of the models' high accuracy is provided in **Fig. 5**, **Fig. 6**, and **Fig. 7**, which plot predicted versus actual axial load capacities. These scatter plots illustrate the relationship between the model-predicted axial load capacities and the actual experimental values. For all three models, the data points are observed to congregate tightly along the line of perfect agreement ($y = \hat{y}$, indicated by the dashed red line), with minimal scatter. This visual evidence strongly corroborates the high R^2 values and low error metrics reported in **Table 3**.

The color-coding of data points by column categories (e.g., Circle-Short, Circle-Long, Rectangular-Short, Rectangular-Long) offers an initial qualitative indication that the models perform consistently across these different column configurations. There are no visually apparent clusters of poorly predicted points specific to any single column category, suggesting a balanced and robust learning process across the diverse dataset.

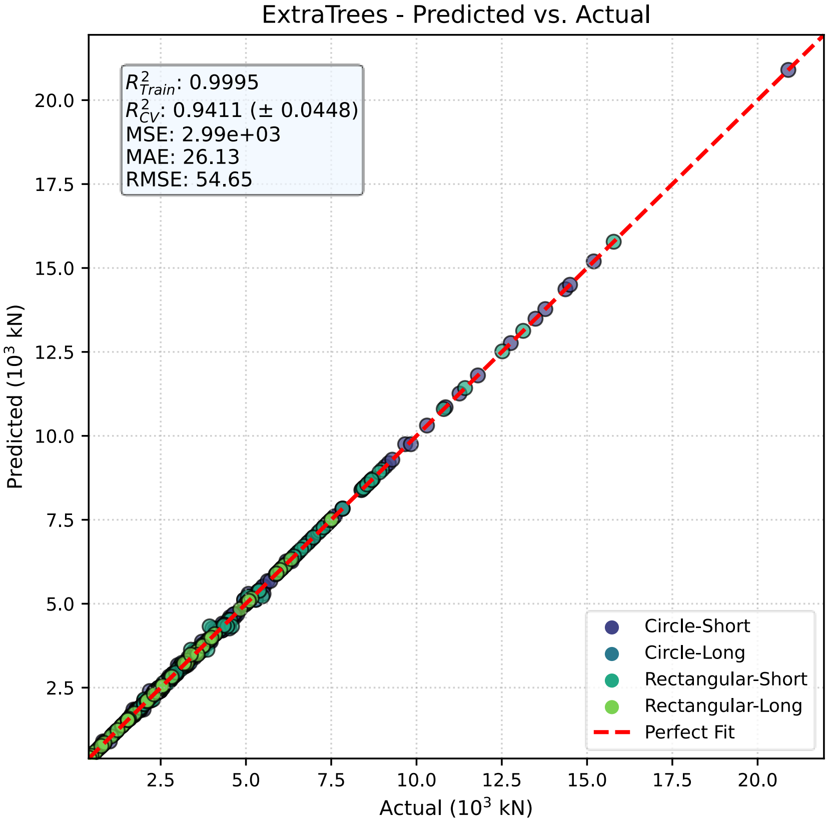

Fig. 5. Predicted vs. actual axial load capacities for the ExtraTrees model.

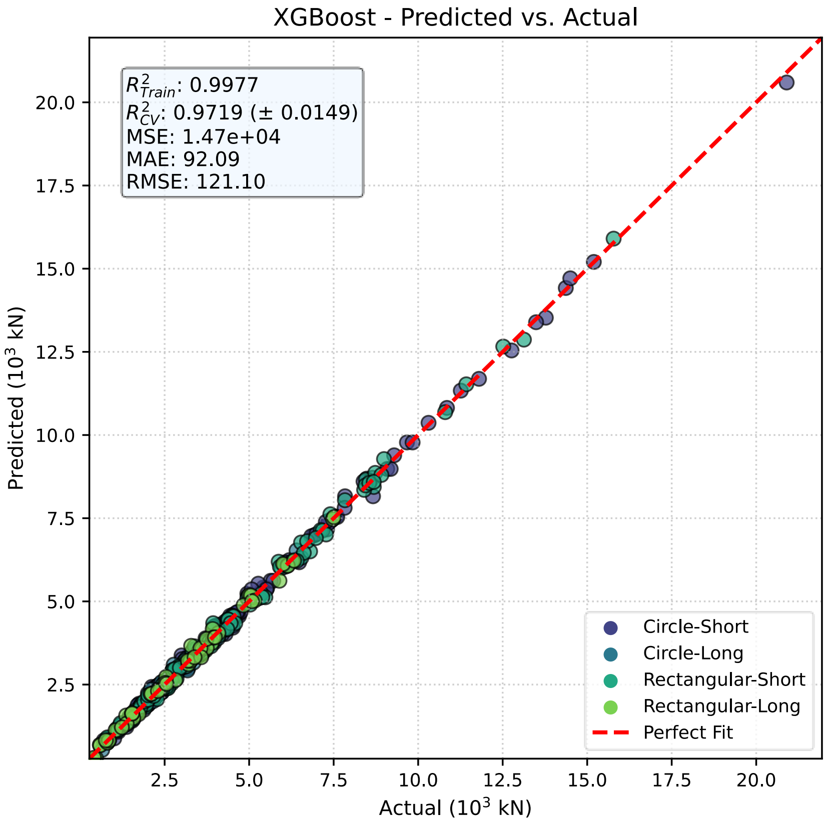

Fig. 6. Predicted vs. actual axial load capacities for the XGBoost model.

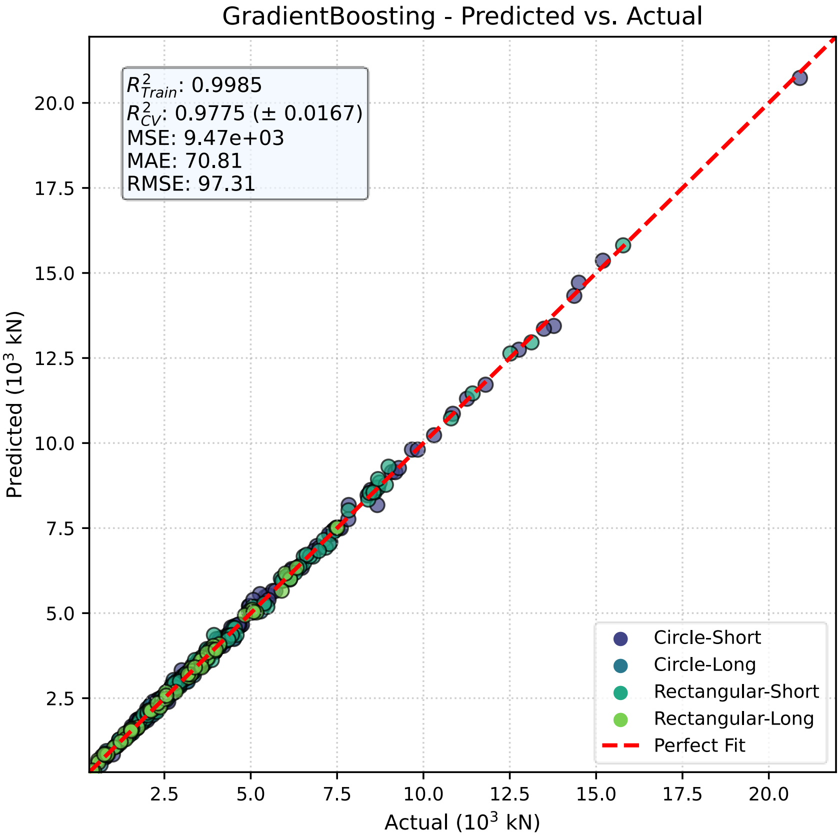

Fig. 7. Predicted vs. actual axial load capacities for the GradientBoosting model.

Table 4. Statistical performance indicators for ML models across diverse column categories

Model	category	R^2	MSE	MAE	RMSE	Count
ExtraTrees	Circle-Short	0.9997	2.36×10^3	27.15	48.62	318
	Circle-Long	0.9971	9.05×10^3	50.38	95.16	24
	Rectangular-Short	0.9994	4.30×10^3	26.99	65.57	139
	Rectangular-Long	0.9999	4.38×10^2	6.02	20.93	51
XGBoost	Circle-Short	0.9982	1.24×10^4	82.86	111.27	318
	Circle-Long	0.9916	2.66×10^4	135.45	163.16	24
	Rectangular-Short	0.9975	1.74×10^4	103.11	131.89	139
	Rectangular-Long	0.9952	1.59×10^4	99.22	125.94	51
GradientBoosting	Circle-Short	0.9988	8.40×10^3	65.31	91.68	318
	Circle-Long	0.9938	1.97×10^4	111.47	140.37	24
	Rectangular-Short	0.9984	1.10×10^4	78.18	104.92	139
	Rectangular-Long	0.9978	7.10×10^3	65.91	84.25	51

To scrutinize model performance under varied column configurations, the predictive metrics were disaggregated by CFST column categories: circular short (C-S), circular long (C-L), rectangular short (R-S), and rectangular long (R-L) columns. **Table 4** and the comparative bar charts in **Fig. 8** present these category-specific metrics. Across all categories, the models generally maintained high accuracy. The ExtraTrees model, for instance, demonstrated exceptional consistency, with R^2 values consistently above 0.997 for all four column categories (e.g., Circle-Short: 0.9997, Circle-Long: 0.9971, Rectangular-Short: 0.9994, Rectangular-Long: 0.9999). The XGBoost and GradientBoosting models also exhibited robust performance, with R^2 values typically exceeding 0.99 for short columns both circle and rectangular, and maintaining excellent, albeit slightly lower, values for long columns. The error metrics (RMSE) remained commendably low across all categories. A slight increase in error metrics was observable for long columns compared to short columns in circular sections, and the opposite in rectangular sections. For example, for the ExtraTrees model, the RMSE for Circle-Short columns was 48.62 kN, while for Circle-Long columns, it was 95.16 kN. For rectangular columns, these values were 65.57 kN and 20.93 kN, respectively.

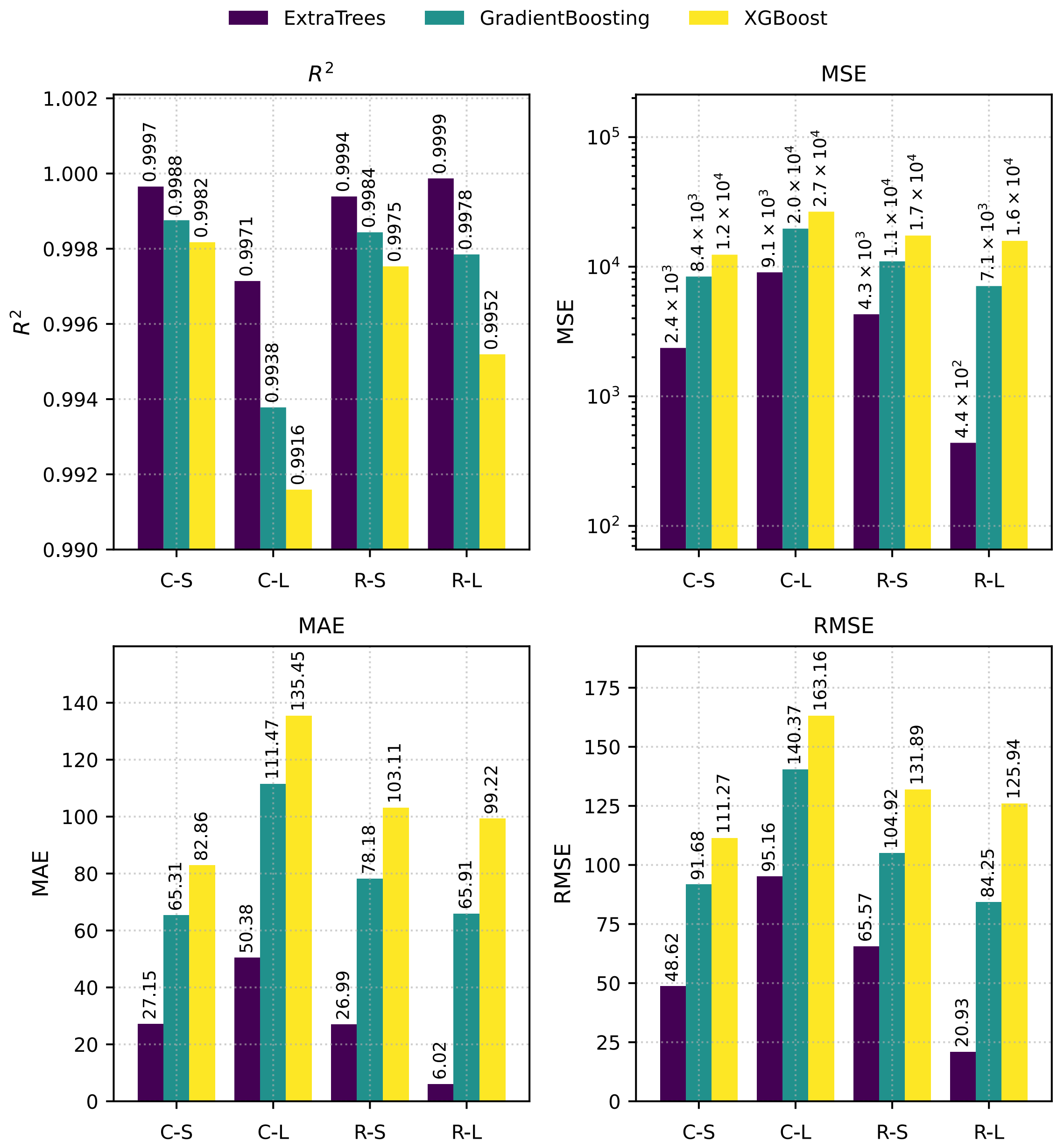

Fig. 8. Comparative bar charts of R², MSE, MAE, and RMSE across ML models.

Fig. 8 visually reinforces these findings, allowing for a direct comparison of R², MSE, MAE, and RMSE across models and column categories. This graphical representation clearly highlights the consistent high performance of all three ensemble methods, with ExtraTrees always showing marginally lower error metrics or higher R² in specific categories.

A detailed analysis of prediction errors ($e_i = y_i - \hat{y}_i$) provides further insights into model behavior. **Fig. 9** presents histograms of these errors for each model, categorized by column categories. It was observed that the error distributions for every model were nearly symmetric and centered near zero, which suggests the absence of significant systematic bias within the predictions. The spread of these distributions, while narrow, varied slightly between models, with ExtraTrees showing the tightest clustering of errors.

To contextualize the predictive accuracy of the developed models, their performance was benchmarked against existing state-of-the-art machine learning models from recent literature. A summary of comparison is presented in Table 5, which compiles reported performance metrics from several relevant studies published in high-impact journals. These studies were selected based on their methodological relevance and the transparent reporting of absolute error metrics like RMSE, which are critical for engineering assessment.

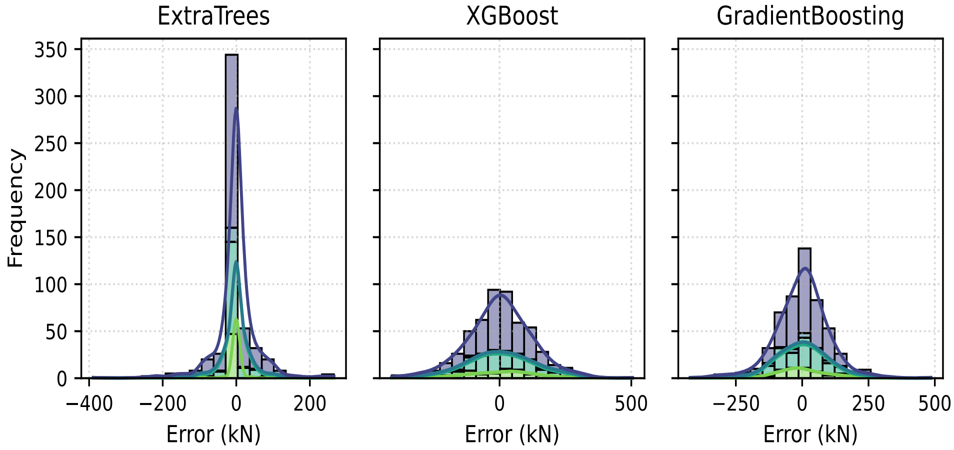

Fig. 9. Histograms of prediction errors for each ML model.

Table 5. Performance benchmark of proposed ML models against existing studies for CFST axial capacity

Study	Year	ML Model(s)	Column Type(s)	Database Size	R^2	RMSE (kN)
This Study	2026	ExtraTrees	Circular/Rectangular, Short/Slender	532	0.9995	54.65
This Study	2026	GradientBoosting	Circular/Rectangular, Short/Slender	532	0.9985	97.31
This Study	2026	XGBoost	Circular/Rectangular, Short/Slender	532	0.9977	121.10
Lai et al. [48]	2024	NGBoost (probabilistic)	Circular	1,127	0.9949	248.09
Zhou et al. [49]	2023	GPR (optimized)	Rectangular	1,641	0.9998	70.69
Wang & Chan [50]	2023	SVR, RFR, NN	Rectangular, Eccentric Loading	403	0.97 (SVR)	143.91
Cakiroglu et al. [51]	2022	CatBoost, XGBoost, RF	Rectangular Stub	719	0.9945 (RF)	205.83
Vu et al. [52]	2021	GTB, RF, SVM	Circular	>1,000	0.965 (SVM)	-
Naser et al. [53]	2021	GA, GEP (symbolic)	Circular/Rectangular, Concentric	1,245 979	R_m=0.897 R_m=0.829	384.73 295.3
Megahed et al. [11]	2023	GPR, SR	Stub (Circular, Rect., Double-skin)	1,041	0.998 (GPR)	-

The analysis of this benchmark reveals that the developed ensemble models, particularly the ExtraTrees variant (R^2 = 0.9995, RMSE = 54.65 kN), exhibit exceptional performance and establish a new precedent in terms of absolute prediction error. A direct comparison with models developed on large databases, such as the probabilistic NGBoost from Lai et al. [48], shows that the models developed herein achieve a significantly lower RMSE (54.65 kN vs. 248.09 kN). This performance is also highly competitive with specialized models, such as the optimized GPR from Zhou et al. [49], which reports an RMSE of 70.69 kN. This highlights the efficacy of the Bayesian Optimization strategy in minimizing absolute prediction error.

While other advanced models like the optimized GPR (Zhou et al. [49]) and ensemble methods from Cakiroglu et al. [51] have demonstrated excellent performance on specific CFST subsets (e.g., rectangular columns), the models in this study achieve a comparable or superior level of accuracy while being validated on a more diverse dataset. The database used herein encompasses both circular and rectangular sections across a full spectrum of slenderness, which underscores their broad generalizability. Furthermore, in contrast to symbolic regression approaches like those in Naser et al.

[53], which report a high RMSE of approximately 385 kN, the ensemble tree methods presented here showcase a markedly superior capability for precise quantitative prediction.

Collectively, this benchmark analysis confirms that the rigorous optimization of advanced ensemble models on a comprehensive and diverse dataset yields predictive tools with state-of-the-art accuracy and broad applicability, providing a clear advantage over many existing approaches in the field.

Fig. 10 illustrates the relationship between prediction errors and actual axial load capacities for the three models. A detailed examination of the error plots in **Fig. 10** reveals critical differences in the models' predictive behavior. The XGBoost and GradientBoosting models exhibit largely homoscedastic error distributions, meaning the variance of their prediction errors remains relatively constant across the full range of actual load capacities. The points are randomly scattered around the zero-error line, indicating no significant systematic bias for either low- or high-capacity columns.

In stark contrast, the ExtraTrees model demonstrates strong heteroscedasticity. For columns with lower axial capacities (approximately 5×10^3 kN), the prediction errors are substantial and widely dispersed. However, as the actual load capacity increases, the errors decrease dramatically, clustering tightly around zero for the highest-capacity columns. This pattern suggests that while the ExtraTrees model achieves exceptional accuracy for robust, high-performance columns, its reliability is significantly diminished for columns in the lower-capacity range. This behavior can be attributed to the model's underlying architecture; the optimized deep trees (max_depth=18) with minimal leaf constraints are exceptionally effective at capturing the complex behaviors of high-strength specimens but may be overfit to noise in the less-represented, lower-capacity regime. This finding underscores the importance of not relying on a single aggregate metric like R^2 and highlights the superior robustness of the boosting-based models for this specific dataset. Future work could explore mitigation strategies for the ExtraTrees model, such as weighted regression, to improve its performance on low-capacity specimens.

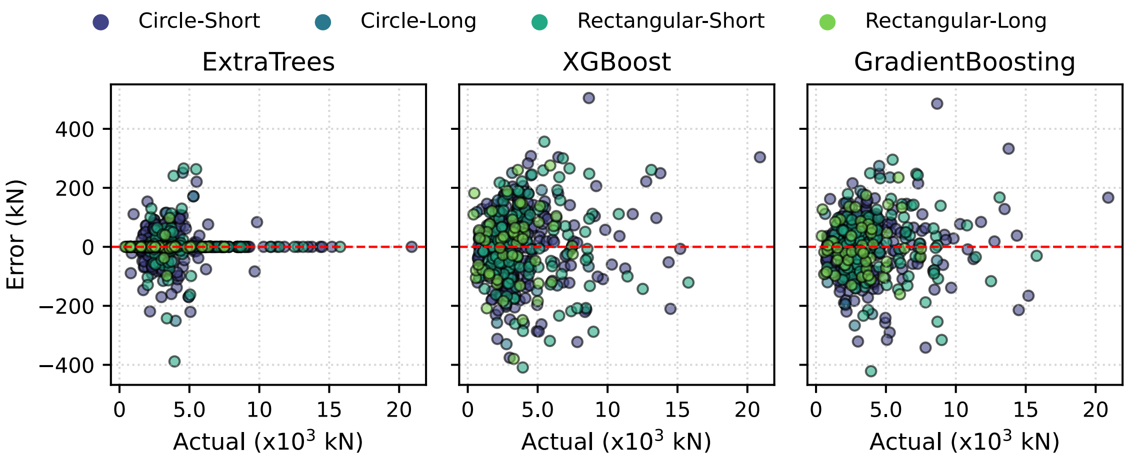

Fig. 10. Relationship between prediction errors and actual experimental tests for ML models.

For reproducing the model, the following information is provided. The predictive models were developed and analyzed within the Python ecosystem, leveraging a suite of specialized open-source libraries that directly implement the advanced methodologies described. The core of the automated hyperparameter tuning was conducted using the BayesSearchCV function from the scikit-optimize (skopt) library. This function systematically implements the Bayesian Optimization protocol by creating a probabilistic surrogate model of the objective function in this case, the 10-fold cross-validated R^2 score, to intelligently and efficiently search the high-dimensional hyperparameter space. For each model, the optimization process was executed for 32 iterations. The ensemble models themselves were implemented using their canonical libraries. This computational framework yielded the optimized hyperparameters for each model, as detailed in **Table 6**, with a specific highlight being the optimized setting for the ExtraTrees model, where the bootstrap hyperparameter was set to False.

Table 6. Optimized hyperparameters for ML models

Model	max_depth	min_samples_leaf	min_samples_split	n_estimators	colsample_bytree	gamma	learning_rate	min_child_weight	subsample
ExtraTrees	18	1	2	300	-	-	-	-	-
XGBoost	3	-	-	300	1	0	0.153	6	0.6
GradientBoosting	3	8	2	300	-	-	0.2	-	1

The optimized hyperparameters in **Table 6** are not arbitrary values; they reflect the fundamental learning strategies of each algorithm and underscore the sophistication of the Bayesian search. For the boosting models (XGBoost and GradientBoosting), the optimization process converged on very shallow trees (max_depth=3). This is a characteristic design principle in boosting algorithms, where individual “weak learners” are created to prevent overfitting, and the ensemble’s predictive power is built by sequentially adding many trees (n_estimators=300) that correct the errors of their predecessors. The moderate learning_rate further tempers this process. XGBoost’s configuration reveals additional regularization: subsample=0.6 introduces stochasticity by training each tree on a random 60% of the data, while min_child_weight=6 prevents the model from creating splits based on very small, potentially noisy, sample groups.

Conversely, the ExtraTrees model’s parameters illustrate a different philosophy. It was optimized for deep trees (max_depth=18) with minimal constraints on leaf size (min_samples_leaf=1), allowing each tree to fit the training data very closely. The key setting bootstrap=False indicates that the entire dataset was used to build each tree. Overfitting is controlled not by weakening the learners, but through two powerful randomization mechanisms: first, selecting a random subset of features at each split, and second, selecting a random threshold for each feature, which is the defining characteristic of “Extremely Randomized Trees.” The final prediction is an average of these diverse, deep trees, which effectively cancels out individual tree variance.

To ensure the scientific integrity of this study, the hyperparameter search space for each model must also be reported. These ranges, which guided the Bayesian optimization, were defined as follows: for XGBoost, n_estimators (50-300), max_depth (3-10), learning_rate (0.01-0.3), subsample (0.6-1.0), colsample_bytree (0.6-1.0), gamma (0-5), and min_child_weight (1-10); for GradientBoosting, n_estimators (50-300), learning_rate (0.01-0.2), max_depth (3-8), min_samples_split (2-10), min_samples_leaf (1-8), and subsample (0.6-1.0); and for ExtraTrees, n_estimators (50-300), max_depth (5-20), min_samples_split (2-10), min_samples_leaf (1-8), and bootstrap ([True, False]). To ensure the reproducibility and scientific integrity of this study, a fixed random seed (random_state=123) was consistently employed across all computational stages, including the KFold cross-validation splits, the model initializations, and the BayesSearchCV optimization process.

The practical implications of the findings for the construction industry are significant. These accurate and reliable predictive tools enable engineers to design more efficient and economical structures without compromising safety by substantially reducing uncertainty in column capacity prediction. By accommodating a wide spectrum of materials from conventional to ultra-high-performance and diverse geometries, the models broaden the applicability of CFST. This is particularly valuable for innovative structural applications where traditional design codes are limited or unvalidated.

5 Explanation of Models

To unravel the intricate decision-making processes embedded within the developed ML models and to rigorously quantify the influence of various input parameters on the predicted axial load capacity of CFST columns, a dual approach of interpretability and sensitivity analysis was undertaken. This comprehensive investigation employed SHAP to provide model-agnostic attribution of feature importance at both global and local levels. Complementing this, Sobol’ indices were utilized to conduct a global sensitivity analysis, discerning the contribution of each input parameter and their interactions to the variance observed in the model outputs.

5.1 Model Interpretability through Feature Importance Analysis (SHAP)

The SHAP framework offers a unified and theoretically sound method for interpreting the predictions of complex ML models. It assigns a specific value of importance, known as the SHAP value, to each input feature for every individual prediction made by the model. The core principle of SHAP is rooted in cooperative game theory, attributing the difference between a specific prediction and the average prediction to the contributions of individual feature values. For an input instance z' comprising M features, the prediction $g(z')$ is explained as an additive combination of the base value ϕ_0 (which represents the average model output over the training dataset) and the sum of SHAP values ϕ_j corresponding to each feature z'_j:

$$g\left(z^{'}\right) = \phi_0 + \sum_{j=1}^{M} \phi_j \tag{22}$$

where ϕ_0 is the base value (average model output over the training dataset) and ϕ_j is the SHAP value for feature j.

The overall importance of each input feature across all predictions was ascertained by calculating the mean absolute SHAP value for that feature. This provides a global ranking of feature influence. **Fig. 11** presents the SHAP summary plots for the ExtraTrees, XGBoost, and GradientBoosting models. These plots are typically rendered as beeswarm plots, where each point represents a SHAP value for a feature for a single prediction. The horizontal axis indicates the SHAP value, signifying the impact on the model output (positive values increase the prediction, negative values decrease it). The color of each point corresponds to the original value of the feature (e.g., red for high values, blue for low values), and the vertical spread of points at a given SHAP value shows the density of predictions with that feature impact.

An examination of **Fig. 11** reveals a remarkably consistent hierarchy of feature importance across all three ML models. The column diameter (D or its equivalent D_{eq} for rectangular sections) consistently emerged as the most influential parameter. For these analyses, the feature D (mm) represents the actual diameter for circular columns and the calculated equivalent diameter (D_{eq}) for rectangular columns. The fact that this unified feature consistently emerged as the most important predictor across all three distinct models strongly validates its effectiveness as a powerful, shape-agnostic descriptor of cross-sectional scale. High values of diameter (typically shown in red on the beeswarm plots) are predominantly associated with large positive SHAP values, indicating a substantial increase in the predicted axial capacity, which aligns perfectly with fundamental mechanics. Similar to diameter, increased steel tube thickness resulted in higher positive SHAP values, signifying its positive contribution to load-bearing capacity.

The column length (L) also exhibited significant influence; longer columns generally corresponded to negative SHAP values, reflecting a reduction in predicted capacity due to increased slenderness and propensity for buckling. Material strengths, specifically the steel yield strength (f_y) and the concrete compressive strength (f_c'), demonstrated considerable impact. Higher values of both f_y and f_c' were typically associated with positive SHAP values, thereby increasing the predicted axial capacity, as expected. The derived slenderness ratios, L/D and D/t, were also found to be influential. Higher L/D ratios generally led to negative SHAP values, decreasing the predicted capacity. The impact of D/t was often more complex, sometimes showing an inverse relationship with capacity (higher D/t implying a more slender, less robust tube wall, leading to negative SHAP values) or interacting significantly with other features. The consistent directionality and magnitude of these feature effects, as visualized in **Fig. 11**, underscore the physical realism captured by the ML models. For instance, in the XGBoost summary plot, the clear separation of red points (high feature value) towards positive SHAP values for D (mm) and t (mm) and towards negative SHAP values for L (mm) is evident.

Beyond assessing global feature importance, the SHAP methodology facilitates a granular understanding of individual predictions through the generation of force plots. **Fig. 12** displays illustrative SHAP force plots for selected individual predictions made by the ML models, encompassing circle short CFST columns. Each force plot visualizes how the specific feature values for a given data instance collectively contribute to “push” or “pull” the model’s output from the base value (the average prediction across the dataset) to the final predicted value for that instance. Features that increase the prediction are shown as red arrows pushing to the right, while features that decrease the prediction are shown as blue arrows pushing to the left; the length of each arrow is proportional to the magnitude of the feature's SHAP value for that prediction.

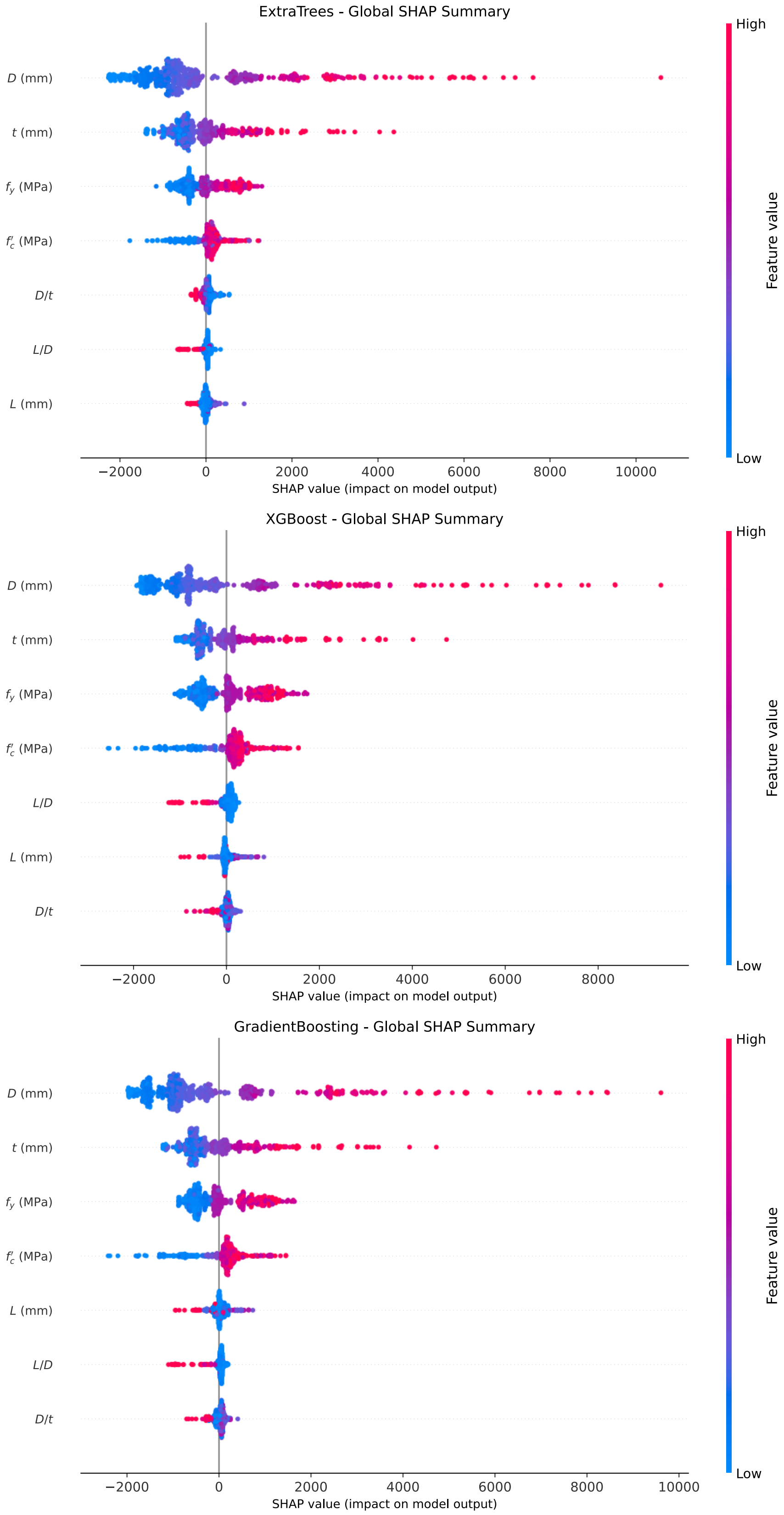

Fig. 11. SHAP summary plots illustrating global feature importance and the distribution of SHAP values for ML models. Features are ranked by mean absolute SHAP value.

Fig. 12. SHAP force plots for individual predictions, illustrating the contribution of each feature to the final predicted axial capacity $\boldsymbol{f(x)}$ for selected samples from ML models.

For example, consider the ExtraTrees model's prediction shown in the figure for a circle short column. The final predicted capacity $f(x)$ is 2328.85 kN, which is lower than the model's base value. The force plot shows that the concrete compressive strength (f_c' = 163.2 MPa) is the feature providing a positive contribution, as indicated by its red arrow pushing the prediction higher. However, this is more than offset by negative contributions from the other features. The diameter (D = 133.0 mm), tube thickness (t = 3.0 mm), and steel yield strength (f_y = 290.0 MPa) all significantly decrease the predicted capacity, as shown by their large blue arrows. The D/t ratio (= 44.3) also exerts a slight negative influence, depicted by a small blue arrow.

5.2 Quantitative Sensitivity Analysis (Sobol Indices)

A global sensitivity analysis that utilized Sobol' indices was performed to more deeply analyze and quantify how each input parameter contributes to the variance in the model's output. This variance-based method decomposes the total variance of the model output into fractions attributable to individual input parameters and to their cooperative interactions.

The first-order Sobol' index, $S1_i$, for an input parameter X_i quantifies the direct contribution of this parameter alone to the total output variance $V(Y)$. It measures the expected reduction in output variance if X_i were to be fixed, averaged over all possible fixed values of X_i. Mathematically, it is defined as:

$$S1_i = \frac{V_{X_j}\left(E_{X_{\sim i}}\left(Y / X_i\right)\right)}{V\left(Y\right)} \tag{23}$$

where $E_{X_{\sim i}}(Y|X_i)$ represents the expected value of the output Y conditional on X_i being fixed, with the expectation taken over all other input parameters $X_{\sim i}$. Var_{X_i} denotes the variance taken over the distribution of X_i.

The total-order Sobol' index, S_{Ti}, for a parameter X_i encapsulates its total effect on the output variance. This includes not only its first-order effect (direct contribution) but also the effects arising from all interactions, of any order, between X_i and any combination of other input parameters. Consequently, S_{Ti} provides a comprehensive measure of a parameter's overall importance. The difference, $S_{Ti} - S1_i$, quantifies the collective contribution of all interaction terms involving X_i to the output variance. A substantial positive value for this difference signifies that the parameter X_i is heavily involved in complex interactions with other input variables.

Fig. 13, **Fig. 14**, and **Fig. 15** present the computed first-order ($S1_i$) and total-order (S_{Ti}) Sobol' indices for the ML models. These analyses were performed systematically for distinct CFST column categories: C-L, R-L, C-S, and R-S.

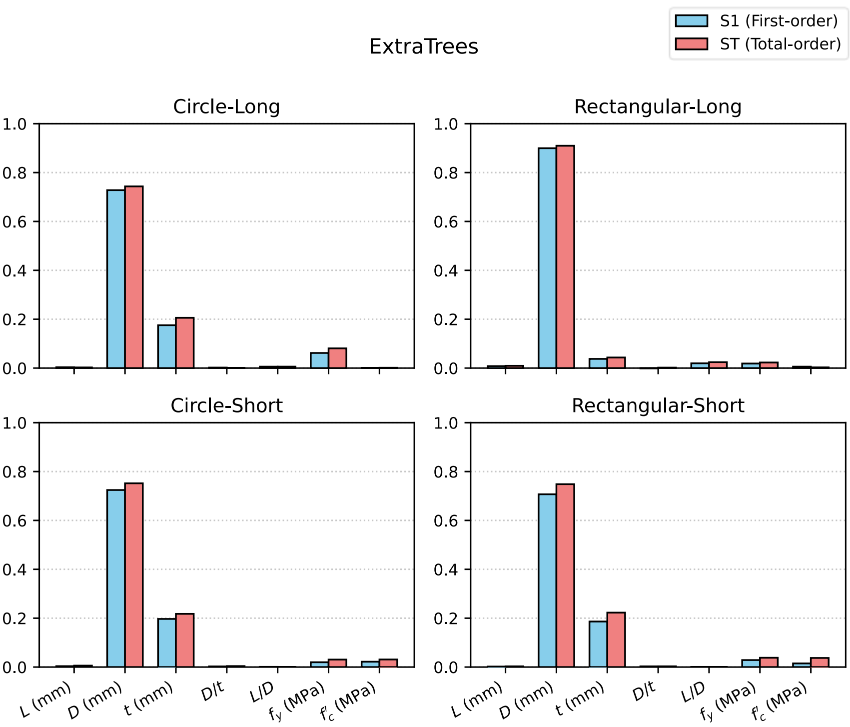

Fig. 13. First-order (**$S1_i$**) and total-order (**S_{Ti}**) Sobol' indices for the ExtraTrees model.

Across all models and column categories, the Sobol' analysis results consistently underscored the preeminent influence of the primary geometric parameters on the predicted axial load capacity's variance. The column diameter (D or D_{eq}) invariably exhibited the highest $S1_i$ and S_{Ti} values, signifying its dominant role in driving the variability of the model predictions. For instance, as depicted in **Fig. 13** for the ExtraTrees model applied to circular long columns, the diameter D shows a first-order index $S1_i \approx 0.68$ and a total-order index $S_{Ti} \approx 0.70$. This indicates that approximately 68% of the variance in predicted capacity for this category can be attributed directly to variations in diameter, with an additional 2% stemming from its interactions. Similar patterns of dominance for diameter are observable in **Fig. 14** and **Fig. 15** for the XGBoost and GradientBoosting models.

The steel tube thickness (t) generally ranked as the second most sensitive parameter in terms of its contribution to output variance. Material strengths, particularly steel yield strength f_y, often demonstrate moderate sensitivity. Concrete compressive strength f_c' generally exhibited lower sensitivity values in the Sobol' analysis across most categories and models, suggesting that while it is an important parameter (as shown by SHAP), its variability within the dataset contributes less to the overall variance of predictions compared to D, t, and often f_y. Parameters such as column length L, and the derived ratios L/D and D/t, typically exhibited lower first-order and total-order Sobol' sensitivities compared to the primary cross-sectional dimensions (D, t) and, in some cases, material strengths. However, Sobol' indices measure the contribution to variance of the output across the entire dataset (or subset for a category). If the range or distribution of D and t in the dataset causes a larger spread in predicted capacities (e.g., P_{exp}) than the range or distribution of L/D, then D and t will show higher Sobol' indices. This does not diminish the local importance of L/D for a specific column's stability, as captured by SHAP, but rather reflects its relative impact on the overall variability of predictions within the dataset's parameter space.

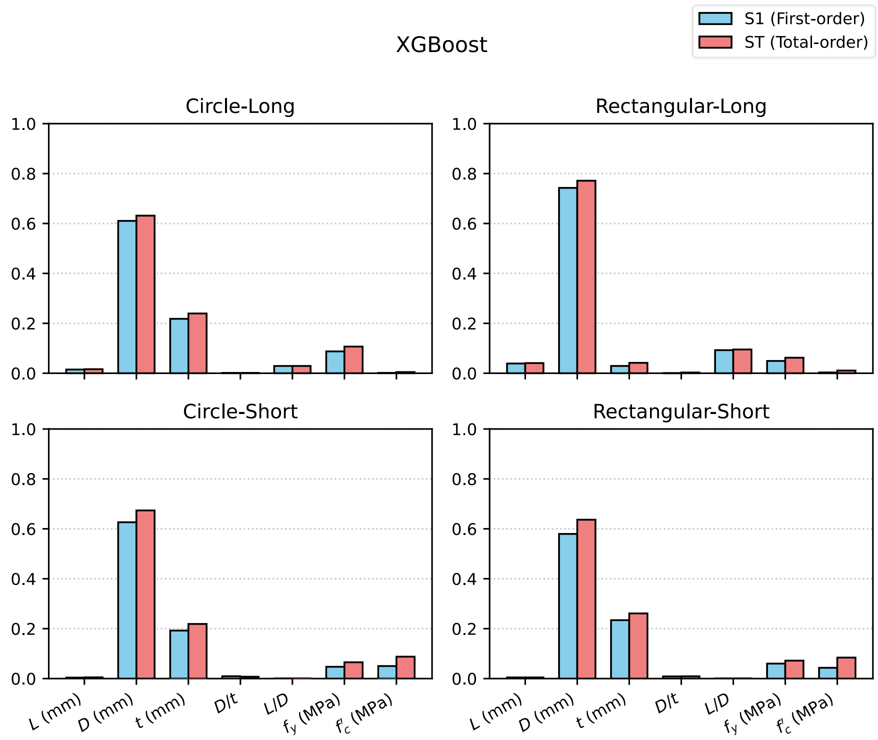

Fig. 14. First-order ($\boldsymbol{S1}_i$) and total-order ($\boldsymbol{S}_{\boldsymbol{T}i}$) Sobol' indices for the XGBoost model.

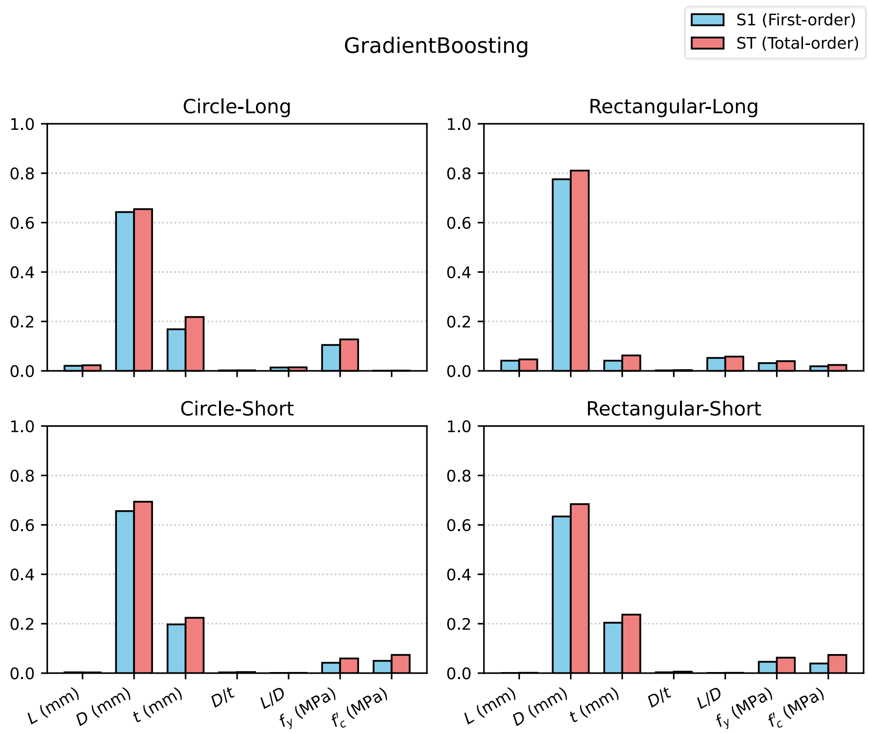

Fig. 15. First-order ($\boldsymbol{S1}_i$) and total-order ($\boldsymbol{S}_{\boldsymbol{T}i}$) Sobol' indices for the GradientBoosting model.

The disparity between S_{Ti} and $S1_i$ for various parameters, evident in many subplots of **Fig. 13-15**, confirms the presence and significance of interaction effects. Parameters like t and f_y frequently showed S_{Ti} values noticeably greater than their corresponding $S1_i$ values, underscoring their involvement in complex, non-additive relationships that influence CFST column behavior. The ability of the ML models to capture these intricate interactions is a key factor in their high predictive performance.

Utilizing SHAP analysis and Sobol' indices revealed that ML models align with structural engineering principles. Geometric properties (column diameter, steel tube thickness) are paramount, followed by material strengths, with the models effectively capturing complex non-linearities. This comprehensive understanding significantly boosts confidence in the models' reliability and applicability for diverse CFST designs.

6 Conclusions

This study successfully developed and validated a suite of Bayesian Optimization-enhanced ML models specifically ExtraTrees, XGBoost, and GradientBoosting for the accurate prediction of axial load capacity in CFST columns. The research addressed a diverse range of column parameters, including circular and rectangular cross-sections, short and slender configurations, and a wide spectrum of material strengths from normal to ultra-high performance concrete and high-strength steel. Based on the comprehensive analysis conducted, the following key conclusions are drawn:

1) Exceptional predictive accuracy was a key finding, with all developed machine learning models achieving Coefficients of Determination (R^2) near unity when evaluated on the entire dataset. Rigorous 10-fold cross-validation confirmed their robustness and generalization capabilities, with mean R^2 values of 0.9411 (ExtraTrees), 0.9719 (XGBoost), and 0.9775 (GradientBoosting). This performance underscores the models' ability to capture the complex, non-linear behavior inherent in CFST columns.

2) The application of Bayesian Optimization proved highly effective in systematically tuning the hyperparameters of the employed ensemble learning algorithms. This sophisticated optimization strategy was instrumental in unlocking the full predictive potential of the models, enabling them to achieve the reported high levels of accuracy and reliability across the diverse dataset.

3) The models exhibited consistent and high-accuracy performance across various sub-categories of CFST columns.

4) SHAP analysis provided crucial insights into the models' decision-making processes. Across all models, geometric parameters, D_{eq} and t consistently emerged as the most influential factors, followed by f_y and f_c'.

5) Sobol' indices corroborated the SHAP findings, with D_{eq} and t showing the highest first-order ($S1_i$) and total-order (S_{Ti}) sensitivity indices, indicating their dominant contribution to the variance in predicted axial load capacity.

This research provides reliable and efficient predictive tools that can facilitate more accurate and economical design of CFST structures by reducing uncertainty in capacity prediction. Although the models performed robustly, future work could enhance their applicability. Research efforts could focus on expanding the database to include CFST members under more complex scenarios such as combined axial load and bending, cyclic loading, or fire exposure. Further investigation into other advanced ML algorithms and the explicit inclusion of parameters like concrete confinement effects would also be beneficial. Ultimately, developing user-friendly software based on these validated models would facilitate their practical implementation by engineers in the construction industry. Finally, and most importantly, the models should be validated against a new, prospectively collected set of experimental data to confirm their predictive accuracy in real-world scenarios, which remains the ultimate test of any predictive tool. It is also critical to acknowledge that the models are most reliable for interpolation within the domain of the training data. Users should exercise caution when extrapolating to material strengths or geometric ratios significantly outside the ranges presented in **Table 1**, as predictive accuracy is not guaranteed.

Nomenclature

Symbol	Description	Units
P_{exp}	Experimental ultimate axial load capacity	kN
f_c'	Concrete compressive cylinder strength (150 mm $\times$ 300 mm cylinder)	MPa
f_y	Steel yield strength	MPa
L	Member length of the CFST column	mm
D	Diameter of circular CFST column. For rectangular sections, D can represent the equivalent diameter D_{eq}.	mm
B	Width of rectangular CFST column	mm
H	Depth (Height) of rectangular CFST column	mm
t	Steel tube thickness	mm
D_{eq}	Equivalent diameter for rectangular sections, calculated as $\sqrt{(4 \times B \times H)/\pi}$	mm
L/D	Slenderness ratio (overall column slenderness; L/D for circular sections, L/H for rectangular sections). The equivalent diameter D_{eq} is NOT used for slenderness classification — only as a unified geometric feature for ML input	Dimensionless
D/t	Diameter-to-thickness ratio (section slenderness for circular sections)	Dimensionless
b/t	Width-to-thickness ratio (section slenderness for rectangular sections, where b is width of rectangular tube element)	Dimensionless
R^2	Coefficient of Determination	Dimensionless
MSE	Mean Squared Error	kN^2
MAE	Mean Absolute Error	Units of target variable
RMSE	Root Mean Squared Error	Units of target variable
$\bar{R}^2_{CV}$	Mean R^2 score from K-Fold cross-validation	Dimensionless
$\text{SD}_{R^2_{\text{CV}}}$	Standard deviation of R^2 scores from K-Fold cross-validation	Dimensionless
y_i	Actual (experimental) value for the i-th observation	Units of target variable
$\hat{y}_i$	Predicted value for the i-th observation by the model	Units of target variable
$\bar{y}$	Mean of actual (experimental) values	Units of target variable
K	Number of folds in K-Fold cross-validation	Dimensionless
$\text{Obj}^{(t)}$	Objective function at iteration t (XGBoost)	Varies (depends on loss and regularization)
$l(y_i, \hat{y}_i)$	Differentiable convex loss function	Varies
$\Omega(f_t)$	Regularization term for tree f_t (XGBoost)	Varies
T	Number of leaves in a tree (XGBoost)	Dimensionless
w_j	Score (weight) of the j-th leaf (XGBoost)	Varies
γ	Regularization parameter (complexity control for number of leaves in XGBoost)	Dimensionless
λ	Regularization parameter (L2 regularization on leaf weights in XGBoost)	Dimensionless
g_i, h_i	First and second-order gradient statistics of the loss function (XGBoost)	Varies
$F_m(\boldsymbol{x})$	Model prediction at m-th iteration (Gradient Boosting)	Units of target variable
ν	Learning rate (or shrinkage parameter) in Gradient Boosting	Dimensionless
$h_m(\boldsymbol{x})$	Prediction of the m-th base learner (decision tree)	Units of target variable
r_{im}	Pseudo-residuals for i-th instance at m-th stage (Gradient Boosting)	Units of target variable
$\mu(\boldsymbol{x})$	Mean prediction of Gaussian Process posterior	Units of objective function
$\sigma^2(\boldsymbol{x})$	Variance of Gaussian Process posterior	(Units of objective function)^2
$\text{EI}(\boldsymbol{x})$	Expected Improvement (acquisition function in Bayesian Optimization)	Units of objective function
ϕ_0	Base value in SHAP (average model output over the training dataset)	Units of target variable
ϕ_j	SHAP value for feature j (contribution of feature j to the prediction difference from base value)	Units of target variable
$S1_i$	First-order Sobol' index for input parameter X_i	Dimensionless
S_{Ti}	Total-order Sobol' index for input parameter X_i	Dimensionless
$V(Y)$	Total variance of the model output Y	kN^2
M	Number of trees in an ensemble (e.g., ExtraTrees)	Dimensionless

Declaration of AI and AI-assisted Technologies in the Writing Process

During the preparation of this work, the author used Gemini 2.5 Pro to improve readability and language. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

Funding Statement

The author received no specific funding for this study.

CRediT authorship contribution statement

Khalid Saqer Alotaibi: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing.

Conflicts of Interest

The author declares that he has no conflicts of interest to report regarding the present study.

Data Availability Statement

Some or all data, models, or codes that support the findings of this study are available from the corresponding author upon reasonable request.

References

[1] Shanmugam NE, Lakshmi B. State of the art report on steel–concrete composite columns. Journal of Constructional Steel Research 2001; 57(10): 1041–1080. https://doi.org/10.1016/S0143-974X(01)00021-9.

[2] Han LH, Li W, Bjorhovde R. Developments and advanced applications of concrete-filled steel tubular (CFST) structures: members. Journal of Constructional Steel Research 2014; 100: 211–228. https://doi.org/10.1016/j.jcsr.2014.04.016.

[3] Li L, Lai Z, Chen B. Ultra-high strength concrete filled steel tube members: classification, experimental database, and design. Engineering Structures 2024; 300: 117210. https://doi.org/10.1016/j.engstruct.2023.117210.

[4] Uy B. Strength of concrete filled steel box columns incorporating local buckling. Journal of Structural Engineering 2000; 126(3): 341–352. https://doi.org/10.1061/(ASCE)0733-9445(2000)126:3(341).

[5] Tao Z, Han LH, Wang D. Strength and ductility of stiffened thin-walled hollow steel structural stub columns filled with concrete. Thin-Walled Structures 2008; 46(10): 1113–1128. https://doi.org/10.1016/j.tws.2008.02.007.

[6] Thai S, Thai HT, Uy B, Ngo T. Concrete-filled steel tubular columns: test database, design and calibration. Journal of Constructional Steel Research 2019; 157: 161–181. https://doi.org/10.1016/j.jcsr.2019.02.024.

[7] Klöppel K, Goder W. Traglastversuche mit ausbetonierten Stahlrohren und Aufstellung einer Bemessungsformel. Der Stahlbau 1957; 26(1): 1–10.

[8] Nishiyama I, Morino S, Sakino K, Nakahara H, Fujimoto T, Mukai A. Summary of research on concrete-filled structural steel tube column system carried out under the U.S.-Japan cooperative research on composite and hybrid structures. Tsukuba, Japan: Building Research Institute 2002.

[9] Han LH, Zhao X, Yang Y, Feng J. Experimental study and calculation on the fire resistance of concrete-filled hollow steel columns. Journal of Structural Engineering 2003; 129(3): 346–356. https://doi.org/10.1061/(ASCE)0733-9445(2003)129:3(346).

[10] Chen B, Wang T, Wu X. Overview of concrete filled steel tube arch bridges in China. Practice Periodical on Structural Design and Construction 2009; 14(2): 70–80. https://doi.org/10.1061/(ASCE)1084-0680(2009)14:2(70).

[11] Megahed K, Mahmoud N, Abd-Rabou S. Application of machine learning models in the capacity prediction of RCFST columns. Scientific Reports 2023; 13(1): 20878. https://doi.org/10.1038/s41598-023-48044-1.

[12] Mursi M, Uy B. Strength of slender concrete filled high strength steel box columns. Journal of Constructional Steel Research 2004; 60(12): 1825–1848. https://doi.org/10.1016/j.jcsr.2004.05.002.

[13] Goode C, Dennis L. Concrete-filled steel tube columns-tests compared with Eurocode 4. Composite Construction in Steel and Concrete VI 2012; 86(16): 317-325. https://doi.org/10.1061/41142(396)26.

[14] Gourley B, Tort C, Denavit M, Schiller P, Hajjar J. A synopsis of studies of the monotonic and cyclic behavior of concrete-filled steel tube members, connections, and frames. Urbana, IL: Newmark Structural Engineering Laboratory, University of Illinois at Urbana-Champaign 2008.

[15] Lai Z, Varma A, Zhang K. Noncompact and slender rectangular CFT members: Experimental database, analysis, and design. Journal of Constructional Steel Research 2014; 101(Oct): 455-468. https://doi.org/10.1016/j.jcsr.2014.06.004.

[16] Ahmadi M, Naderpour H, Kheyroddin A. Utilization of artificial neural networks to prediction of the capacity of CCFT short columns subject to short term axial load. Archives of Civil and Mechanical Engineering 2014; 14(3): 510–517. https://doi.org/10.1016/j.acme.2014.01.006.

[17] Liew JY, Xiong M, Xiong D. Design of concrete filled tubular beam-columns with high strength steel and concrete. Structures 2016; 8: 213–226. https://doi.org/10.1016/j.istruc.2016.05.005.

[18] Varma A. Seismic behavior, analysis and design of high strength square concrete filled steel tube (CFT) columns [Doctoral dissertation]. Bethlehem, PA: Lehigh University 2000.

[19] Portolés J, Serra E, Romero M. Influence of ultra-high strength infill in slender concrete-filled steel tubular

columns. Journal of Constructional Steel Research 2013; 86: 107–114. https://doi.org/10.1016/j.jcsr.2013.03.016.
[20] Xiong M, Xiong D, Liew JY. Axial performance of short concrete filled steel tubes with high- and ultra-high- strength materials. Engineering Structures 2017; 136: 494–510. https://doi.org/10.1016/j.engstruct.2017.01.037.
[21] Gardner NJ, Jacobson ER. Structural behavior of concrete filled steel tubes. ACI Journal Proceedings 1967; 64(7): 404–413. https://doi.org/ 10.14359/7575.
[22] Schneider SP. Axially loaded concrete-filled steel tubes. Journal of Structural Engineering 1998; 124(10): 1125–1138. https://doi.org/10.1061/(ASCE)0733-9445(1998)124:10(1125).
[23] Fujimoto T, Nishiyama I, Mukai A, Baba T. Test results of eccentrically loaded short columns - square CFT columns, In: Proceedings of Second Joint Technical Coordinating Committee Meeting, US-Japan Cooperative Research Program, Phase 5 Composite and Hybrid Structures. Honolulu, HI: National Science Foundation 1995; 443–457.
[24] Han L. Tests on stub columns of concrete-filled RHS sections. Journal of Constructional Steel Research 2002; 58(3): 353–372. https://doi.org/10.1016/S0143-974X(01)00059-1.
[25] Hajjar JF, Schiller PH, Molodan A. A distributed plasticity model for concrete-filled steel tube beam-columns with interlayer slip. Engineering Structures 1998; 20(8): 663–676. https://doi.org/10.1016/S0141-0296(97)00107-7.
[26] Varma A, Ricles J, Sause R, Lu L. Seismic behavior and modeling of high-strength composite concrete-filled steel tube (CFT) beam-columns. Journal of Constructional Steel Research 2002; 58(5–8): 725–758. https://doi.org/10.1016/S0143-974X(01)00099-2.
[27] Varma A, Sause R, Ricles J, Li Q. Development and validation of fiber model for high-strength square concrete-filled steel tube ceam-columns. ACI Structural Journal 2005; 102(1): 73–85. https://doi.org/10.14359/13532.
[28] Lee J, Fenves GL. Plastic-damage model for cyclic loading of concrete structures. Journal of Engineering Mechanics 1998; 124(8): 892–900. https://doi.org/10.1061/(ASCE)0733-9399(1998)124:8(892).
[29] American Institute of Steel Construction. Specification for Structural Steel Buildings (ANSI/AISC 360-22). Chicago, IL: American Institute of Steel Construction 2022.
[30] European Committee for Standardization. Eurocode 4: design of composite steel and concrete structures – Part 1-1: general rules and rules for buildings (EN 1994-1-1:2004). Brussels: European Committee for Standardization (CEN) 2004.
[31] Standards Australia/Standards New Zealand. AS/NZS 2327:2017 Composite structures – Composite steel-concrete construction in buildings. Sydney, Australia: Standards Australia/Standards New Zealand 2017.
[32] Leon RT, Kim DK, Hajjar JF. Limit state response of composite columns and Part 1: formulation of design provisions for the 2005 AISC specification. Engineering Journal, AISC 2007; 44(4): 341–358. https://doi.org/10.62913/engj.v44i4.919.
[33] American Institute of Steel Construction. Specification for Structural Steel Buildings (ANSI/AISC 360-10). Chicago, IL: American Institute of Steel Construction 2010.
[34] Adeli H. Neural networks in civil engineering: 1989–2000. Computer-Aided Civil and Infrastructure Engineering 2001; 16(2): 126–142. https://doi.org/10.1111/0885-9507.00219.
[35] Salehi H, Burgueño R. Emerging artificial intelligence methods in structural engineering. Engineering Structures 2018; 171: 170–189. https://doi.org/10.1016/j.engstruct.2018.05.084.
[36] Al-Jamimi H, Al-Kutti W, Alwahaishi S, Alotaibi K. Prediction of compressive strength in plain and blended cement concretes using a hybrid artificial intelligence model. Case Studies in Construction Materials 2022; 17: e01238. https://doi.org/10.1016/j.cscm.2022.e01238.
[37] Alotaibi K, Islam A. Symbolic regression model for predicting compression strength of prismatic masonry columns confined by FRP. Buildings 2023; 13(2): 509. https://doi.org/10.3390/buildings13020509.
[38] Alotaibi K. Predicting the strength of fiber reinforced polymer materials externally bonded to masonry using artificial intelligent techniques. Frontiers of Structural and Civil Engineering 2025; 19(2): 242–261. https://doi.org/10.1007/s11709-025-1136-0.
[39] Alotaibi K, Almohammed F. Forecasting interfacial bond strength in FRP-reinforced concrete using soft computing techniques. Construction and Building Materials 2025; 473: 140827. https://doi.org/10.1016/j.conbuildmat.2025.140827.
[40] Ren Q, Li M, Zhang M, Shen Y, Si W. Prediction of ultimate axial capacity of square concrete-filled steel tubular short columns using a hybrid intelligent algorithm. Applied Sciences 2019; 9(14): 2802. https://doi.org/10.3390/app9142802.
[41] Nadeem K, Ahmad R, Mohsin T, Khan Q, Raza A. Predictions of ANN model for axial strength of FRP-confined CFST columns, In: 3rd International Conference on Advances in Civil and Environmental Engineering (ICACEE-2024). Taxila, Pakistan: University of Engineering & Technology Taxila 2024; 91–101.

[42] Du Y, Chen Z, Zhang C, Cao X. Research on axial bearing capacity of rectangular concrete-filled steel tubular columns based on artificial neural networks. Frontiers of Computer Science 2017; 11(5): 863–873. https://doi.org/10.1007/s11704-016-5113-6.

[43] Snoek J, Larochelle H, Adams R. Practical Bayesian optimization of machine learning algorithms. Advances in Neural Information Processing Systems 2012; 25: 2951–2959.

[44] Lundberg S, Lee S-I. A Unified approach to interpreting model predictions, In: 31st Conference on Neural Information Processing Systems (NIPS 2017). Long Beach, CA, USA 2017; 30: 4765–4774.

[45] Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. Machine Learning 2006; 63(1): 3–42. https://doi.org/10.1007/s10994-006-6226-1.

[46] Chen T, Guestrin C. XGBoost: a scalable tree boosting system, In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY: ACM 2016; 785–794. https://doi.org/10.1145/2939672.2939785.

[47] Friedman J. Greedy Function Approximation: A Gradient Boosting Machine. Annals of Statistics 2001; 29(5): 1189–1232. https://doi.org/10.1214/aos/1013203451.

[48] Lai D, Wei J, Contento A, Xue J, Briseghella B, Albanesi T, Demartino C. Machine learning-based probabilistic predictions for concrete filled steel tube (CFST) column axial capacity. Structures 2024; 70: 107543. https://doi.org/10.1016/j.istruc.2024.107543.

[49] Zhou X, Hou C, Feng W. Optimized data-driven machine learning models for axial strength prediction of rectangular CFST columns. Structures 2023; 47: 760–780. https://doi.org/10.1016/j.istruc.2022.11.030.

[50] Wang C, Chan T. Machine learning (ML) based models for predicting the ultimate strength of rectangular concrete-filled steel tube (CFST) columns under eccentric loading. Engineering Structures 2023; 276: 115392. https://doi.org/10.1016/j.engstruct.2022.115392.

[51] Cakiroglu C, Islam K, Bekdaş G, Isikdag U, Mangalathu S. Explainable machine learning models for predicting the axial compression capacity of concrete filled steel tubular columns. Construction and Building Materials 2022; 356: 129227. https://doi.org/10.1016/j.conbuildmat.2022.129227.

[52] Vu Q, Truong V, Thai H. Machine learning-based prediction of CFST columns using gradient tree boosting algorithm. Composite Structures 2021; 259: 113505. https://doi.org/10.1016/j.compstruct.2020.113505.

[53] Naser M, Thai S, Thai H. Evaluating structural response of concrete-filled steel tubular columns through machine learning. Journal of Building Engineering 2021; 34: 101888. https://doi.org/10.1016/j.jobe.2020.101888.