Electronic Journal of Structural Engineering

Original Article

Cite this: DOI:10.56748/ejse.24833

Received Date: 26 May 2024
Accepted Date: 20 August 2025

1443-9255
https://ejsei.com/ejse


# Hybrid and Ensemble Machine Learning Approaches for Predicting Axial Load Capacity in Rectangular CFST Stub Columns

**Zhibo Wang** [a*]

[a] Department of Architectural Engineering, Henan College of Transportation, Zhengzhou, Henan, 450461, China.
*Corresponding author: zdwzb2000@163.com

## Abstract

Concrete-filled steel tubular (CFST) columns are widely utilized in Structural Engineering because of their outstanding Load-Bearing Capacity (LBC) and ductility. Current design codes offer inconsistent predictions for Axial Load Capacity (ALC), particularly for high-strength Rectangular CFST Stub Columns (R-CFST columns), leading to uncertainty in practical applications. This study addresses this gap by developing interpretable and accurate Machine Learning (ML) models for predicting the ALC of such columns. A Database of 719 experimental results was compiled, encompassing six input features related to geometry and material properties. The core ML algorithm used is Histogram Gradient Boosting Regression (HGBR), which is further enhanced using two metaheuristic optimization algorithms: the Lotus Effect Optimization Algorithm (LEOA) and the Emperor Penguin Optimization Algorithm (EPOA). An ensemble strategy based on Dempster–Shafer theory (D–S theory) is also proposed. Model performance is evaluated using $R^2$, RMSE, MSE, MRAE, and RSR metrics. The hybrid HGBR–LEOA model (HGLA) achieved the best performance with $R^2$ = 0.9933 and RMSE = 202.728 in the test set. A sensitivity analysis using the Pearson Correlation Coefficient (PCC) identified Wall Thickness (t) and Section Width (B) as the most influential features. The outcomes illustrate that the suggested ML models significantly outperform traditional design code predictions and offer a fast, reliable alternative for early-stage structural design. This document provides a practical, data-driven framework that bridges the gap between empirical behavior and design code limitations for R-CFST columns.

## Keywords

**Concrete-filled steel tubular, Machine Learning, Pearson correlation coefficient, Histogram gradient boosting regression**

## 1. Introduction

High-strength composite columns offer greater strength, ductility, and stiffness, along with enhanced energy absorption in comparison with traditional steel columns or Reinforced Concrete (RC) (Damico & Conti, 2024; Khajavi et al., 2025; Liu & Gho, 2005; Naeim, Akbarzadeh, et al., 2024; Naeim, Khiavi, et al., 2024; Sakino et al., 2004a; Zhong, 2006). CFSTs leverage the optimal attributes of concrete and steel. Encasing the concrete core, the Steel Tube (ST) builds a tri-axial stress environment that boosts strength and ductility. Consequently, the solid core substantially reduces the probability of premature local buckling of the ST (Hatzigeorgiou, 2008). Additionally, CFST columns are highly effective in streamlining the construction process and lowering costs, as there is no need for formwork. Consequently, CFST columns are increasingly sought after in building and bridge construction.

To decrease the usage of construction materials, significant attention has been directed towards utilizing high-yield-strength steel and high-compressive-strength concrete in rectangular CFST stub column configurations. Research conducted by Liew et al. (Liew et al., 2016) and Xiong et al. (Xiong et al., 2017) evaluated these configurations, using steel yield strengths ($f_y$) and concrete compressive strengths ($f_c'$) values of 193 MPa and 779 MPa, respectively, while Khan et al. (Khan et al., 2017) provided similar findings with $f_y$ at 762 MPa and $f_c'$ at 113 MPa. Mursi and Uy (Mursi & Uy, 2004; Uy, 2001) and Sakino et al. (Sakino et al., 2004b) also replicated these tests, achieving $f_y$ values reaching 761 MPa and 853 MPa, respectively. The design instructions for CFST configurations are outlined in several design codes, including those from the United States (AISC 360–16 (*Specification for Structural Steel Buildings (ANSI/AISC 360-16) - 2016 | American Institute of Steel Construction*, n.d.)), the United Kingdom (BS 5400 (Institution, 1982)), and Europe (Eurocode 4 (Johnson & Anderson, 2001)).

The codes mentioned do not address the whole range of material strengths. Additionally, there is a disparity in the results achieved by these codes for the same problems, thus hindering the precise determination of CFST column strength. These limitations linked to the codes can potentially be addressed by ML algorithms, which analyze the database and identify correlations between input and output variables (Almustafa & Nehdi, 2020; Olalusi & Awoyera, 2021; Solhmirzaei et al., 2020). Lai and Varma (Lai & Varma, 2015) assembled a comprehensive collection of data on robust rectangular CFST columns. Drawing from an extensive parametric investigation using Finite Element (FE) analysis, they also proposed new formulas for design and a resistance factor (ϕ) to evaluate the strength of the Cross Section (C/S). The steel tube and concrete filling's fundamental stress-strain relationships were demonstrated. Creating validated FE schemes for CFST columns can be extremely challenging and time-consuming, although FE methods can accurately simulate their behavior. Thus, creating precise and dependable data-oriented methods that can steadily forecast the capability of CFST columns is essential.

Scholars have proposed many ML tactics to help predict Failure Modes (FMs), analyze progressive collapse, estimate capacity, design, and compute structural demands (Fu, 2020; Jeon et al., 2014; Mangalathu, Hwang, et al., 2020; Rahman et al., 2021; Seo et al., 2012). Among the utilization of ML schemes in SE are identifying FMs in circular RC bridge columns (Mangalathu & Jeon, 2019), evaluating bridge vulnerability to fire (Kodur & Naser, 2021), quickly predicting earthquake destruction to structures (Mangalathu, Sun, et al., 2020), classifying FMs, and forecasting shear strength in beam-column joints, shear walls, and RC beams (Keshtegar et al., 2022; Mangalathu, Jang, et al., 2020; Mangalathu & Jeon, 2018). Previous studies have also utilized data-oriented schemes to examine the CFST columns' capacity. Tran et al. (Tran et al., 2019, 2020) compiled a substantial dataset of circular and R-CFST columns and deployed Artificial Neural Networks (ANN) to determine the maximum strength of these columns. Zarringol et al. (Zarringol et al., 2020) similarly, an ANN-based approach was utilized to forecast the capacity of CFST columns, round and oblong, in terms of eccentric and concentric loading conditions. Naser et al. (Naser et al., 2021) examined 3,103 CFST section test results and applied Genetic Algorithms (GA) alongside Gene Expression Programming (GEP) to create data-driven models. All these explorations found that the precision of the predictive schemes significantly exceeded that of the compressive capacity estimates based on the code. Memarzadeh et al. (Memarzadeh et al., 2023) examined the LBC of columns featuring square and circular shapes, utilizing GEP in conjunction with ANN techniques. In their research, Chen Wang and Tak-Ming Chan (Wang & Chan, 2023) used three ML methods—Neural Networks (NN), Support Vector Regression (SVR), and Random Forest Regression (RFR)—to enhance models that forecast the final strength of CFSTs subjected to eccentric loads.

It should be pointed out that having a highly predictive potential and accuracy is not necessarily an indicator of the reliability of an ML scheme. The ML schemes' construction ought to be well-explainable and interpretable. The goal of this exploration is to develop interpretable ML schemes to project the ALC of R-CFST columns. The authors focus solely on projecting the ALC in R-CFST columns in this paper. HGBR, along with LEOA and EPOA, is used in this study to set up single and hybrid ML

schemes. Furthermore, an ensemble method on the basis of D–S theory is utilized to integrate the HGBR with two optimizers. A sensitivity analysis based on the PCC was also done to demonstrate the correlation of compressive strength with six features of the exploration.

## 2. Data Collection

The database employed in this research was sourced from Kaggle (https://www.kaggle.com/datasets/celalakrolu/rectangularstubcolumn) , comprising 719 data points dedicated to assessing the Axial Load-Bearing Capacity (ALBC) of Stub Columns (SCs), on the basis of their material and geometry characteristics. The ALC of columns is predicted based on six different features, including the width of the cross-section (B), Height of the cross-section (H), steel casing wall thickness (t), length of the specimen (L), yield stress of steel $(f_y)$, and $(f_c')$. **Error! Reference s ource not found.** illustrates the distribution and ranges of input and output values based on five primary statistical metrics: Median, Minimum (Min), Maximum (Max), Average, and Standard Deviation (SD). According to the table, the Max and Min values for both B and H are 400 and 60 mm, and the average values for t and L are 4.69 mm and 522.02 mm, respectively. Also, the SD for $f_y$, $f_c'$ and ALC are 167.87, 31.98, and 2049.18. Furthermore, **Error! Reference source not found.** i llustrates the frequency distribution of inputs and output values. For instance, the most frequent features B and H happen between 100 and 200 mm, and most of the output data is distributed between 0 and 4000 MPa.

***Table 1. Numerical properties of input and output variables***

| Components | Properties | | | | |
|---|---|---|---|---|---|
| | Max | Min | Average | Median | St. Dev |
| B (mm) | 400 | 60 | 149.30 | 130 | 59.46 |
| H (mm) | 400 | 60 | 161.36 | 150 | 57.69 |
| t (mm) | 12.5 | 0.7 | 4.69 | 4.38 | 2.31 |
| L (mm) | 1514 | 60 | 522.02 | 480 | 211.95 |
| $f_y$ (MPa) | 835 | 115 | 397.13 | 340.1 | 167.87 |
| $f_c$ (MPa) | 164.1 | 8.52 | 53.98 | 44.9 | 31.98 |
| Axial load capacity (KN) | 14116 | 275 | 2528.48 | 1845 | 2049.18 |

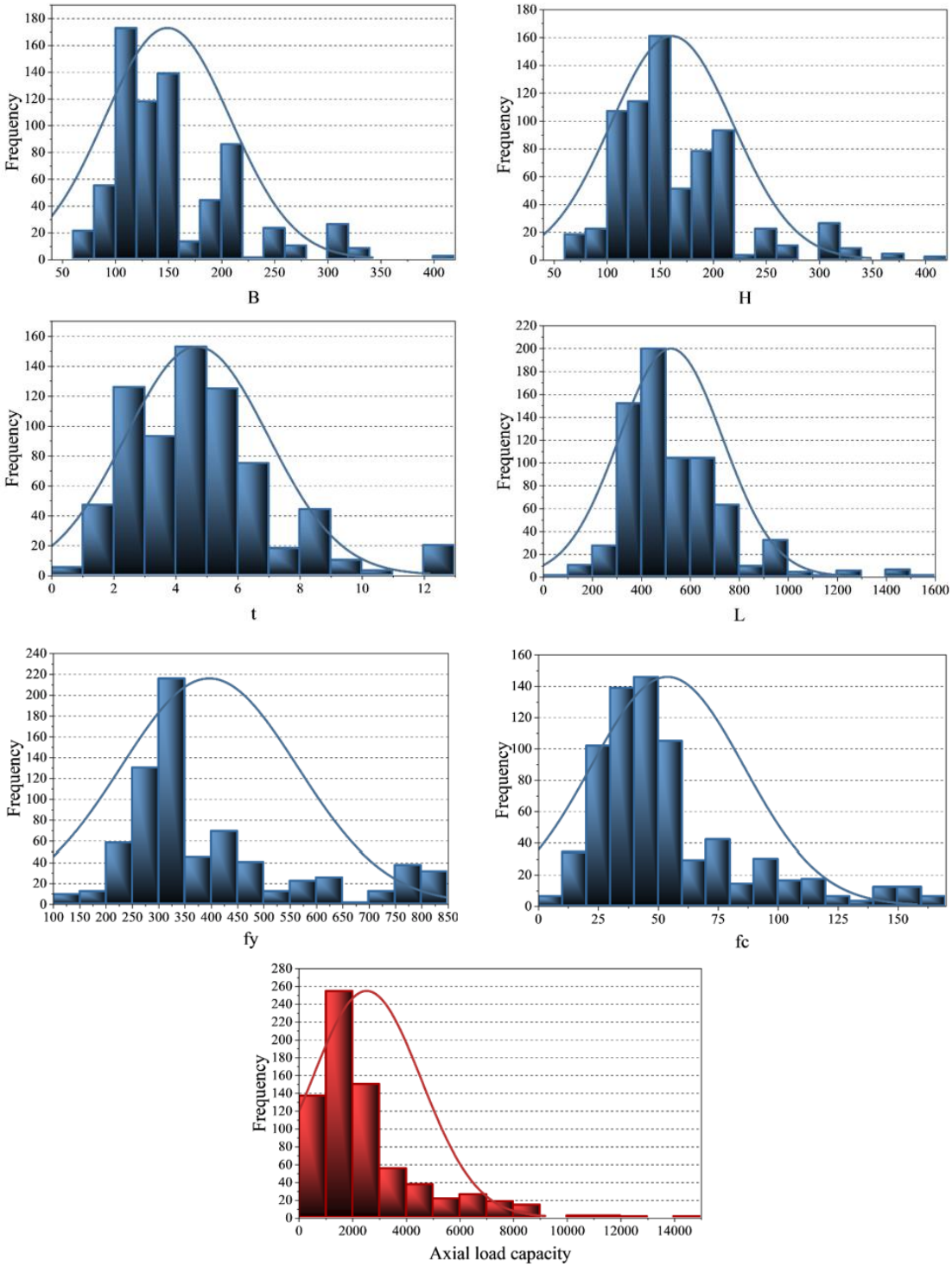


***Fig. 1 Frequency distribution plots for input and output factors***

Fig.2 indicates the correlation matrix between the ALC of columns and six key Input Parameters (IPs): B, H, t, L, $f_y$, and $f'_c$. The heatmap reveals that B and H show moderate to strong positive correlations with the ALC, indicating that larger C/S dimensions significantly enhance a column's ability to resist compressive forces. This is consistent with fundamental structural mechanics, which states that axial strength increases with the C/S area. The steel casing t shows a moderate positive correlation, suggesting that thicker steel walls contribute to better confinement and increased resistance to axial forces, particularly in composite column systems. In contrast, L exhibits a weak negative or near-zero correlation with ALC, implying that longer columns may be more prone to buckling and instability, which negatively impacts their effective axial performance. Additionally, both the $f_y$ and the $f'_c$ demonstrate moderate to strong positive correlations with axial capacity, reflecting their crucial roles in enhancing structural strength. Higher $f_y$ values increase the column's resistance to deformation under load, while stronger concrete mixes (higher $f'_c$ directly contribute to bearing capacity. Overall, this analysis reveals that parameters increasing geometric size and material strength have a positive influence on ALC, while geometric instability factors, such as excessive length, may have an adverse or limited effect.

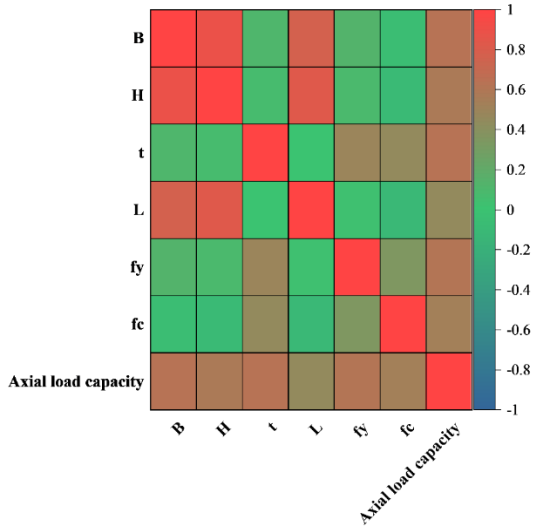


*Fig. 2 Correlation between parameters of the DB*

# 3. Machine Learning

## 3.1 HBGR

The procedure initiates by fitting a regressor to the DB, then continues with training additional regressors on the residuals left by the initial model. These weaker models are subsequently combined to form the final algorithm, which aims to minimize the loss function defined in Eq. (1). HGBR, a variant of Gradient-Boosting Regressor, distinguishes itself by employing histograms to enhance the efficiency of gradient and Hessian calculations for the loss function.

$$K = \sum_{i=1}^{N} \left(S_i - \hat{S}_i\right)^2 \quad (1)$$

At each cycle, a weak learner $P_t$ (h) is trained using the residual errors from earlier regressors. The DB is first categorized into bins according to the weak learner's decision tree and the input attributes' values. Next, it calculates values for the gradients and Hessians directly from the data through histograms, eliminating any approximations related to the involved loss. The weight assigned to the learner is calculated accurately using these gradients and Hessians. A significant advantage of histogram gradient boosting is its natural capacity to handle missing values and categorical features by effectively creating new groupings for each unique category or any missing data. The final model is developed by calculating a weighted average of each weak learner, as demonstrated in Eq. (2).

$$\hat{S}(h) = \sum_{t=1}^{T} \beta_t P_t(h) \quad (2)$$

$\beta_t$ represents the learner's weight assigned to $t_{th}$ weak learner (Li, 2024). **Error! Reference source not found.** illustrates the process of HGBR.

*Fig. 3 The design of the HGBR (Li, 2024)*

## 3.2 LEOA

The pollination process of the lotus flower and the distinct characteristics of its leaves have motivated the development of the proposed LEOA. This algorithm combines a localized water-like search on lotus leaves with the pollination-inspired principles of the dragonfly algorithm, enhancing its effectiveness. The stages of exploration and exploitation in LEAO are described below:

Exploration, as insects such as dragonflies assist in the dispersal of seeds, and their flight patterns can be tailored for this function (refer to the dragonfly optimization algorithm).

Exploitation: As blossoms open, concentrate on a key center; they offer inspiration for nearby exploration by using a multi-group strategy that arranges the search elements (based on lotus flowers).

Exploitation Guiding water across the plant's foliage, allowing it to exit through the nearest pores, might motivate a focused examination within the algorithm to determine the best spots (using local search techniques, namely, the Hill Climbing Algorithm—HCA) (Dalirinia et al., 2024). Fig. 4 illustrates the LEOA procedure.

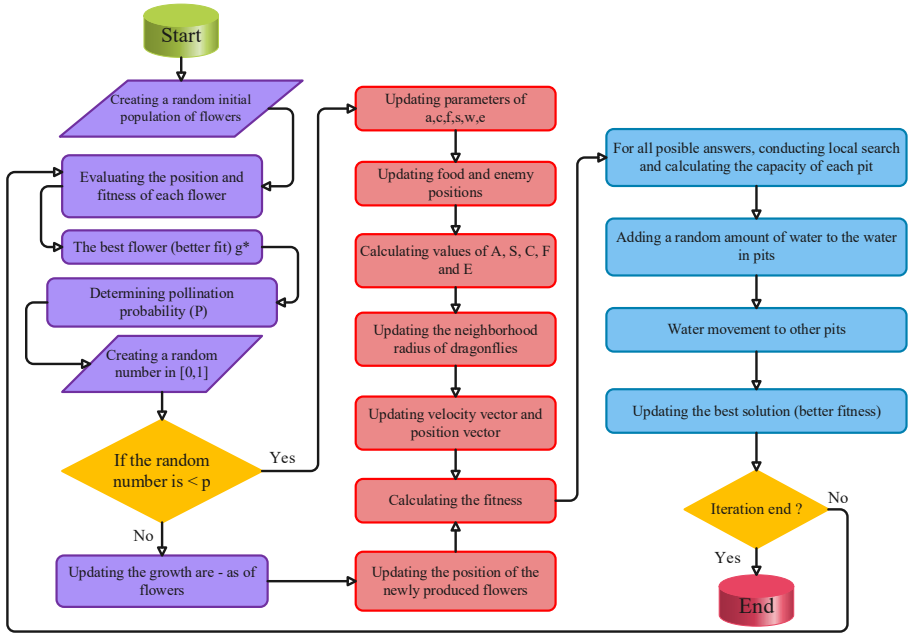


*Fig. 4 The diagram of the LEOA (Dalirinia et al., 2024)*

### 3.3 EPOA

The Emperor Penguin (EP) (Dhiman & Kumar, 2018), scientifically designated as Aptenodytes forsteri, is recognized as the most significant and heaviest type. The plumage and size of male and female EPs are alike. Their backs and heads are black, while their bellies are white, displaying a pale-yellow chest and bright yellow accents near the ears. EPs are unique in that they gather together to survive the severe Antarctic winter. The huddling actions observed in EPs can be separated into four distinct phases.

1. Establish the boundary of the EP huddle.
2. Assess the temperature distribution surrounding the huddle.
3. Measure the spacing between the EPs.
4. Identify and reposition the effective mover.

This approach mainly aims to locate a capable facilitator. The huddle is positioned on a 2-D L-shaped polygonal surface. First, EPs haphazardly outline the huddle's border. Subsequently, the spread of temperature surrounding the huddle is analyzed. The distance between the EPs is evaluated to facilitate further investigation and utilization. Ultimately, the efficient relocation specialist, or optimal solution, is discovered, and the huddle border is recalibrated on the basis of the updated positions of the EPs. Fig. 5 illustrates the diagram of the optimization framework of EPOA.

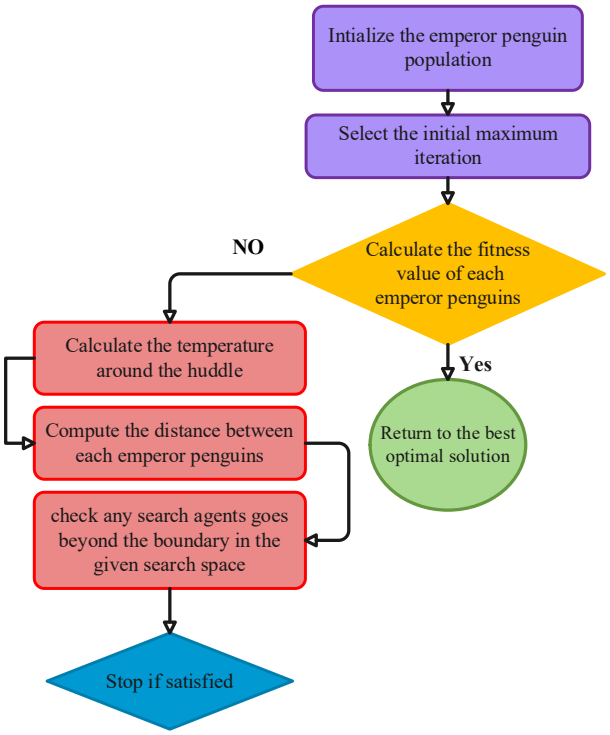


*Fig. 5 The flowchart of the EPOA (Dhiman & Kumar, 2018)*

### 3.4 Dempster-Schafer Theory for Optimizers

Evidence theory, also called D–S theory (Zhao et al., 2022), extends the concept of Bayesian subjective probability. It represents a structure for uncertain thinking and shares connections with other theories, such as probability and possibility theory. Dempster's rule combines belief degrees, fusing them, and bases the theory on subjective probabilities, taking data assumptions and the number of responses into account. The D-S evidence hypothesis is fundamentally on the basis of 2 significant thoughts: 1. Subjective probabilities regarding linked matters are used to infer levels of confidence; 2. Degrees of belief, constructed on various parts of proof, are aggregated utilizing Dempster's rule. Belief degrees rely on subjective probabilities for corresponding issues or hypotheses. This indicates the degree to which present evidence supports a particular assertion but does not necessarily assign exact probabilities. This ability to accommodate flexibility enables the theory to handle uncertainty without demanding prior knowledge. Dempster's combination rule (Shafer, 2016), being the foundation of the theory, is a process of combining several sources of evidence to adjust the levels of belief. It compiles various evidence sources by superimposing their impacts while diminishing inconsistencies. The degree of belief assigned to a proposition reflects how strong the evidence in its support is and its coherence with other evidence.

### 3.5 Sensitivity Analysis Based on PCC

PCC (Zhou et al., 2016), gauges the linear relationship between a pair of Random Variables (RV) (real-valued vectors). The PCC is traditionally the earliest method developed for gauging correlation and remains a widely used metric for assessing relationships. The covariance of variables X and Y is split by the product of their SDs, serving as a normalization factor, to describe the PCC between them. Another alternative definition can also be provided for it:

$$r_{xy} = \frac{\sum(x_i - \bar{x})\sum(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}} \quad (3)$$

The mean values of $y$ and $x$ are represented by $\bar{y}$ and $\bar{x}$ respectively, and the correlation coefficient $r_{xy}$, which ranges from $-1$ to 1, stays unaffected by linear transformations of either variable.

PCC determines the intensity of the linear link between the 2 RVs, $x$ and $y$. A positive correlation coefficient indicates that the variables are positively correlated, whereas a negative one indicates an inverse correlation. No relationship between $x$ and $y$, when $r_{xy}$ equals 0. As the absolute value of $|rxy|$ approaches 1, the strength of the linear association between the variables increases. This is because the measure of association indicates the direction of the tendency for changes in one variable to be predictive of changes in the corresponding variable between the two profiles. Fig. 6 illustrates the PCC sensitivity analysis results to illustrate the impact of inputs on compressive strength, which is the target value of the exploration. As is denoted in the figure, the inputs B and t are the features with the highest coefficients of 0.63332 and 0.62726, and the lowest coefficient belongs to feature L, which indicates the weakest impact on the target value.

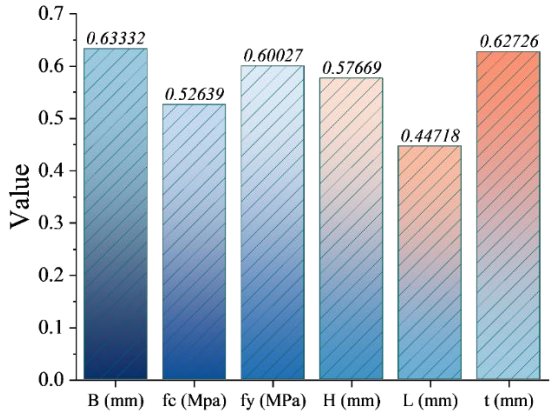


*Fig. 6 The influence of IPs on the ALC drawing on PCC Sensitivity analysis*

### 3.6 K-Fold Cross-Validation (KCV)

KCV is widely deployed by practitioners in model selection and estimation of errors in classifiers. KCV asks for a DB to be split into k portions; later on, via a cycle, a portion of these segments is deployed to train the scheme, while others are deployed to test the scheme's performance. In this study, all the algorithms proposed were enhanced by applying a k-fold strategy with k=5, in which various segments of the data are employed for learning and evaluation processes. Fig. 7 illustrates the process of fold selection, drawing on $R^2$ and RMSE values. Drawing on the figure, the highest precision and least error belong to fold K4, with approximate values of 0.975 and 625 for $R^2$ and RMSE, respectively.

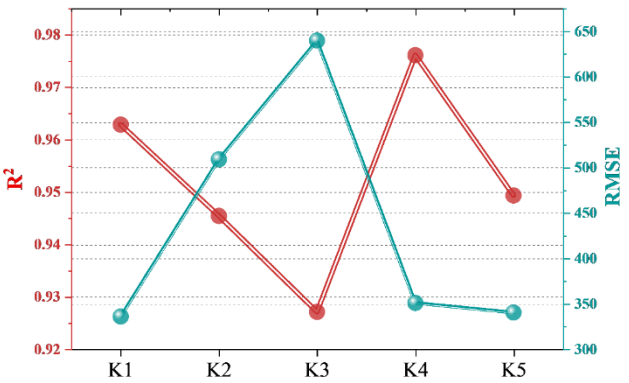


*Fig. 7 The results of 5-fold cross-validation*

## 4. Evaluation Procedures

For evaluation of the results of the predictions, numerical metrics of $R^2$, RMSE, MSE, MRAE, and Residual Standard Error to Standard Deviation Ratio (RSR) have been used (Asgarkhani et al., 2023).

$R^2$ gauges the degree of fluctuation in the result-dependent variable, which is expressed by inputs (independent variables) within a regression framework, and it falls between 0 and 1:

- $R^2$=1: Perfect fit. The scheme describes 100% of the variation in the data.

- $R^2$=0: The scheme describes none of the variability, implying a poor model.

RMSE is one of the most deployed measures to evaluate the precision of a regression model. RMSE quantifies differences between expected and real values, giving an understanding of the typical size of the error in forecasts. The estimate of how well the model's prediction approximates the actual data is particularly facilitated by RMSE, without any alteration to the units of the original data.

This is a commonly used measure in regression analysis that assesses the mean squared difference between the actual and predicted values. By squaring the differences, it emphasizes larger errors compared to smaller ones, making it notably responsive to outliers.

The MRAE is used as a measure to find out the goodness of regression schemes. Usually, the meaning of the observed ones serves as the baseline model, and the average absolute prediction error is calculated relative to its absolute error.

RSR is a statistical measure applied to test the productivity of schemes by examining RSE compared to the observed data SD.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Actual_i - predicted_i)^2}{\sum_{i=1}^{n}(Actual_i - predicted_i)^2} \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Actual_i - predicted_i)^2} \quad (4)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Actual_i - predicted_i)^2 \quad (5)$$

$$MRAE = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|Actual_i - predicted_i\right|}{\left|Actual_i - Actual_{avg}\right|} \quad (6)$$

$$RSR = \frac{RMSE}{\sigma_y} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Actual_i - predicted_i\right)^2}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(Actual_i - Actual_{avg})}} \quad (7)$$

# 5. Results and Discussion

## 5.1 Outcomes of Hyperparameter Tuning and Convergence Curves

Hyperparameters are the options that determine how the ML model is learned. In contrast to model elements, which are trained from data, hyperparameters are determined before training and significantly affect the productivity of the scheme. Various tactics have been employed to adjust hyperparameters and improve model outcomes. In this study, hyperparameter optimization was conducted using the random search technique implemented in Python. This method samples combinations of hyperparameters from predefined ranges and evaluates model performance, offering efficient exploration in high-dimensional parameter spaces. Compared to exhaustive grid search, random search enables broader coverage with fewer evaluations, which is particularly advantageous for complex models such as HGBR and its hybrid variants. Furthermore, more sophisticated hyperparameter tuning tactics, such as Bayesian optimization, GAs, and gradient-based enhancement, can also be employed to further optimize hyperparameter selection. Table 2 illustrates the results of hyperparameter optimization for two hybrid schemes, HGLA (HGBR optimized using the Lotus Effect Optimization Algorithm (LEOA)) and HGEA (HGBR optimized using the Emperor Penguin Optimization Algorithm (EPOA)).

***Table 2. Results of hyperparameters for developed hybrid schemes***

| Optimal Hyperparameters | Schemes | |
|---|---|---|
| | HGLA | HGEA |
| Max_depth | 3 | 1 |
| Min_samples_leaf | 21 | 27 |

For the evaluation of the speed of hybrid schemes in obtaining an optimum solution convergence process, based on $R^2$ and RMSE, has been done for two hybrid schemes of the exploration, HGLA and HGEA, and illustrated in Fig. 8. According to the figure, HGLA takes about 190 cycles to obtain an RMSE of about 200, which is the optimum RMSE, and the cycle number to reach the same level for HGEA is 180, which is slightly lower. In the case of $R^2$, it can be seen that the cycle numbers for obtaining the optimum value for HGLA are slightly higher. In this respect, HGEA indicates a better performance.

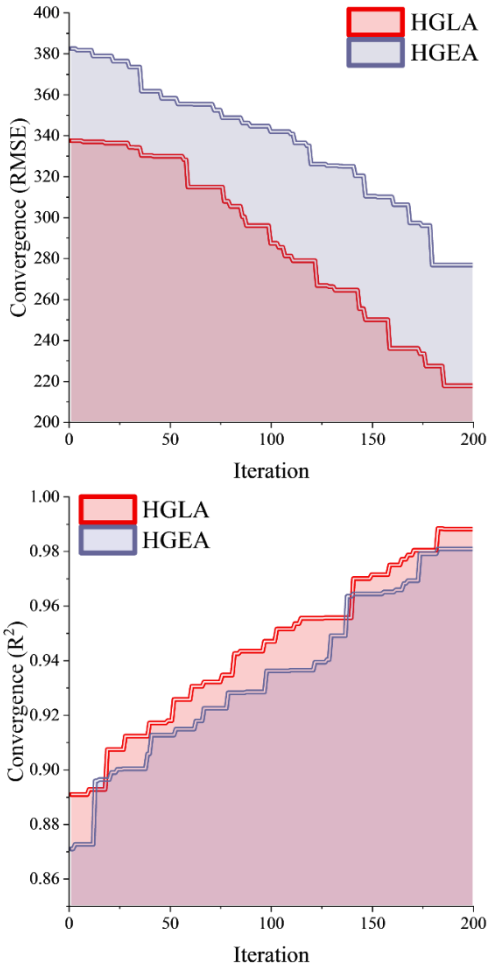


***Fig. 8 Results of the convergence process for hybrid schemes***

## 5.2 Comparison of Schemes' Performance Based on Evaluation Metrics

Table 3 illustrates the results of the five metrics of $R^2$, RMSE, MSE, MRAE, and RSR for a single model of HGBR, hybrid schemes of HGLA (HGBR+LEOA) and HGEA (HGBR+EPOA), and the ensemble approach HGLE (HGBR+LEOA+EPOA). For a more reliable comparison between schemes, the results of the test phase are considered. By exploring test phase results, it is seen that the hybrid model of HGLA with values of 0.9933 and 202.728, respectively, for $R^2$ and RMSE, has the best performance among all schemes. The ensemble approach, HGLE, is the second-best model with lower values for error metrics of MSE, MRAE, and RDR, and higher $R^2$ compared to the single model of HGBR and the hybrid model of HGEA, and the worst performance belongs to the single model of HGBR, with higher error-based values and a higher $R^2$.

***Table 3. The outcomes of the appraisal metrics for the developed schemes***

| Phase | Schemes | Data Metrics | | | | |
|---|---|---|---|---|---|---|
| | | $R^2$ | RMSE | MSE | MRAE | RSR |
| Train | HGBR | 0.9754 | 314.540 | 98935.33 | 0.6479 | 0.1582 |
| | HGLA | 0.9881 | 217.357 | 47244.25 | 0.4480 | 0.1093 |
| | HGEA | 0.9809 | 276.532 | 76469.94 | 0.5699 | 0.1391 |
| | HGLE | 0.9823 | 266.457 | 70999.19 | 0.5500 | 0.1340 |
| Validation | HGBR | 0.9793 | 292.004 | 85266.21 | 0.4521 | 0.1498 |
| | HGLA | 0.9901 | 201.777 | 40713.74 | 0.3124 | 0.1035 |
| | HGEA | 0.9839 | 256.824 | 65958.71 | 0.3976 | 0.1317 |
| | HGLE | 0.9851 | 247.285 | 61150.00 | 0.3829 | 0.1268 |
| Test | HGBR | 0.9858 | 293.439 | 86106.20 | 0.3834 | 0.1194 |
| | HGLA | 0.9933 | 202.728 | 41098.50 | 0.2647 | 0.0825 |
| | HGEA | 0.9891 | 257.942 | 66534.01 | 0.3379 | 0.1050 |
| | HGLE | 0.9899 | 248.478 | 61741.52 | 0.3250 | 0.1011 |

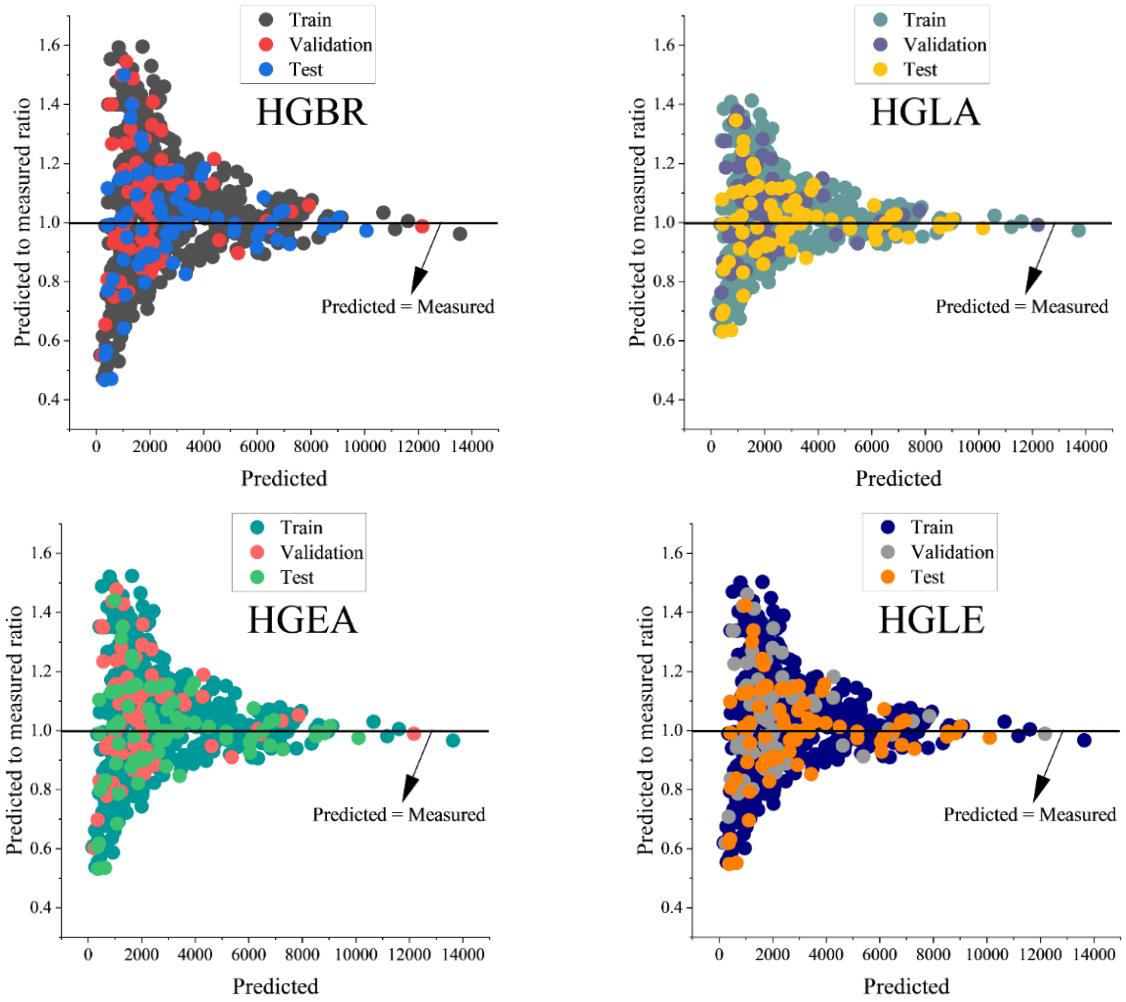


***Fig. 9 Scatter plots for correlation between predicted and measured values of ALC***

Fig. 9 is a representation of four scatter plots of the studied schemes, and they are plots of the values predicted against the ratio of expected to be measured. The nearer an expected value is to the line where expected/measured values equal 1, the more accurate the prediction. According to the figure, almost all the schemes have the same distribution on the plot. For all the schemes, more than seventy percent of the expected values fall between 0 and 4000 MPa. HGLA distribution is the most accurate since the predicted/measured ratio is between 0.6 and 1.4, which is the lowest among all the schemes.

Table 4 illustrates the results of five statistical metrics for prediction schemes in three phases of Validation, Training, and Test, compared with the measured values of the exploration. In the Test phase maximum (Max) and minimum (Min) of measured values are 10357 and 382, respectively. The average and median of expected values by the single model of HGBR are 3136.5 and 2511.31, respectively, and the SD for the ensemble approach HGLE is 2433.87.

Table 5 displays the distribution of errors made by studied schemes via 5 primary statistical metrics. Positive amounts of errors in the maximum (MAX) section show that expected values are bigger than measured, and negative values in the minimum (Min) section indicate expected values being lower than measured ones. In the Test phase, the highest and lowest number of errors belong to a single model of HGBR with values of 49.971 and -53.3398, respectively; the average and median of errors in HGLA are positive, and for the remaining schemes are negative. Furthermore, the lowest SD is for HGLA, 4.7207. Fig. 10 illustrates a box plot for the normal distribution of prediction errors in the test phase. As depicted in Table 5, HGLA has the least SD, meaning that the number of prediction errors is distributed around the average and median. However, for the other three schemes, more dispersity is seen in the plot.

***Table 4. Statistical properties for comparison of schemes***

| Phase | Schemes | Properties | | | | |
|---|---|---|---|---|---|---|
| | | Max | Min | Average | Median | St. Dev |
| Train | Measured | 14116.00 | 275.00 | 2493.01 | 1815.00 | 1988.08 |
| | HGBR | 13562.70 | 151.35 | 2506.58 | 1940.00 | 2003.12 |
| | HGLA | 13732.81 | 189.92 | 2502.40 | 1916.11 | 1993.20 |
| | HGEA | 13629.21 | 166.27 | 2504.93 | 1926.71 | 1998.70 |
| | HGLE | 13647.10 | 170.26 | 2504.49 | 1927.06 | 1997.74 |
| Validation | Measured | 12307.00 | 374.00 | 2180.93 | 1760.00 | 1949.66 |
| | HGBR | 12150.60 | 210.23 | 2259.66 | 1815.55 | 1943.64 |
| | HGLA | 12199.19 | 263.47 | 2235.26 | 1795.61 | 1941.15 |
| | HGEA | 12169.70 | 229.89 | 2250.11 | 1791.86 | 1942.23 |
| | HGLE | 12174.70 | 236.26 | 2247.56 | 1792.82 | 1941.92 |
| Test | Measured | 10357.00 | 382.00 | 3159.26 | 2376.50 | 2457.33 |
| | HGBR | 10068.30 | 319.16 | 3136.50 | 2511.31 | 2432.39 |
| | HGLA | 10157.36 | 407.31 | 3143.45 | 2479.06 | 2436.41 |
| | HGEA | 10102.12 | 358.88 | 3139.34 | 2521.16 | 2433.54 |
| | HGLE | 10112.10 | 366.62 | 3139.94 | 2514.68 | 2433.87 |

*Table 5. Statistical properties of the error of schemes*

| Phase | Schemes | Properties | | | | |
|---|---|---|---|---|---|---|
| | | Max | Min | Average | Median | St. Dev |
| Train | HGBR | 59.5547 | -52.6419 | 0.6282 | 0.6719 | 20.8803 |
| | HGLA | 49.7727 | -46.8824 | 0.7345 | 0.1581 | 5.9004 |
| | HGEA | 52.3364 | -46.2808 | 0.5530 | 0.5827 | 18.3488 |
| | HGLE | 50.2987 | -44.5511 | 0.5311 | 0.5619 | 17.6778 |
| Validation | HGBR | 54.4197 | -44.9649 | 6.4072 | 4.4940 | 19.9244 |
| | HGLA | 41.9966 | -9.7435 | 1.4285 | -0.0460 | 7.6361 |
| | HGEA | 47.6827 | -39.8187 | 5.6241 | 3.9430 | 17.5147 |
| | HGLE | 46.0319 | -38.1513 | 5.4216 | 3.7933 | 16.8697 |
| Test | HGBR | 49.9710 | -53.3398 | -0.9236 | -0.4422 | 18.4143 |
| | HGLA | 13.9508 | -14.8186 | 0.4837 | 0.0224 | 4.7207 |
| | HGEA | 44.0570 | -46.8779 | -0.8028 | -0.3907 | 16.1874 |
| | HGLE | 42.3571 | -45.2037 | -0.7820 | -0.3771 | 15.5930 |

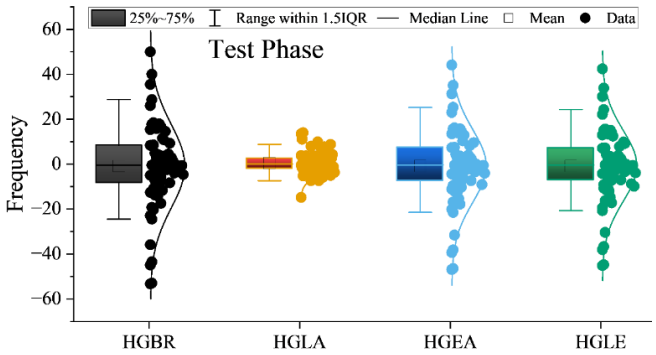


*Fig. 10 Box plot normal distribution for errors of predictive schemes*

# 6. Practical Implications and Limitations

## 6.1 Practical Application

The developed ML-based models, especially the HGLA and HGLE hybrid approaches, offer promising utility for structural engineers involved in the construction and assessment of CFST columns. Practically speaking, engineers can utilize these models to quickly measure the ALBC of R-CFST columns. This removes the necessity for lengthy and computationally demanding FE modeling, or the application of excessively cautious and variable design guidelines. Once trained, the ML models produce predictions in milliseconds, offering a valuable cost- and time-saving tool during preliminary design stages or parametric studies. Moreover, these models improve accuracy by learning directly from experimental data, thus capturing empirical behavior more effectively than analytical formulations in design codes.

## 6.2 Model Boundaries

It is crucial to point out that the trained models are only valid within the input range covered by the dataset. Based on Table 1, the acceptable input ranges are:

- Width (B): 60 mm – 400 mm
- Height (H): 60 mm – 400 mm
- Wall thickness (t): 0.7 mm – 12.5 mm
- Length (L): 60 mm – 1514 mm
- Steel yield strength ($f_y$): 115 MPa – 835 MPa
- Concrete compressive strength (fc): 8.52 MPa – 164.1 MPa

Users must ensure that their IPs fall within these boundaries to avoid extrapolation errors. The model is trained specifically on R-CFST columns; hence, predictions for circular or long-slender columns are outside the intended scope.

## 6.3 Limitations

Although the outcomes suggest strong predictive precision, the model's effectiveness is fundamentally linked to the standard and variety of the training data. Any bias or sparsity in the data, particularly in underrepresented parameter combinations, can reduce the generalizability of the model. Additionally, the ML schemes are data-driven and do not explicitly model complex nonlinear FMs, such as local buckling, bond-slip, or post-peak behavior, which may arise under different loading or boundary conditions. These aspects are typically addressed more thoroughly in FE analysis. Finally, while the PCC-based sensitivity analysis offers some insights into feature importance, the model does not yet incorporate formal explainability tools, which would be valuable in future work to enhance interpretability.

# 7. Conclusion

This study develops interpretable ML models for forecasting the ALBC of R-CFST columns. Utilizing a DB of 719 new records and six geometric and material parameters, a core ML algorithm, HGBR, was enhanced with two bio-inspired optimizers (LEOA and EPOA), as well as a Dempster-Shafer theory-based ensemble strategy. To quantify the results, five statistical metrics are utilized: the coefficient of determination ($R^2$), RMSE, MSE, MRAE, and RSR. Dempster-Shafer theory-based ensemble is also employed to combine the two hybrid schemes to achieve an ensemble approach. PCC is also utilized in performing a sensitivity analysis to illustrate the compressive strength correlation with six of the features extracted in the exploration. Considering the results, it can be concluded that:

1. The findings indicated that the geometry features B, H, and L are strongly correlated among themselves, while feature t is strongly correlated with $f_y$ and $f_c'$. A fairly uniform correlation between ALC and IPs was seen for all the features.
2. Comparative results of prediction schemes based on the five mentioned metrics demonstrate that the best prediction results and best distribution of expected values belong to the HGLA hybrid model, which is a combination of HGBR and LEOA. $R^2$ and RMSE values assigned to HGLA were 0.9933 and 202.728, respectively. Also, the ensemble approach of HGLE (HGBR + LEOA +EPOA) can be mentioned as the second-best model with lower errors and a higher $R^2$ value.
3. Sensitivity assessment is done to recognize the effect of inputs on the output, and outcomes depict that features B and t have the highest coefficients of 0.63332 and 0.62726, respectively, while feature L has the lowest coefficient, indicating the smallest impact on the intended value.

Employing ML to forecast how CFST structures act enhances the accuracy, efficiency, and affordability of structural design and analysis. ML schemes ensure dependable CFST performance predictions with little requirement for complex computations and extensive physical tests. They maximize material utilization, enhance safety analysis, and facilitate real-time monitoring of structural health for early damage detection. With ongoing learning from data, ML improves the precision of designs, minimizes cost, and speeds up the development process. Future research can integrate ML with deep learning, FE analysis, and digital twins to make predictions precise and optimize CFST structures.

# References


Almustafa, M. K. & Nehdi, M. L. (2020). Machine learning model for predicting structural response of RC slabs exposed to blast loading. Engineering Structures, 221, 111109. https://doi.org/10.1016/j.engstruct.2020.111109

Asgarkhani, N., Kazemi, F. & Jankowski, R. (2023). Machine learning-based prediction of residual drift and seismic risk assessment of steel moment-resisting frames considering soil-structure interaction. Computers & Structures, 289, 107181. https://doi.org/10.1016/j.compstruc.2023.107181

Dalirinia, E., Jalali, M., Yaghoobi, M. & Tabatabaee, H. (2024). Lotus effect optimization algorithm (LEA): a lotus nature-inspired algorithm for engineering design optimization. The Journal of Supercomputing, 80(1), 761-799. https://doi.org/10.1007/s11227-023-05513-8

Damico, B. & Conti, M. (2024). Estimation of the Ultimate Bearing Capacity of the Rocks via Utilization of the AI-Based Frameworks. Advances in Engineering and Intelligence Systems, 003(04), 67-81. https://doi.org/10.22034/aeis.2024.488597.1252

Dhiman, G. & Kumar, V. (2018). Emperor penguin optimizer: A bio-inspired algorithm for engineering problems. Knowledge-Based Systems, 159, 20-50. https://doi.org/10.1016/j.knosys.2018.06.001

Fu, F. (2020). Fire induced progressive collapse potential assessment of steel framed buildings using machine learning. Journal of Constructional Steel Research, 166, 105918. https://doi.org/10.1016/j.jcsr.2019.105918

Hatzigeorgiou, G. D. (2008). Numerical model for the behavior and capacity of circular CFT columns, Part I: Theory. Engineering Structures, 30(6), 1573-1578. https://doi.org/10.1016/j.engstruct.2007.11.001

Institution, B. S. (1982). Steel, concrete, and composite bridges. BSI.

Jeon, J., Shafieezadeh, A. & DesRoches, R. (2014). Statistical models for shear strength of RC beam-column joints using machine-learning techniques. Earthquake Engineering & Structural Dynamics, 43(14), 2075-2095. https://doi.org/10.1002/eqe.2437

Johnson, R. P. & Anderson, D. (2001). EN1994 Eurocode 4: Design of composite steel and concrete structures. Proceedings of the Institution of Civil Engineers-Civil Engineering, 144(6), 33-38. https://doi.org/10.1680/cien.2001.144.6.33

Keshtegar, B., Nehdi, M. L., Kolahchi, R., Trung, N.-T. & Bagheri, M. (2022). Novel hybrid machine leaning model for predicting shear strength of reinforced concrete shear walls. Engineering with Computers, 1-12. https://doi.org/10.1007/s00366-021-01302-0

Khajavi, E., Taghavi Khanghah, A. R. & Javadzade Khiavi, A. (2025). An efficient prediction of punching shear strength in reinforced concrete slabs through boosting methods and metaheuristic algorithms. Structures, 74, 108519. https://doi.org/10.1016/j.istruc.2025.108519

Khan, M., Uy, B., Tao, Z. & Mashiri, F. (2017). Behaviour and design of short high-strength steel welded box and concrete-filled tube (CFT) sections. Engineering Structures, 147, 458-472. https://doi.org/10.1016/j.engstruct.2017.06.016

Kodur, V. K. & Naser, M. Z. (2021). Classifying bridges for the risk of fire hazard via competitive machine learning. Advances in Bridge Engineering, 2(1), 2. https://doi.org/10.1186/s43251-020-00027-2

Lai, Z. & Varma, A. H. (2015). Noncompact and slender circular CFT members: Experimental database, analysis, and design. Journal of Constructional Steel Research, 106, 220-233. https://doi.org/10.1016/j.jcsr.2014.11.005

Li, S. (2024). Estimating Stock Market Prices with Histogram-based Gradient Boosting Regressor: A Case Study on Alphabet Inc. International Journal of Advanced Computer Science & Applications, 15(5). https://doi.org/10.14569/IJACSA.2024.0150553

Liew, J. Y. R., Xiong, M. & Xiong, D. (2016). Design of concrete filled tubular beam-columns with high strength steel and concrete. Structures, 8, 213-226. https://doi.org/10.1016/j.istruc.2016.05.005

Liu, D. & Gho, W.-M. (2005). Axial load behaviour of high-strength rectangular concrete-filled steel tubular stub columns. Thin-Walled Structures, 43(8), 1131-1142. https://doi.org/10.1016/j.tws.2005.03.007

Mangalathu, S., Hwang, S.-H. & Jeon, J.-S. (2020). Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach. Engineering Structures, 219, 110927. https://doi.org/10.1016/j.engstruct.2020.110927

Mangalathu, S., Jang, H., Hwang, S.-H. & Jeon, J.-S. (2020). Data-driven machine-learning-based seismic failure mode identification of reinforced concrete shear walls. Engineering Structures, 208, 110331. https://doi.org/10.1016/j.engstruct.2020.110331

Mangalathu, S. & Jeon, J.-S. (2018). Classification of failure mode and prediction of shear strength for reinforced concrete beam-column joints using machine learning techniques. Engineering Structures, 160, 85-94. https://doi.org/10.1016/j.engstruct.2018.01.008

Mangalathu, S. & Jeon, J.-S. (2019). Machine learning-based failure mode recognition of circular reinforced concrete bridge columns: Comparative study. Journal of Structural Engineering, 145(10), 4019104. https://doi.org/10.1061/(ASCE)ST.1943-541X.0002402

Mangalathu, S., Sun, H., Nweke, C. C., Yi, Z. & Burton, H. V. (2020). Classifying earthquake damage to buildings using machine learning. Earthquake Spectra, 36(1), 183-208. https://doi.org/10.1177/8755293019878137

Memarzadeh, A., Sabetifar, H. & Nematzadeh, M. (2023). A comprehensive and reliable investigation of axial capacity of Sy-CFST columns using machine learning-based models. Engineering Structures, 284, 115956. https://doi.org/10.1016/j.engstruct.2023.115956

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

Naeim, B., Akbarzadeh, M. R. & Jahangiri, V. (2024). Machine learning-based prediction of seismic response of elevated steel tanks. Structures, 70, 107649. https://doi.org/10.1016/j.istruc.2024.107649

Naeim, B., Khiavi, A. J., Dolatimehr, P. & Sadaghat, B. (2024). Novel Optimized Support Vector Regression Networks for Estimating Fresh and Hardened Characteristics of SCC.

Naser, M. Z., Thai, S. & Thai, H.-T. (2021). Evaluating structural response of concrete-filled steel tubular columns through machine learning. Journal of Building Engineering, 34, 101888. https://doi.org/10.1016/j.jobe.2020.101888

Olalusi, O. B. & Awoyera, P. O. (2021). Shear capacity prediction of slender reinforced concrete structures with steel fibers using machine learning. Engineering Structures, 227, 111470. https://doi.org/10.1016/j.engstruct.2020.111470

Rahman, J., Ahmed, K. S., Khan, N. I., Islam, K. & Mangalathu, S. (2021). Data-driven shear strength prediction of steel fiber reinforced concrete beams using machine learning approach. Engineering Structures, 233, 111743. https://doi.org/10.1016/j.engstruct.2020.111743

Sakino, K., Nakahara, H., Morino, S. & Nishiyama, I. (2004a). 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)

Sakino, K., Nakahara, H., Morino, S. & Nishiyama, I. (2004b). 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)

Seo, J., Dueñas-Osorio, L., Craig, J. I. & Goodno, B. J. (2012). Metamodel-based regional vulnerability estimate of irregular steel moment-frame structures subjected to earthquake events. Engineering Structures, 45, 585-597. https://doi.org/10.1016/j.engstruct.2012.07.003

Shafer, G. (2016). Dempster's rule of combination. International Journal of Approximate Reasoning, 79, 26-40. https://doi.org/10.1016/J.IJAR.2015.12.009

Solhmirzaei, R., Salehi, H., Kodur, V. & Naser, M. Z. (2020). Machine learning framework for predicting failure mode and shear capacity of ultra high-performance concrete beams. Engineering Structures, 224, 111221. https://doi.org/10.1016/j.engstruct.2020.111221

Specification for Structural Steel Buildings (ANSI/AISC 360-16) - 2016 | American Institute of Steel Construction. (n.d.). Retrieved March 8, 2025, from https://www.aisc.org/Specification-for-Structural-Steel-Buildings-ANSIAISC-360-16-Download

Tran, V.-L., Thai, D.-K. & Kim, S.-E. (2019). Application of ANN in predicting ACC of SCFST column. Composite Structures, 228, 111332. https://doi.org/10.1016/j.compstruct.2019.111332

Tran, V.-L., Thai, D.-K. & Nguyen, D.-D. (2020). Practical artificial neural network tool for predicting the axial compression capacity of circular concrete-filled steel tube columns with ultra-high-strength concrete. Thin-Walled Structures, 151, 106720. https://doi.org/10.1016/j.tws.2020.106720

Uy, B. (2001). Strength of short concrete filled high strength steel box columns. Journal of Constructional Steel Research, 57(2), 113-134. https://doi.org/10.1016/S0143-974X(00)00014-6

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

Xiong, M.-X., Xiong, D.-X. & Liew, J. Y. R. (2017). Axial performance of short concrete filled steel tubes with high-and ultra-high-strength materials. Engineering Structures, 136, 494-510. https://doi.org/10.1016/j.engstruct.2017.01.037

Zarringol, M., Thai, H.-T., Thai, S. & Patel, V. (2020). Application of ANN to the design of CFST columns. Structures, 28, 2203-2220. https://doi.org/10.1016/j.istruc.2020.10.048

Zhao, K., Li, L., Chen, Z., Sun, R., Yuan, G. & Li, J. (2022). A survey: Optimization and applications of evidence fusion algorithm based on Dempster-Shafer theory. Applied Soft Computing, 124, 109075. https://doi.org/10.1016/j.asoc.2022.109075

Zhong, S. T. (2006). Unified theory of CFST: research and application. Tsinghua University Press Beijing, China.

Zhou, H., Deng, Z., Xia, Y. & Fu, M. (2016). A new sampling method in particle filter based on Pearson correlation coefficient. Neurocomputing, 216, 208-215. https://doi.org/10.1016/j.neucom.2016.07.036

## Disclaimer