



Development of effective XGB model to predict the Axial Load Capacity of circular CFST columns

Indra Prakash¹, Raghvendra Kumar², Thuy-Anh Nguyen^{3,*}, Phuong-Thao Vu⁴

¹Dy. Director General (R), Geological Survey of India, Gandhinagar 82010, India.

²Department of Computer Science and Engineering, GIET University, Gunupur-765022, India.

³University of Transport Technology, Hanoi 100000, Vietnam.

⁴University of Transport and Communications, Hanoi 100000, Vietnam.

Article info

Type of article:

Original research paper

DOI:

<https://doi.org/10.58845/jstt.utt.2022.en.2.4.26-42>

*Corresponding author:

E-mail address:

anhnt@utt.edu.vn

Received: 05/12/2022

Revised: 20/12/2022

Accepted: 22/12/2022

Abstract: The Axial Load Capacity (ALC) of Concrete-Filled Steel Tubular (CFST) structural members is regarded as one of the most crucial technical factors for the design of these composite structures. This work proposes the development and application of the Extreme Gradient Boosting (XGB) model to forecast the ALC of circular CFST structural components using the affecting input parameters, namely column diameter, steel tube thickness, column length, steel yield strength, and concrete compressive strength. A dataset of 2073 experimental results from the literature was used for the model development. The performance of the XGB model was evaluated using statistical criteria such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Coefficient of Determination (R^2), and Mean Absolute Percentage Error (MAPE). The five-fold cross-validation technique and Monte Carlo simulation method were used to evaluate the model's performance. The results show good performance of the XGB model ($R^2 = 0.999$, RMSE = 242.757 kN, MAE = 157.045 kN, and MAPE = 0.057) in predicting the circular CFST's ALC.

Keywords: Concrete-filled steel tube; axial load capacity; machine learning, Extreme gradient boosting.

1. Introduction

Concrete-Filled Steel Tube (CFST) columns are a type of composite structure made of hollow steel tubes filled with concrete. Because of many advantages over hollow steel columns and reinforced concrete columns [1–4], this type of structure is prevalent in modern construction. These advantages include high axial bearing capacity, good ductility and strength, large energy absorption capacity, convenient construction, material savings, and high fire resistance [5–7]. In addition, because there is no need for formwork, the construction process is quicker. It also costs

less to construct and they are more environmentally friendly because steel tubes can be reused along with recycled aggregates in concrete [8–10]. According to several studies [11,12], CFST columns exhibit excellent efficiency under compression. As a result, the cross-section of the chosen CFST column is frequently symmetrical, such as a circular, square, or rectangle. The circular CFST column is the most often utilized due to its excellent confinement performance, higher stiffness, and yield strength [13–15].

Numerous investigations have been

conducted over the past decades to assess Axial Load Capacity (ALC) and CFST columns' behavior. Several experiments have been performed on CFST circular columns, including examination of the effects of loads, the strength of concrete [16,17], the diameter-to-thickness ratio of the tube [18,19], or bond action among steel tubes and concrete [4,20]. The first contribution was Knowles and Park's work [21], carried out in the late 1960s to assess the behavior of CFST columns under eccentric and centered loads. In addition, the behavior of CFST columns under cyclic dynamic loads is evaluated in a study by Liu and Goel [22]. In another attempt, the impact of employing high-strength concrete in CFST columns is investigated by Kilpatrick and Rangan [23]. On 114 CFST columns, Sakino et al. [24] investigated the effects of steel pipe shape and strength, tube diameter to thickness ratio, and concrete strength and proposed design formulae to determine their ultimate ALC. It is worth noting that in the literature, many works have attempted to compile the outcomes of these investigations into different databases. However, obtaining the long time data is the main challenge, as it needs a significant investment in terms of funds, high-end test equipment systems, and a considerable amount of time and labor.

The behavior of CFST columns under axial compression is also studied using numerical modeling. For instance, to model compressive CFST stub columns, Dai et al. [25] used the Finite Element Model (FEM) created by an ABAQUS solver. Choi et al. [26] put out a numerical approach to examine the axial behavior of CFST columns and estimate various interactions between the steel tube and concrete. However, the models lack the ability to estimate the behavior of these members with an appropriate level of precision since it is difficult to consider all the complicated boundary conditions and mechanical characteristics of the material in numerical simulations [26].

In addition, CFST column calculation provisions have been suggested in published

design standards such as EC4 [27], ACI [28], AISC [29], and AS/NZS 2327 [30]. Their usefulness is, however, limited to CFST columns with a certain section slenderness ratio and material grade. Due to their restricted applicability, none of the above-mentioned methods have been extensively adopted. Therefore, creating a standardized and precise procedure for designing circular CFST columns is necessary.

In recent years, with the rapid advancement of computer science, Machine Learning (ML) techniques have become pervasive in all scientific domains, including Civil Engineering. ML approaches are methods that construct complicated mathematical models with great precision to reflect the connection between the input and output parameters of a given data set. Based on this perspective, numerous scientists currently utilize ML to identify the structures' behavior [32–38]. The application of ML to forecast the ALC of circular CFST columns has also been the subject of substantial research [31–35]. Specifically, Ahmadi et al. propose the ANN model to estimate the ALC of the circular CFST column under the effect of axial load based on a dataset of 268 experimental results and obtain a forecast performance of $R = 0.899$. In the study by Sarir et al., a gene expression program (GEP) is developed using 303 experimental results and five input parameters to estimate the ALC of the circular CFST column. The best predictive model is selected with model performance $R^2 = 0.939$ and $RMSE = 606.28$ kN. In a recent study by Liu et al., a PSO-ANN hybrid model consisting of an artificial neural network (ANN) optimized using a particle swarm algorithm (PSO) has been proposed to predict the ALC of a circular CFST column with a dataset of 227 experimental results. The model's performance is equivalent to $R = 0.989$. The above studies have shown that machine learning algorithms are powerful numerical tools to predict the ALC of circular CFST columns. However, these studies have not evaluated the effect of input factors on the ALC of columns nor considered a limited amount of data. Moreover, the predictive

potential of these investigations needs additional development.

As a result, this research aims to propose an ML model, the Extreme Gradient Boosting (XGB) model, to predict the ALC of circular CFST columns. A data set of experimental findings containing 2073 circular CFST column samples was employed to train and test the developed model. The database includes parameters such as the structural members' geometry and the component materials' mechanical properties. This dataset is the largest ever created, providing solid results when training and testing the model. Simultaneously, feature importance and sensitivity analysis are studied using one-dimensional partial dependence plots (PDP). The results of the model were evaluated using standard statistical measures, namely Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Coefficient of Determination (R^2), and Mean Absolute Percentage Error (MAPE).

2. Database description and analysis

In this study, the 2073 data points on circular CFST columns studies are collected from the published literature, including 1305 data from two well-known databases of Denavit [36] and Goode [37], and 768 finite element results of high-strength concrete columns from Tran's study [38]. Fig. 1 depicts the experimental setup to determine the ALC of CFST columns in general. Initial imperfections in column geometry and residual stresses during member production are disregarded and not considered input parameters in this database due to their insignificant effect on the CFST column [39]. For each CFST specimen, several geometric and material parameters are gathered. The geometric characteristics consist of the physical parameters of CFST columns, i.e., column length (L), tube thickness (t), and tube diameter (D). The material properties include steel yield strength (f_y) and concrete compressive strength (f_c). Table 1 presents the primary material and geometric characteristics of the collected database. Notably, the concrete compressive

strength determined by the tests is based on both cylinder and cube specimens, and the cube strength is converted into cylinder strength for use in calculations. Table 1 shows that the cross-section diameter ranges from 44.5 mm to 1020 mm, with an average value of 264.87 mm and a standard deviation of 176.58 mm. The thickness of the steel tube ranges from 0.52 mm to 30 mm, with an average of 8.38 mm and a standard variation of 6.75 mm. The length of the member spans from 152.35 mm to 5560 mm, with an average of 1658.31 mm and a standard deviation of 1287.19 mm. Steel tube yield strength ranges from 178.28 MPa to 1153 MPa, with an average value of 342.59 MPa and a standard variation of 105.59 MPa. The compressive strength of concrete ranges from 7.01 MPa to 200 MPa, with an average value of 84.79 MPa and a standard variation of 57.79 MPa. The observed axial load varies from 45.2 to 75194.86 kN, with an average value of 12574.56 kN and a standard deviation of 16560.77 kN.

Fig. 2 depicts histograms of inputs and output parameters used in this study. In addition, a correlation study between input and output variables is also carried out to investigate the linear statistical correlation between the variables in the database. The Pearson technique is used to calculate the correlation coefficient R . Fig. 3 depicts the correlation matrix between the pairs of parameters, in which the bottom triangle reflects the correlation coefficient value and the top triangle depicts the correlation based on the intensity and size of the circles. The diagonal represents the connection between the variables. Tabachnick et al. [40] define strongly correlated parameter pairs as having an absolute value of R greater than 0.75. The greatest absolute value of R in the gathered input space is 0.74, indicating that it is suitable to use the existing input space to create the ML model in this study.

The dataset is randomly divided into two sub-datasets, including the first part (70% of the data) used to train the model, called the training part. The second part (the remaining 30% of data) is used to

verify the model, called the testing part. This split ratio is chosen to ensure efficiency during training

and testing, as the relevant literature suggested [41].

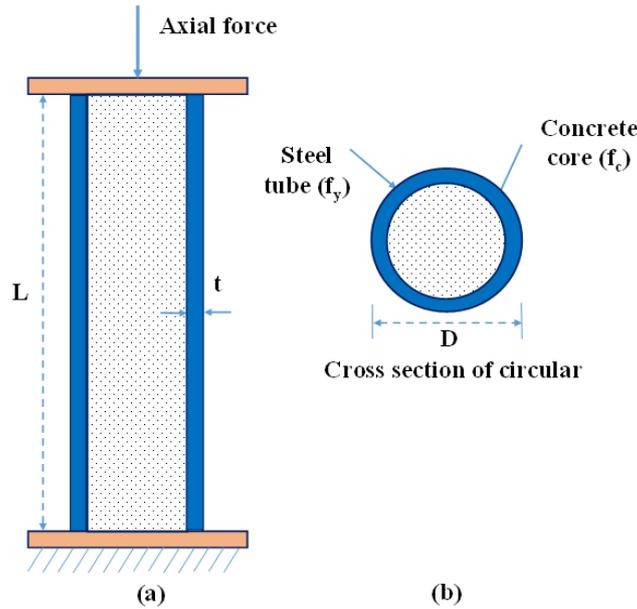
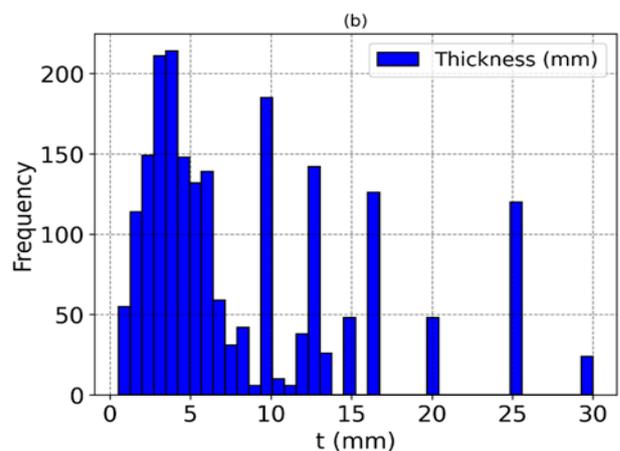
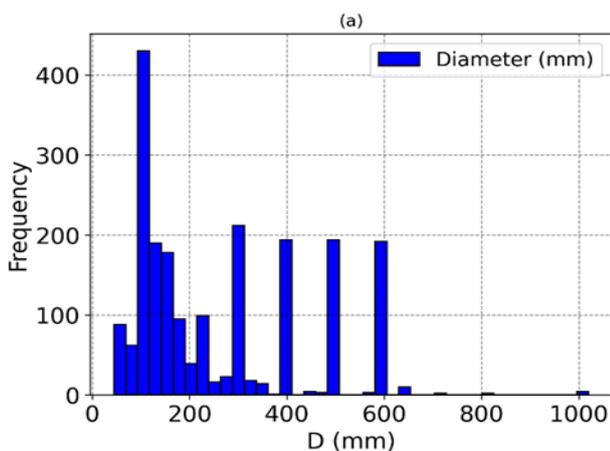


Fig 1. Schematic diagram showing experimental set up of (a) the CFST columns under axial force, (b) the cross-section of circular column

Table 1. Statistical characteristics of the input and output parameters in the database.

Parameter	Unit	Mean	Std	Min	25%	50%	75%	Max
Input								
Diameter of tube (D)	mm	264.87	176.58	44.45	114.30	190.70	400.00	1020.00
Thickness of steel tube (t)	mm	8.38	6.75	0.52	3.35	5.84	12.5	30.00
Length of column (L)	mm	1658.3	1287.1	152.35	661.50	1200.0	2400.0	5560.00
Yield strength of steel tube (f_y)	MPa	342.59	105.59	178.28	275.00	332.02	374.00	1153.00
Compression strength concrete (f_c)	MPa	84.79	57.79	7.01	34.85	59.00	140.00	200.00
Output								
ALC (P_u)	kN	12574.	16560.	45.20	945.00	2500.6	21048.	75194.8

Std=Standard deviation;



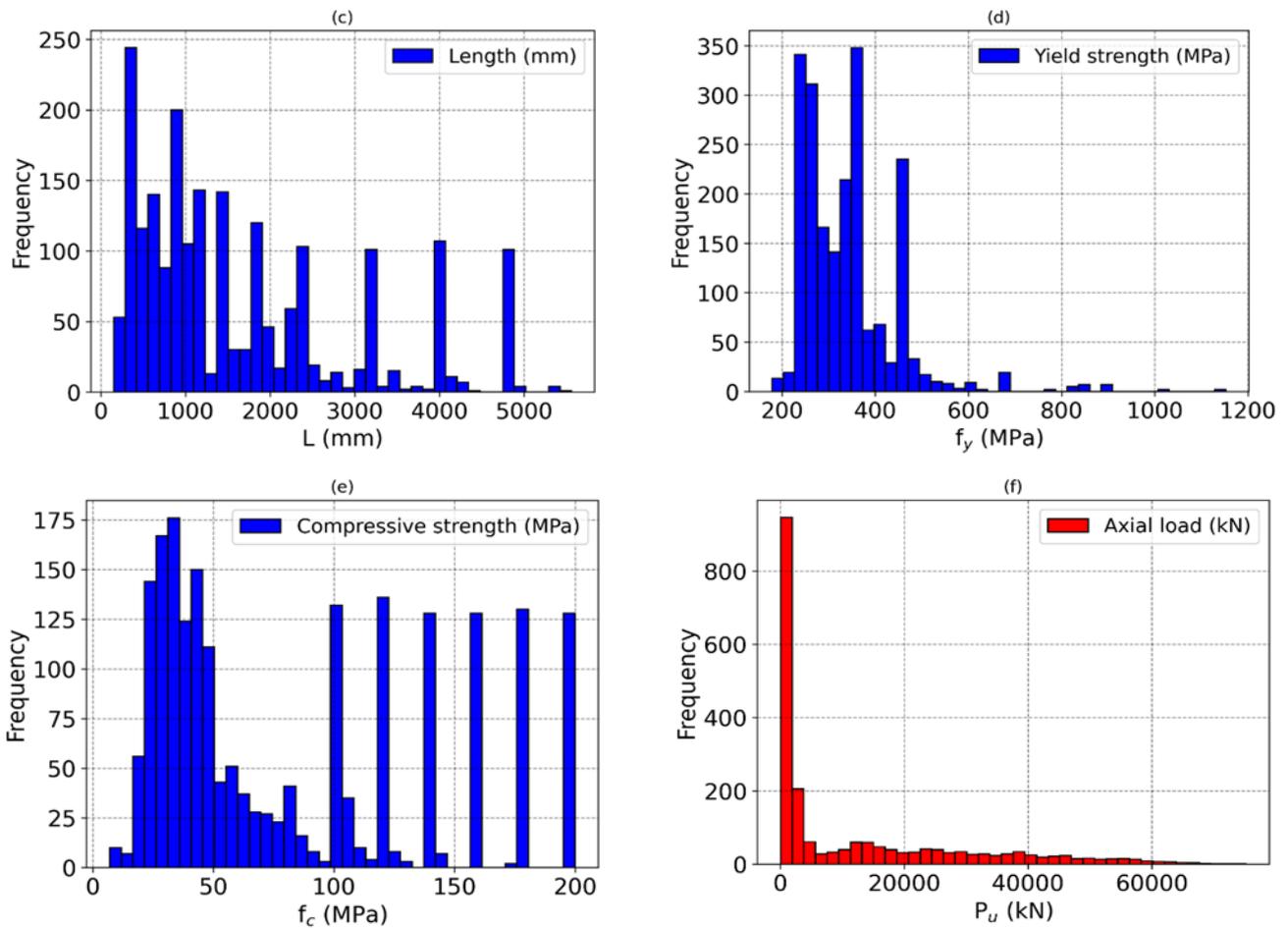


Fig 2. Histograms of the parameters used in the database for the model study

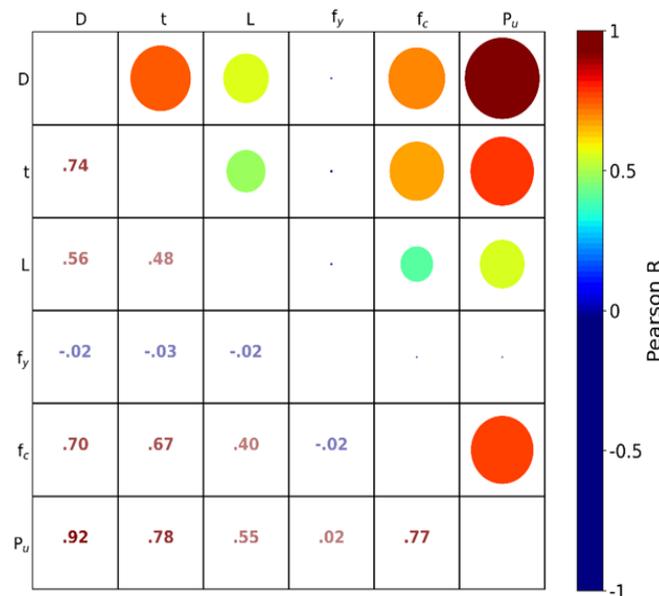


Fig 3. Correlation of input and output parameters of the database

3. Method used

3.1. Machine learning methods

In this study, the Extreme Gradient Boosting (XGB), is an ensemble machine-learning technique

that Chen and Guestrin created in 2016 [42], has been used for the prediction of ALC. This approach is an improved gradient-boosting decision tree algorithm that aims to produce high accuracy with

little chance of overfitting. In addition to effectively building gradient-boosting trees, XGB can also tackle regression and classification issues while operating in parallel. This is an efficient and easy-to-use algorithm that delivers high performance and accuracy as compared to other algorithms. The XGB model introduces a component to the loss function as an enhancement over the gradient boosting approach. As such, the loss function of the XGB model takes the form:

$$Obj(\Phi) = L(\Phi) + \Omega(\Phi) \tag{1}$$

where L is the loss function in the training process, Ω is the rule set of the decision tree. The loss function or error rate is used to measure the model's performance during training (model building). Rule sets are used to control the complexity of the model and avoid redundancy or lack of information in the data. There are different methods used to determine the complexity of the model. However, the complexity of each decision tree is usually determined by the following formula:

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

where T is the number of leaves on the decision tree, ω is the vector of scores on the leaves (of the decision tree). The core of the XGB algorithm is the objective function, which is determined by the following equation:

$$l = \sum_{j=1}^T \left[G_j \omega_j + \frac{1}{2} (H_j + \lambda) \omega_j^2 \right] + \gamma T \tag{3}$$

where ω_j are independent variables. The goal of real XGB math is to minimize the *Obj* function so that the error rate is minimal.

3.2. Performance indices of models

The RMSE, MAE, R^2 , and MAPE are the performance metrics used in this study to assess the effectiveness and precision of the XGB model in predicting the ALC of the CFST column. The following formulas are used to determine these performance indicators:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_{tt,i} - Y_{db,i})^2} \tag{4}$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |Y_{tt,i} - Y_{db,i}| \tag{5}$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{Y_{tt,i} - Y_{db,i}}{Y_{tt,i}} \right| \times 100\% \tag{6}$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (Y_{tt,i} - Y_{db,i})^2}{\sum_{i=1}^N (Y_{tt,i} - \bar{Y}_{tt})^2} \tag{7}$$

where $Y_{tt,i}$ is the actual test value of the i^{th} sample, \bar{Y}_{tt} is the average of the actual test values, $Y_{db,i}$ is the predicted value corresponding to the i^{th} sample, calculated according to the model's predictions, whereas N denotes the sample numbers. Out of these, the optimum value of performance metrics is achieved when RMSE, MAE, and MAPE are equal to 0, whereas R^2 is equal to 1.

4. Results and Discussion

4.1. Hyperparameter tuning of the XGB model

The model's parameters are the most important notions in machine learning, and the training model identifies the optimal ones for improved performance. The parameters are separated into model parameters and hyperparameters, with the hyperparameters requiring setting before training the ML model because they determine its architecture. Tuning hyperparameters is crucial since they enhance learning models' performance. In this context, an extensive trial and error test is used to adjust the XGB model's hyperparameters. Four hyperparameters, n estimators, max depths, learning rate, and minimum child weight, that have a significant impact on the XGB algorithm are chosen for tweaking based on past research [43,44]. Table 2 presents the associated search domains for hyperparameters. The XGB package utilizes the default values for the remaining hyperparameters in Python.

To prevent overfitting and increase the reliability of the training operations, a 5-fold Cross-Validation (CV) methodology is used. The training

dataset is randomly split into five equal folds. The cross-validation training is then carried out by selecting four folds randomly to serve as the training component and the remaining fold as the validation part. The performance metrics are calculated using the mean after training and verifying five times. The validation dataset created by such 5-fold CV method tests and compares the model's performance throughout the training phase. The performance of the model is assessed using RMSE criteria. When the model has the lowest RMSE, the optimal hyperparameters for the XGB model are selected.

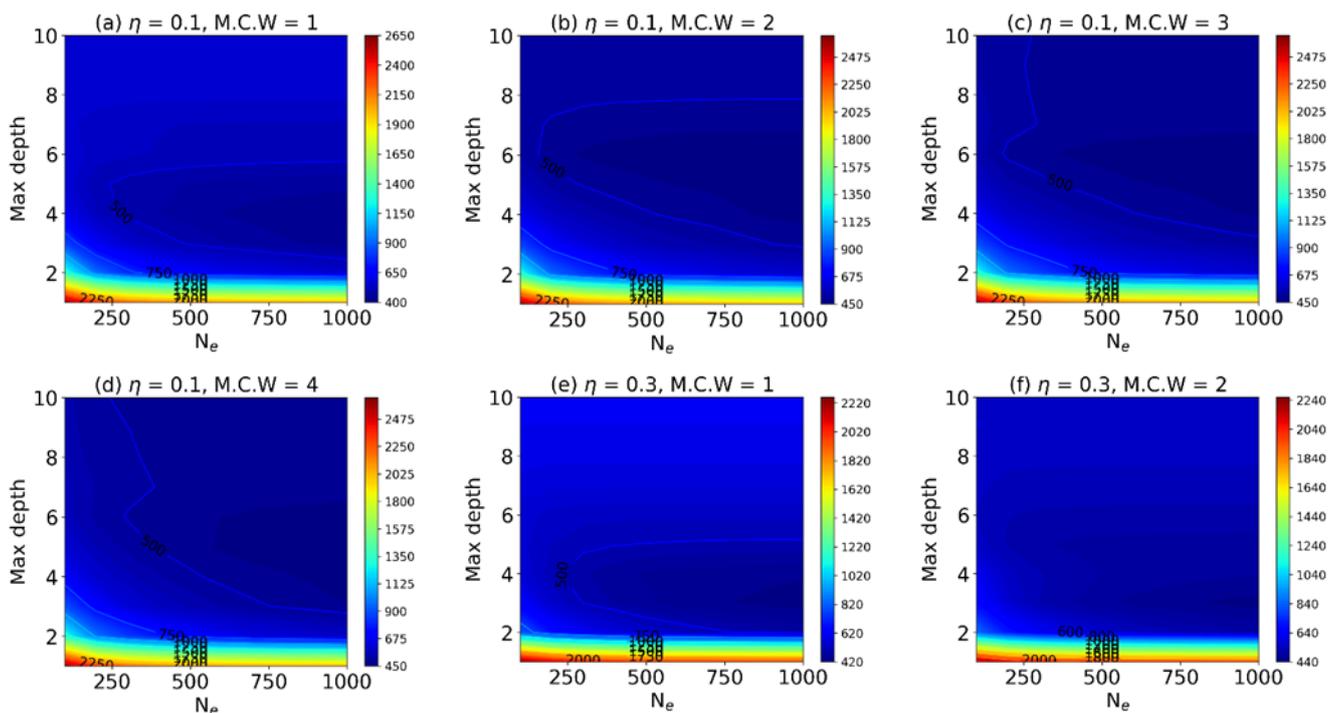
Fig. 4 illustrates the four hyperparameters' effect on the model's performance on the validation dataset. It can be observed that the lowest RMSE value achieved is 400 kN, hence, this RMSE value may be used to identify the optimal XGB model. An ML model, on the other hand, is termed stable if it performs well on both the validation and test data sets. As a result, five models performing well on the validation dataset are chosen for further

examination to establish the best predictive model. These five models are carefully tested on the test dataset to compare their predicting ability. Table 3 presents the hyperparameters of the five models that are chosen. The performance of five ML models is assessed using standard statistical criteria.

Fig. 5 depicts the models' performance using two criteria, RMSE and R^2 . In addition to the validation dataset, the training and validation datasets are utilized to compare model performance. Table 4 shows the detailed values of the statistical criteria relating to the 5 XGB models. Based on these values, it is obvious that the XGB-05 model outperforms the others, including both validation and test datasets. Interestingly, this model does not have the best validation score on the validation set. The other four models perform well, but are less accurate on the testing set. As a consequence, XGB-05 is chosen as the best model, and typical results are shown in the following section.

Table 2. Search domain of 4 hyperparameters tuned in XGB model

n_estimators (N_e)	learning_rate (η)	max_depth	min_child_weight (M.C.W)
100 - 1000	0.1, 0.3, 0.5	1 - 10	1, 2, 3, 4



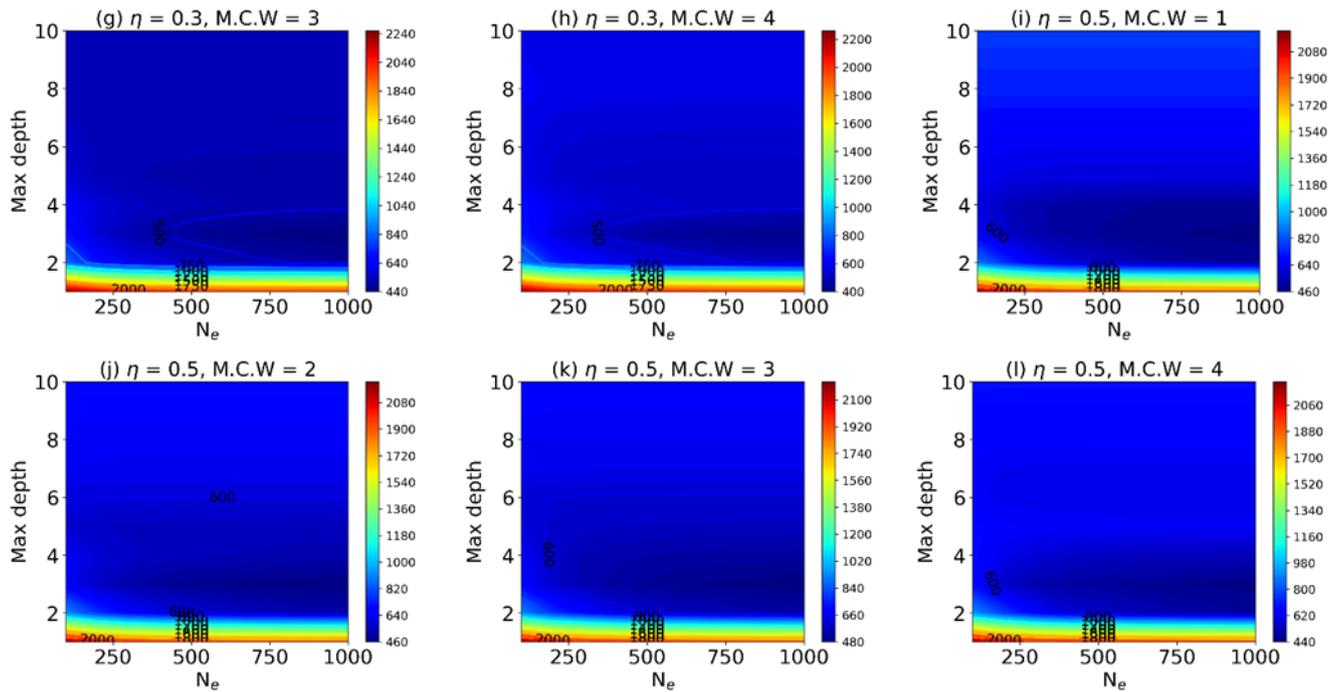


Fig 4. RMSE validation scores of different XGB models

Table 3. Optimal hyperparameters of 5 XGB models

Model	XGB_01	XGB_02	XGB_03	XGB_04	XGB_05
n_estimators (N_e)	700	800	1000	800	1000
learning_rate	0.1	0.3	0.3	0.3	0.3
max_depth	4	3	3	3	3
min_child_weight	1	4	1	3	2

Table 4. Summary of prediction results of 5 XGB models.

	RMSE (kN)	MAE (kN)	R ²	MAPE
XGB_01				
5-fold CV score	424.002	216.145	0.999	0.068
Training part	97.252	60.993	0.999	0.019
Testing part	329.212	186.995	0.999	0.059
XGB_02				
5-fold CV score	419.678	223.866	0.999	0.090
Training part	103.464	64.804	0.999	0.022
Testing part	309.067	182.719	0.999	0.066
XGB_03				
5-fold CV score	430.664	214.422	0.999	0.078
Training part	99.006	61.577	0.999	0.019
Testing part	281.085	176.423	0.999	0.066
XGB_04				
5-fold CV score	442.268	218.178	0.999	0.090
Training part	122.477	77.251	0.999	0.026
Testing part	300.474	191.672	0.999	0.073
XGB_05				
5-fold CV score	451.234	216.891	0.999	0.090
Training part	97.903	61.212	0.999	0.019
Testing part	277.050	173.223	0.999	0.066

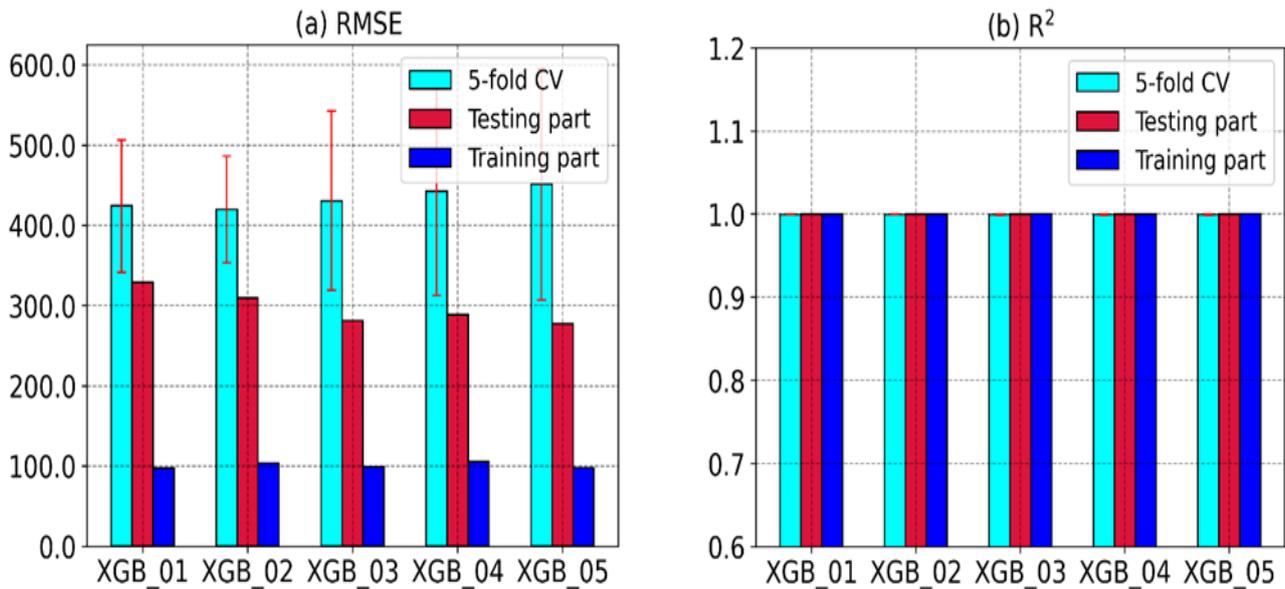


Fig 5. Performance comparison of 5 XGB models

4.2. Representative prediction results

This section presents the typical outcomes of 200 Monte Carlo simulations of the XGB 05 model. The Monte Carlo simulation aims to generate several separate training and testing parts so that the model's generalizability can be evaluated [45].

The comparison between the predicted value and the actual ALC of the circular CFST column is depicted in Fig. 6 under the regression graph, in which Fig. 6a represents the training part, and Fig. 6b represents the testing part. The horizontal axis reflects the ALC based on the circular CFST column's actual values, while the vertical axis represents the predicted values. The solid black line shows the diagonal, whereas the black dotted lines and blue dashed lines represent 10% and 20% error contours, respectively. Most data points lie along the $y=x$ line, indicating that the suggested model achieves near-absolute performance in the training and testing parts. The robust prediction ability of the XGB_05 model could thus be demonstrated using the results obtained in such

regression analysis.

In addition, the distribution histogram and cumulative distribution of the error between the predicted and actual ALC of the circular CFST columns for the training part (Fig. 7a) and testing part (Fig. 7b) clearly show this excellent prediction ability. Among the 1451 training and 623 testing data samples, there is a significant error value of around 0. Only six samples in the training part (representing 0.4% of the training part) have errors beyond the range [-300, 300] kN, which is an insignificant proportion compared to the total number of samples. The testing part has a greater error with a maximum error value of 1500 kN with only two samples. In addition, the quantitative values of the XGB_05 model performance evaluation criteria are provided in Table 5. It can be observed that the model's prediction ability is robust. Therefore, using the XGB_05 model to accurately predict the ALC of circular CFST columns is conceivable, thereby saving time and cost on trials.

Table 5. Statistical criteria values for typical results of XGB_05 model

	RMSE (kN)	MAE (kN)	R ²	MAPE
Training part	90.086	56.502	0.999	0.018
Testing part	242.757	157.045	0.999	0.057
All dataset	152.848	86.670	0.999	0.030

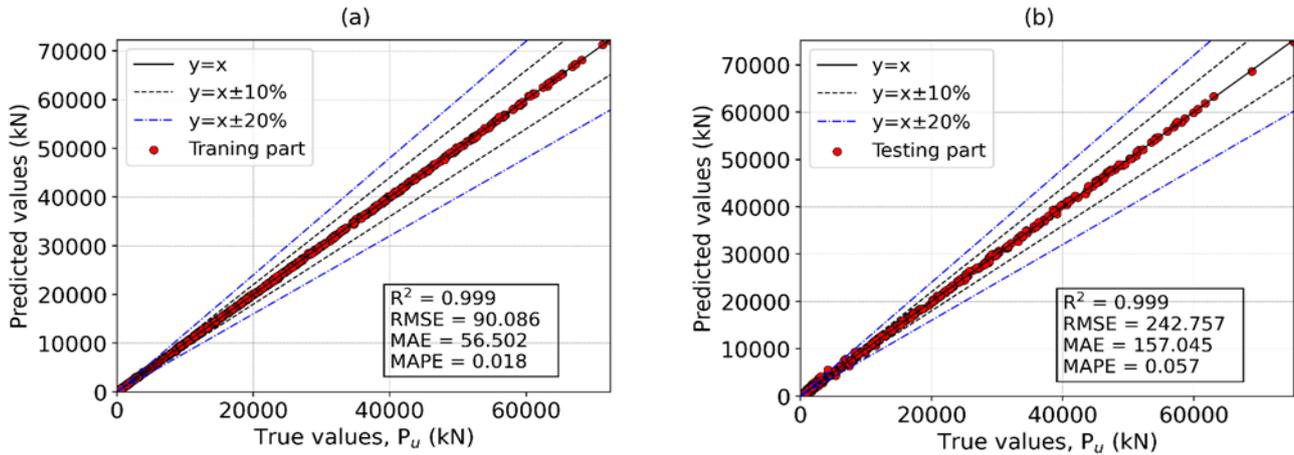


Fig 6. Regression charts between experimental and true values of ALC of circular CFST columns for the training part; and testing part

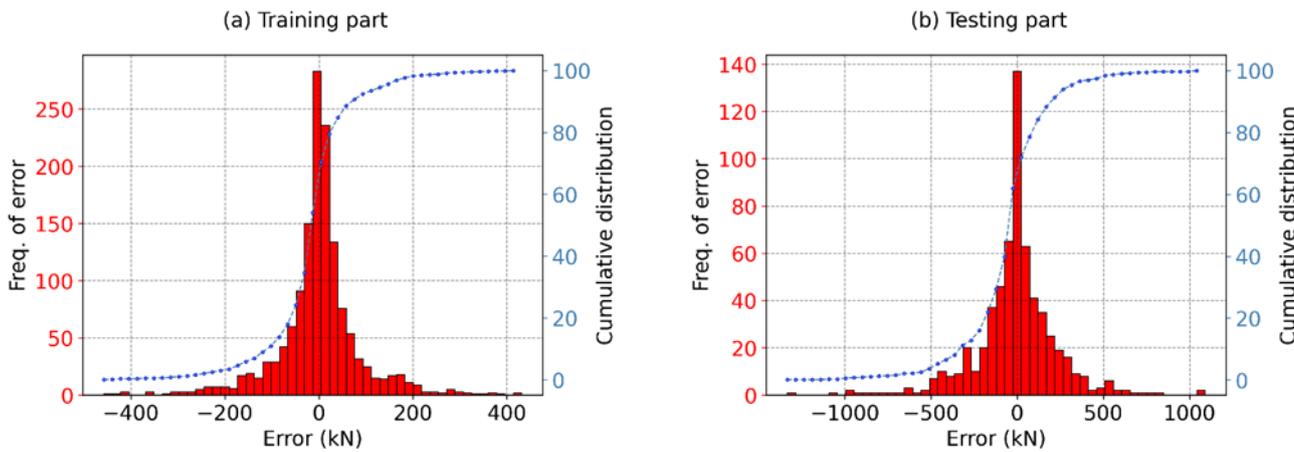
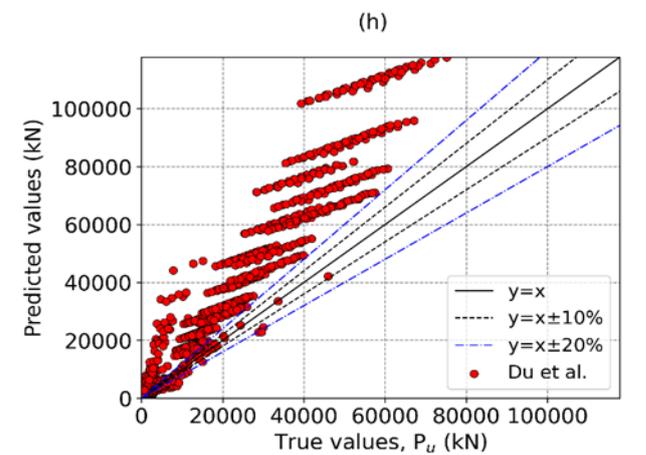
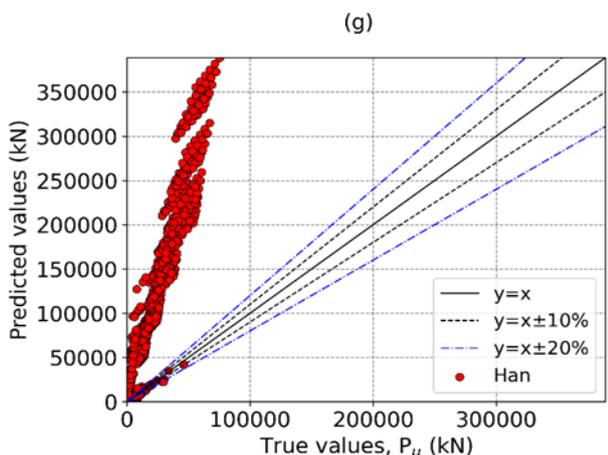
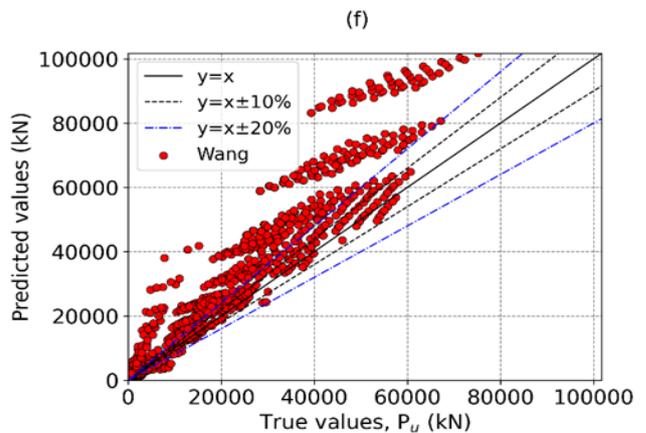
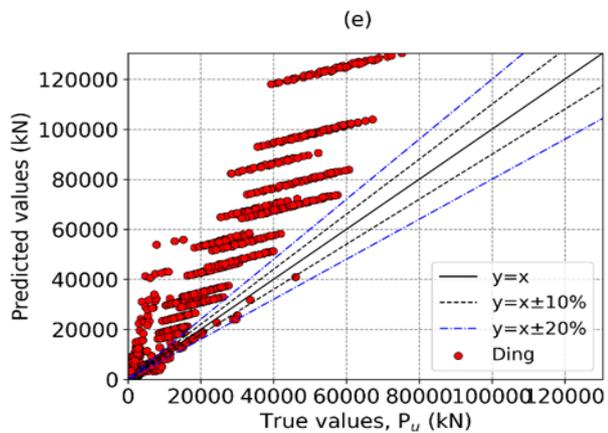
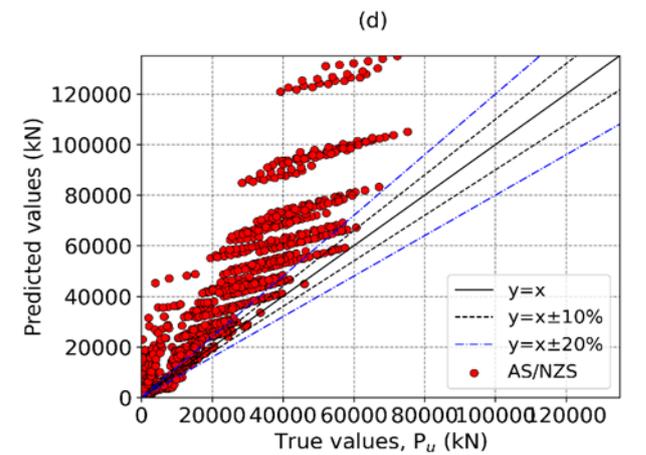
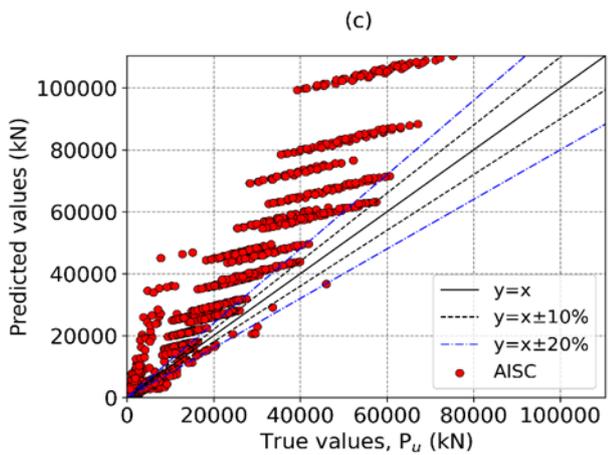
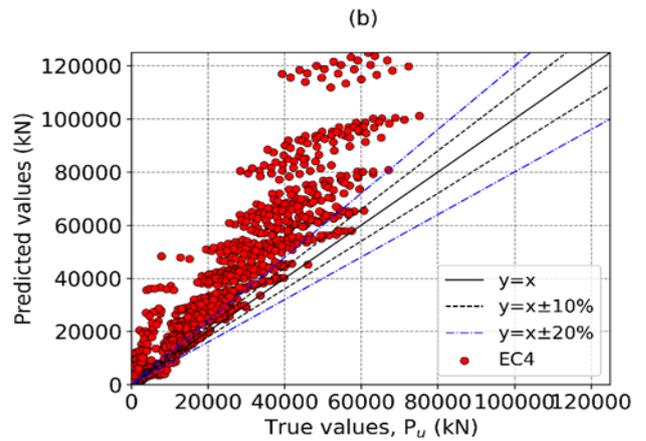
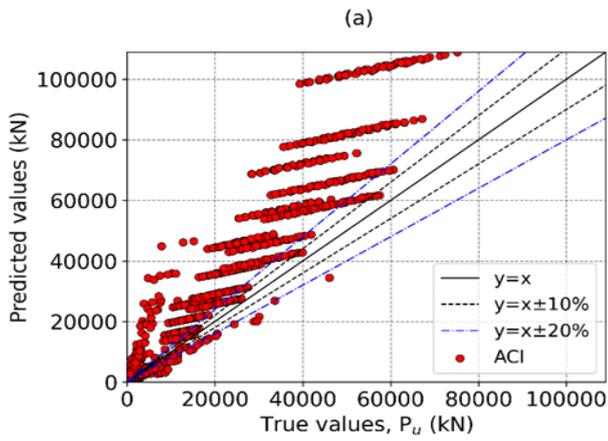


Fig 7. Error results (a) training part; and (b) testing part

4.3. Comparison with design codes and empirical equations

To further validate the proposed ML-based prediction model and illustrate its superiority, the performance of the XGB_05 model is compared to that of four existing design codes and four empirical equations developed by other researchers. Four design codes are investigated, including EC4 [27], ACI [28], AISC [29], AN/NZS [30], and the empirical equations of Ding [46], Wang [47], Han [48], and Du et al. [49]. This study employs all collected data for comparison purposes, using R^2 , RMSE, MAE, and MAPE to assess the performance of the models. Fig. 8 illustrates the regression analysis, whereas Table 6 indicates the prediction results between the actual and output data using design codes and empirical equations. As can be seen, all previously proposed design codes and experimental formulae have a large deviation

between the actual value and the predicted output. Compared to the experimental model with minimal error, i.e. Wang's model, the XGB_05 model improved in R^2 , RMSE, and MAE by 3.09%, 98.31%, and 98.12%, respectively. There are a variety of causes for the considerable variance in error: (1) the design codes and empirical equations frequently have constraints on the calculation assumptions, such as the condition of the column's slenderness, the condition of the steel's strength, and the compressive strength of the concrete; (2) ML models are capable of learning the effect of column length, whereas the design code and empirical equations are not, although column length has a significant impact on the axial load capacity of the column in general. Again, this demonstrates the proposed ML approach's advantages in improving prediction accuracy and generalization.



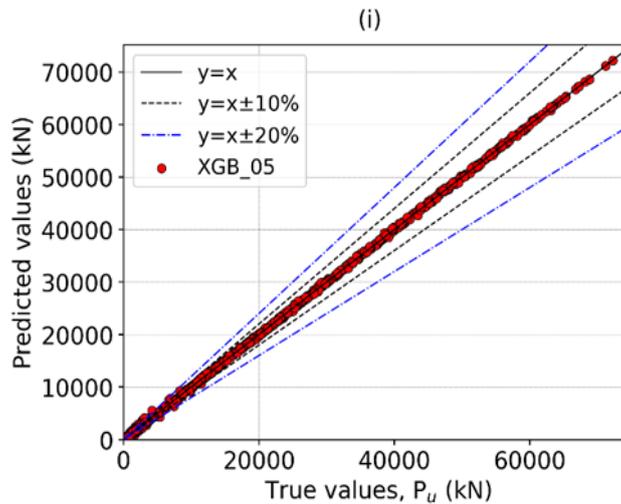


Fig 8. Regression model of design codes, empirical formulas, and XGB_05 model

Table 6. Statistical criteria values of design codes and empirical formulas in predicting the axial load of CFST columns

Models	Criteria			
	RMSE (kN)	MAE (kN)	R ²	MAPE
ACI [28]	12692.379	6892.101	0.966	0.701
EC4 [27]	13725.388	7141.012	0.956	1.081
AISC [29]	13089.701	7119.080	0.968	0.700
AN/NZS [30]	15729.399	8485.992	0.949	1.715
Ding [46]	18803.657	10345.129	0.966	0.896
Wang [47]	9053.863	4615.104	0.969	0.537
Han [48]	86636.629	48881.878	0.968	3.124
Du et al. [49]	14887.188	8255.142	0.974	0.755
This study	152.848	86.670	0.999	0.030

4.4. Feature importance and partial dependence analysis

The developed ML model has a significant capacity to predict the ALC of circular CFST columns. However, it is also vital to interpret or explain the model predictions [50] because ML models are "black boxes" in general. Model interpretation may guide the model's development and decision-making techniques while fostering user confidence in the trained model. This is accomplished by feature importance analysis and partial dependence analysis [51].

Analysis of feature significance is the most popular method for explaining model results. It provides a direct ranking of each feature's effect on

the target. The greater the influence of a feature on the model's prediction, the greater its importance. Fig. 9 gives the relative feature importance results for the XGB_05 model for the axial load of the circular CFST column. It can be observed that the outer diameter of steel tube D plays the most critical role. It is also the most dominant input parameter compared to the other four. This finding is expected because the outer tube diameter D is the main contributor to the cross-sectional area of circular CFST columns. Furthermore, the compressive strength of concrete (f_c) is the second most essential input parameter, followed by steel tube thickness (t) and yield strength of steel (f_y). Finally, the column length is the least relevant aspect of the circular CFST column's ALC.

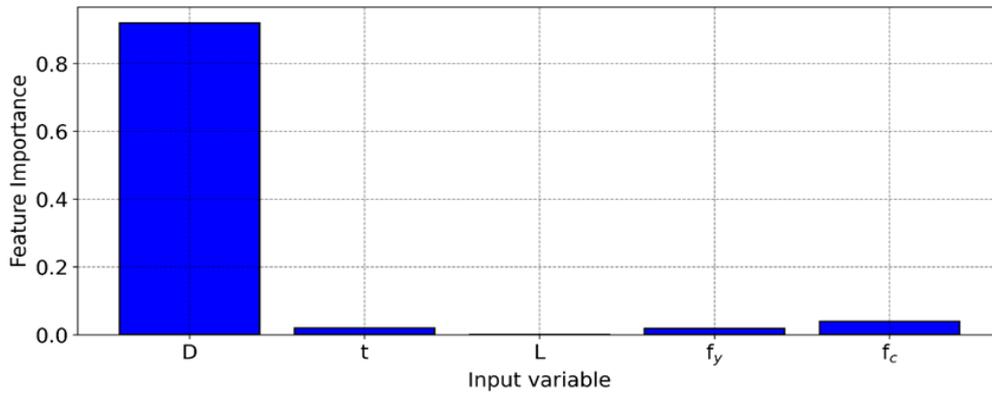


Fig 9. Feature importance analysis results

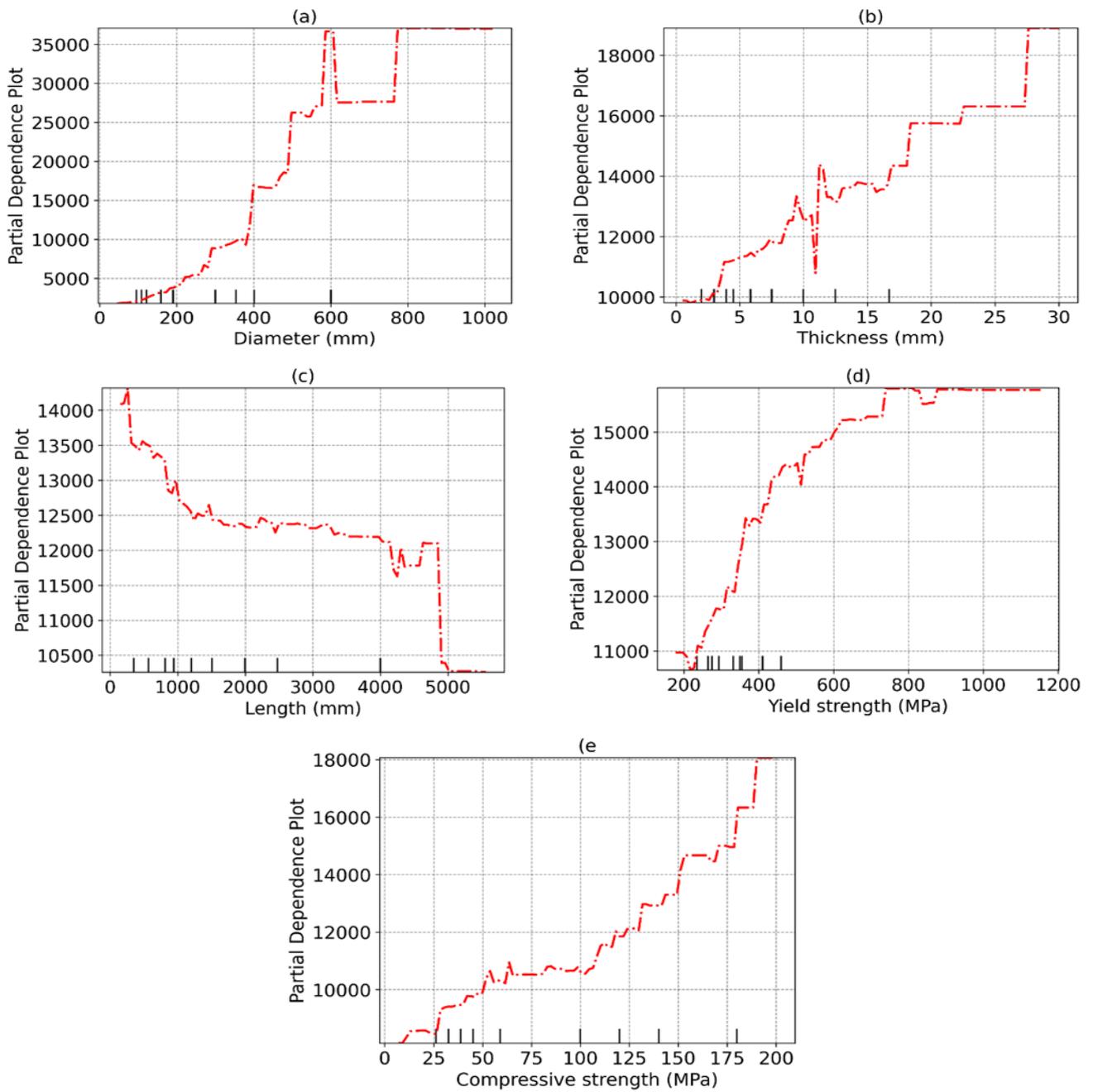


Fig 10. Partial dependence plots (PDP) analysis of the input variables used in this study

Feature importance analysis demonstrates whether or not the characteristics are important, but does not explain how the features impact the target. Partial dependence analysis is used to overcome the current limitation of feature importance analysis. It can show how the model's final output varies with the selected feature, and reveal if the feature has a positive or negative influence on the results. The model output's partial dependency on a specific feature X is determined by fixing the remaining features while feature X varies.

Fig. 10 depicts the PDP for the five input parameters over all samples in the dataset. The horizontal axis displays input parameter change, while the vertical axis represents the PDP value of each parameter. Although column length has a negative effect, the remaining factors tend to positively influence the ALC of the circular CFST column. When the column diameter (D) is increased from 44 mm to 600 mm, the ALC of the column increases significantly, and remains stable, with D being more than 800 mm. The ALC of the column varies in three different phases corresponding to the column diameter ranging from 0 to 600 mm, 600 to 800 mm, and above 800 mm. Besides, it is clear that the column diameter has the most significant impact on the axial load capacity of the column. This is based on the difference in the PDP value ranging from 1954.195 to 36681.478. This change is much larger than the PDP change of the remaining parameters. This observation is in good agreement with the findings of the feature importance analysis. Regarding the tube thickness (t), when t increases, so do the ALC of the CFST circular column, especially when t is in the [0.52, 11.5] mm range. However, when t varies in a short range, from 11.5 mm to 12 mm, the column's axial load capacity tends to decrease. The axial load capacity of columns tends to remain constant or slightly increase, especially with thicker columns.

When the column length (L) is considered, it is clear that the ALC decreases as L increases,

especially for long columns. This result is consistent with the literature [52], since as column length grows, the column becomes more narrow, limiting bearing capacity. Finally, the PDP curve of two factors, the compressive strength of concrete and the yield strength of steel, shows that as these two values increase, so does the ALC of the column. This corresponds to the structure's carrying capacity and remarks in the literature [52]. In summary, partial dependency plots clearly explain how factors impact prediction while avoiding high processing costs.

5. Conclusions

In this study, an XGB model is developed to estimate the ALC of circular CFST structural components subjected to compression. The dataset of 2073 experimental results of circular CFST columns subjected to axial loads is utilized to develop the model. Two types of input parameters are employed to train the model, including the geometric dimensions of the section and the material's mechanical properties. The results reveal that the XGB model delivers reliable prediction results with $R^2 = 0.999$, $RMSE = 242.757$ kN, $MAE = 157.045$ kN, and $MAPE = 0.057$. The proposed model was also compared with existing empirical equations and published standards, which shows higher accuracy. In addition, the sensitivity analysis employing partial dependence plots analysis and feature importance is performed to determine the influence of each input variable on the output. This study's findings can facilitate and simplify the design of circular CFST columns based on input parameters. The optimum values provide a rapid and accurate assessment of the ALC of a circular CFST column for practical applications in the construction of civil engineering structures.

Conflict of Interest: The authors declare no conflict of interest.

Acknowledgments: This research is funded by the University of Transport Technology, Thanh Xuan, Hanoi, Vietnam (UTT), under grant number DTTD2022-05.

Funding: This research did not receive any specific grant from public, commercial, or not-for-profit funding agencies.

Availability of data and material: Data will be made available on request.

References

- [1] L.-H. Han, D.-Y. Ma, K. Zhou. (2018). Concrete-encased CFST structures: behaviour and application. *Proceedings of the 12th International Conference on Advances in Steel-Concrete Composite Structures. ASCCS 2018, Editorial Universitat Politècnica de València*, pp 1-10.
- [2] L.-H. Han, W. Li, R. Bjorhovde. (2014). Developments and advanced applications of concrete-filled steel tubular (CFST) structures: Members. *Journal of Constructional Steel Research*, 100, 211-228.
- [3] J.R. Liew, M. Xiong, D. Xiong. (2016). Design of concrete filled tubular beam-columns with high strength steel and concrete. *Structures, Elsevier*, 8(Part2), 213-226.
- [4] X. Zhou, J. Liu. (2019). Application of steel-tubed concrete structures in high-rise buildings. *International Journal of High-Rise Buildings*, 8(3), 161-167.
- [5] F. Alatshan, S.A. Osman, F. Mashiri, R. Hamid. (2019). Explicit simulation of circular CFST stub columns with external steel confinement under axial compression. *Materials*, 13(1), 23.
- [6] A. Akbar, F. Farooq, M. Shafique, F. Aslam, R. Alyousef, H. Alabduljabbar. (2021). Sugarcane bagasse ash-based engineered geopolymer mortar incorporating propylene fibers. *Journal of Building Engineering*, 33, 101492.
- [7] B. Rong, S. Liu, Z. Li, R. Liu. (2018). Experimental and numerical studies of failure modes and load-carrying capacity of through-diaphragm connections. *Transactions of Tianjin University*, 24, 387-400.
- [8] T. Zhang, F. Ding, X. Liu, Z. Yu. (2021). Seismic behavior of terminal stirrup-confined concrete-filled elliptical steel tube columns: experimental investigation. *Thin-Walled Structures*, 167, 108251.
- [9] J. Zhang, Y. Liu, Z. Chen, R. Cai, X. Li. (2021). Seismic behavior of high-strength concrete-filled square steel tube columns reinforced with ultrahigh-strength reinforcing bars. *Structures, Elsevier*, 34, 3125-3140.
- [10] W. Feng, Y. Wang, J. Sun, Y. Tang, D. Wu, Z. Jiang, J. Wang, X. Wang. (2022). Prediction of thermo-mechanical properties of rubber-modified recycled aggregate concrete. *Construction and Building Materials*, 318, 125970.
- [11] H.Q. Nguyen, H.-B. Ly, V.Q. Tran, T.-A. Nguyen, T.-T. Le, B.T. Pham. (2020). Optimization of artificial intelligence system by evolutionary algorithm for prediction of axial capacity of rectangular concrete filled steel tubes under compression. *Materials*, 13(5), 1205.
- [12] Q.H. Nguyen, H.-B. Ly, V.Q. Tran, T.-A. Nguyen, V.-H. Phan, T.-T. Le, B.T. Pham. (2020). A novel hybrid model based on a feedforward neural network and one step secant algorithm for prediction of load-bearing capacity of rectangular concrete-filled steel tube columns. *Molecules*, 25(15), 3486.
- [13] P. Zhao, Y. Huang, Z. Liu, H. Wang, Y. Lu. (2022). Experimental research on seismic performance of steel fiber-reinforced recycled concrete-filled circular steel tube columns. *Journal of Building Engineering*, 54, 104683.
- [14] X. Wang, F. Fan, J. Lai. (2022). Strength behavior of circular concrete-filled steel tube stub columns under axial compression: A review. *Construction and Building Materials*, 322, 126144.
- [15] H. Jahangir, A. Soleymani, M.R. Esfahani. (2022). Investigating the confining effect of steel reinforced polymer and grout composites on compressive behavior of square concrete columns. *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, 47, 775-791.
- [16] J.B. Chen, T.M. Chan, J.M. Castro. (2017).

- Parametric study on the flexural behaviour of circular rubberized concrete-filled steel tubes. *Tubular Structures XVI, CRC Press*, pp 51-59.
- [17] H.-S. Hu, Y. Liu, B.-T. Zhuo, Z.-X. Guo, B.M. Shahrooz. (2018). Axial compressive behavior of square CFST columns through direct measurement of load components. *Journal of Structural Engineering*, 144(11), 04018201.
- [18] G. Li, B. Chen, Z. Yang, Y. Feng. (2018). Experimental and numerical behaviour of eccentrically loaded high strength concrete filled high strength square steel tube stub columns. *Thin-Walled Structures*, 127, 483-499.
- [19] C.-C. Chen, J.-W. Ko, G.-L. Huang, Y.-M. Chang. (2012). Local buckling and concrete confinement of concrete-filled box columns under axial load. *Journal of Constructional Steel Research*, 78, 8-21.
- [20] M.F. Javed, N.R. Sulong, S.A. Memon, S.K.U. Rehman, N.B. Khan. (2017). FE modelling of the flexural behaviour of square and rectangular steel tubes filled with normal and high strength concrete. *Thin-Walled Structures*, 119, 470-481.
- [21] R.B. Knowles, R. Park. (1969). Strength of concrete filled steel tubular columns. *Journal of the Structural Division*, 95(12), 2565-2588.
- [22] Z. Liu, S.C. Goel. (1988). Cyclic load behavior of concrete-filled tubular braces. *Journal of Structural Engineering*, 114(7), 1488-1506.
- [23] A.E. Kilpatrick, B.V. Rangan. (1999). Tests on high-strength concrete-filled steel tubular columns. *Structural Journal*, 96(2), 268-274.
- [24] K. Sakino, H. Nakahara, S. Morino, I. Nishiyama. (2004). Behavior of centrally loaded concrete-filled steel-tube short columns. *Journal of Structural Engineering*, 130(2), 180-188.
- [25] X. Dai, D. Lam. (2010). Numerical modelling of the axial compressive behaviour of short concrete-filled elliptical steel columns. *Journal of Constructional Steel Research*, 66(7), 931-942.
- [26] H.T. Duong, H.C. Phan, T.-T. Le, N.D. Bui. (2020). Optimization design of rectangular concrete-filled steel tube short columns with Balancing Composite Motion Optimization and data-driven model. *Structures, Elsevier*, 28, 757-765.
- [27] Eurocode 4, CEN, EN1994-1. (2004). Design of composite steel and concrete structures-Part 1-1: general rules and rules for buildings, Brussels, Belgium. (n.d.).
- [28] A.C.I. Committee. (2019). ACI 318-19: Building Code Requirements for Structural Concrete and Commentary, American Concrete Institute: Farmington Hills, MI, USA.
- [29] AISC. (2016). Specification for structural steel buildings ANSI/AISC 360-16, American Institute of Steel Construction, Chicago. (n.d.).
- [30] W. AS5100. (2004). Bridge design, Part 6: steel and composite construction, (Australian Standard).
- [31] M. Ahmadi, H. Naderpour, A. Kheyroddin. (2017). ANN model for predicting the compressive strength of circular steel-confined concrete. *International Journal of Civil Engineering*, 15, 213-221.
- [32] J. Moon, J.J. Kim, T.-H. Lee, H.-E. Lee. (2014). Prediction of axial load capacity of stub circular concrete-filled steel tube using fuzzy logic. *Journal of Constructional Steel Research*, 101 184-191.
- [33] P. Sarir, J. Chen, P.G. Asteris, D.J. Armaghani, M.M. Tahir. (2021). Developing GEP tree-based, neuro-swarm, and whale optimization models for evaluation of bearing capacity of concrete-filled steel tube columns. *Engineering with Computers*, 37, 1-19.
- [34] X. Liu, Y. Wu, Y. Zhou. (2022). Axial Compression Prediction and GUI Design for CCFST Column Using Machine Learning and Shapley Additive Explanation. *Buildings*, 12(5), 698.
- [35] E.M. Güneş, A. Gültekin, K. Mermerdaş. (2016). Ultimate capacity prediction of axially loaded CFST short columns. *International*

- Journal of Steel Structures*, 16, 99-114.
- [36] M. Denavit. (2019). Steel-Concrete Composite Column Database. *Journal of Constructional Steel Research*.
- [37] C. Goode. (2018). Composite column tests-database and comparison with Eurocode 4. *Proceedings of the 12th International Conference on Advances in Steel-Concrete Composite Structures. ASCCS 2018, Editorial Universitat Politècnica de València*, pp. 763-767.
- [38] V.-L. Tran, D.-K. Thai, D.-D. Nguyen. (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.
- [39] Q. Ren, M. Li, M. Zhang, Y. Shen, W. Si. (2019). Prediction of ultimate axial capacity of square concrete-filled steel tubular short columns using a hybrid intelligent algorithm. *Applied Sciences*, 9(14), 2802.
- [40] B.G. Tabachnick, L.S. Fidell, J.B. Ullman. (2007). Using multivariate statistics, *Pearson Boston, MA*.
- [41] Q.H. Nguyen, H.-B. Ly, L.S. Ho, N. Al-Ansari, H.V. Le, V.Q. Tran, I. Prakash, B.T. Pham. (2021). Influence of data splitting on performance of machine learning models in prediction of shear strength of soil. *Mathematical Problems in Engineering*, 2021, 4832864, 1-15.
- [42] T. Chen, C. Guestrin. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, pp. 785-794.
- [43] J. Elith, J.R. Leathwick, T. Hastie. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4), 802-813.
- [44] G.G. Moisen, E.A. Freeman, J.A. Blackard, T.S. Frescino, N.E. Zimmermann, T.C. Edwards Jr. (2006). Predicting tree species presence and basal area in Utah: a comparison of stochastic gradient boosting, generalized additive models, and tree-based methods. *Ecological Modelling*, 199(2), 176-187.
- [45] C.P. Robert, G. Casella, G. Casella. (1999). Monte Carlo statistical methods. *Springer*.
- [46] F. Ding, L. Luo, J. Zhu, L. Wang, Z. Yu. (2018). Mechanical behavior of stirrup-confined rectangular CFT stub columns under axial compression. *Thin-Walled Structures*, 124, 136-150.
- [47] Z.-B. Wang, Z. Tao, L.-H. Han, B. Uy, D. Lam, W.-H. Kang. (2017). Strength, stiffness and ductility of concrete-filled steel columns under axial compression. *Engineering Structures*, 135, 209-221.
- [48] L.-H. Han, G.-H. Yao. (2018). Tests on Hollow Structural Steel (HSS) Columns Filled with Self-Consolidating Concrete (SCC). *Thin-Walled Structures, CRC Press*, pp. 901-908.
- [49] Y. Du, Z. Chen, M.-X. Xiong. (2016). Experimental behavior and design method of rectangular concrete-filled tubular columns using Q460 high-strength steel. *Construction and Building Materials*, 125, 856-872.
- [50] S. Mangalathu, S.-H. Hwang, J.-S. Jeon. (2020). Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach. *Engineering Structures*, 219, 110927.
- [51] X. Zhao, R. Lovreglio, D. Nilsson. (2020). Modelling and interpreting pre-evacuation decision-making using machine learning. *Automation in Construction*, 113, 103140.
- [52] T.-T. Le, H.C. Phan. (2020). Prediction of ultimate load of rectangular CFST columns using interpretable machine learning method. *Advances in Civil Engineering*, 2020, 8855069, 1-16.