Strength prediction of concrete-filled steel tubular columns using Categorical Gradient Boosting algorithm

Seunghye Lee[a], Thuc P. Vo[b,*], Huu-Tai Thai[c], Jaehong Lee[a], Vipulkumar Patel[d]

[a]*Deep Learning Architecture Research Center, Sejong University*
209, Neungdong-ro, Gwangjin-gu, Seoul 05006, Republic of Korea
[b]*School of Engineering and Mathematical Sciences, La Trobe University, Bundoora, VIC 3086, Australia*
[c]*Department of Infrastructure Engineering, The University of Melbourne, Parkville VIC 3010, Australia*
[d]*School of Engineering and Mathematical Sciences, La Trobe University, Bendigo, VIC 3552, Australia*

Abstract

Due to complexities from the interaction between steel tube and concrete filling of concrete-filled steel tubular (CFST) columns, their strengths are very complicated, which is a highly nonlinear relation with material strengths and geometry. Categorical gradient Boosting (CatBoost), which is advanced boosting machine, is presented to solve the problems. A total of 3103 tests, which is divided in four datasets, is trained and tested the learners to determine the ultimate axial strength as the output variable while the strength of materials (concrete and steel) and geometry (e.g., diameters/width/heights, thickness, effective length, eccentricities) are the input ones. The comparison of the present results from 10-fold cross validation and those from the code predictions (AISC 360-16, Eurocode 4 and AS/NZS 2327) and previous study shows very high prediction accuracy in terms of coefficient of determination (R^2), which is the lowest value (R^2=0.964) for Dataset 2 and the highest one (R^2=0.996) for Dataset 1. While the predictions from three codes beyond material limit and slenderness are less conservative than those within it, CatBoost provides nearly similar experiment results with the mean values as unity without any limits. This algorithm can be used to predict an accurate strength of CFST columns.

Keywords: Concrete-filled steel tubular columns, Categorical gradient Boosting (CatBoost), Code predictions, Material strengths, Slenderness ratio

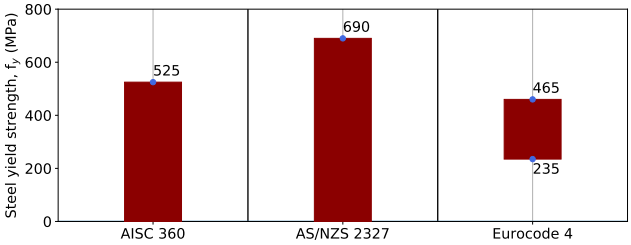
1. Introduction

Concrete-Filled Steel Tubular (CFST) columns, which have excellent structural performance, provide not only high load carrying and energy absorption capacity but also reduce

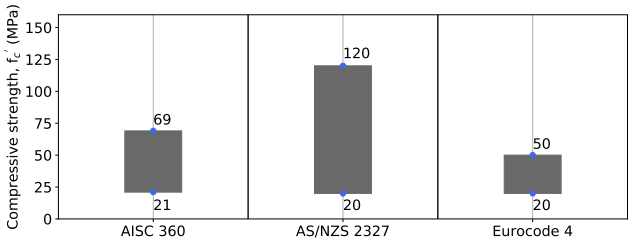
*Corresponding author
Email address: t.vo@latrobe.edu.au (Thuc P. Vo)

construction costs. Due to these advantages, they have been widely used in many structural applications such as high rise building, bridges, offshore structures, etc. Their strength prediction has been attracted to many researchers using experimental, simulation modelling and theoretical studies. Only some representative works are mentioned here and more papers related to this topic can be found in the recent comprehensive work by Thai et al. [1]. Since the first experimental study by Kloppel and Goder [2], a large number of tests have been carried out to study their structural responses and then collected as databases by many authors ([3–8]). In order to reduce the use of construction materials, there also has been a significant interest in using high steel yield strength (f_y) and concrete compressive strength (f'_c) in the CFST columns. Liew et al. [9] and Xiong et al. [10] studied these structures with f_y and f'_c up to 779 MPa and 193 MPa, as well as those by Khan et al. [11] up to 762 MPa and 113 MPa, respectively. Their tests were also conducted by Mursi and Uy ([12], [13]) and Sakino et al. [14] with f_y up to 761 MPa and 853 MPa. The guidelines for the design of the CFST structures have been documented in various design codes such as American (AISC 360-16 [15]), British (BS 5400 [16]), European (Eurocode 4 [17]), Japan (AIJ [18]), China (GB 50936 [19] and Australian/New Zealand (AS/NZS 2327 [20]). Only three typical codes of AISC 360-16, Eurocode 4 and AS/NZS 2327 are considered in this study. Among them, AS/NZS 2327 is the only code that allows the use of high strength steel and concrete (with f_y and f'_c up to 690 MPa and 120 MPa). It is clear that these above codes cannot cover the whole range of materials strengths as shown in Fig. 1. Besides, there are some deviations among the obtained results from these codes for the same problem and thus it is difficult to predict precisely the strength of the CFST columns.

The above code limitations can be overcome by using machine learning algorithms, which learn from the dataset and provide relationships between the input and output variables ([21–23]). There have been some efforts to utilise them to predict the performances of the CFST columns using genetic algorithms (GA) by Naser et al. [24] and Artificial Neural Network (ANN) based-method ([25–29]). Ren et al. [30] combined a metaheuristic algorithm with a Support Vector Machine (SVM) to predict the ultimate axial capacity of square CFST short columns. Mai et al. [31] employed the neural network as a predictive model to solve this problem. However, most previous researches used simple ANNs to train the regression problems by using experimental dataset, while there are some ensemble learning models [32], which are more accurate, robust and powerful. By using the training data, these models train several



(a) Steel yield strength (f_y)



(b) Concrete compressive strength (f_c')

Figure 1: Ranges of material strengths in the CFST columns.

weak learners and then integrate them to construct a strong one. Therefore, their accuracy and robustness is improved.

In order to pursue such goal, this paper explores Categorical gradient Boosting (CatBoost) [33], which is the most typical ensemble learning models, to predict the capacity of CFST columns. The set of hyper-parameters, which maximises the effective performance, is determined by using a grid search method. CatBoost is trained and validated against a comprehensive database of 3103 tests covering a wide range of material and geometric properties as well as loading configurations. The comparison of the present results from 10-fold cross validation and those from the code predictions (AISC 360-16, Eurocode 4 and AS/NZS 2327) and previous study shows very high prediction accuracy. Finally, the effects of the training dataset amount and input variable sensitivities as well as input variable number are discussed.

2. Categorical gradient Boosting (CatBoost)

The dataset of CFST columns can be a regression problem (Fig. 2), which is the process of learning relationships between the input and output variables [34]. There is a variety of regression algorithms that can be considered as candidates to predict the capacity of CFST columns. In particular, more effective ensemble machine learning algorithms have been proposed incorporating many of the recent advances in machine learning [35]. The ensemble methods combine

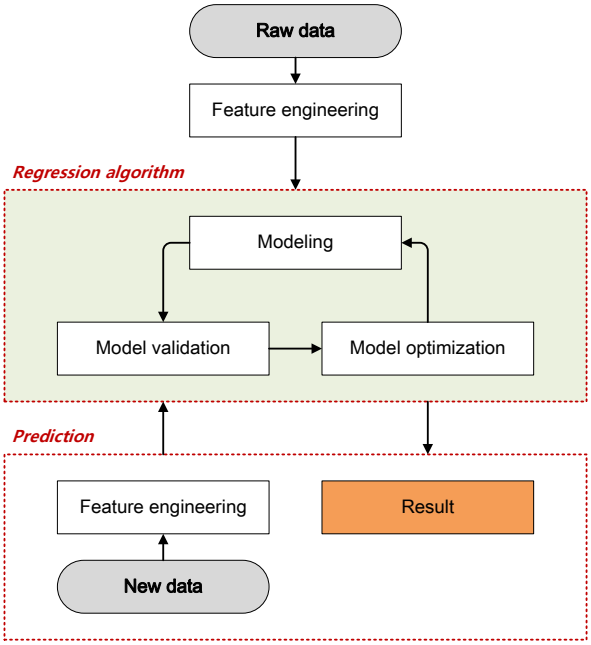


Figure 2: Generic flowchart for regression algorithms.

multiple differential machine learning models to improve the accuracy [36]. The most common approaches to create ensemble models are bagging [37] , boosting [38], and stacking [39]. This study focuses on the boosting in general and Gradient Boosting Machines (GBM) [40] in particular, which combines a set of models, whose individual predictions are aggregated [41]. CatBoost proposed by Yandex [42] is a new type of GBM that can handle a categorical features well. Let $\{(X_k, Y_k)\}_{k=1}^{n}$ be the given input dataset, then $\mathbf{X}_i = (x_{i,1}, \cdots, x_{i,m})$ and $Y_i \in \mathbb{R}$ are a vector of m feature and a label value, respectively. The simplest way is to substitute $x_{i,k}$ with the average label value $\frac{\sum_{j=1}^{n}[x_{j,k}=x_{i,k}]\cdot Y_j}{\sum_{j=1}^{n}[x_{j,k}=x_{i,k}]}$ on the whole train dataset, where $[\cdot]$ denotes Iverson brackets (if $x_{j,k} = x_{i,k}$, then $[x_{j,k} = x_{i,k}]$ equals 1 and 0 otherwise). However, this procedure leads to overfitting problems. Unlike other GBM learning algorithms, CatBoost has the following aspects:

- **Level-wise tree :** by using level-wise symmetric trees (Fig. 3), it employs the same features to split learning instances for each level. The structure implements a vectorized representation of the tree, which can be evaluated rapidly.

- **Ordered boosting and categorical feature combinations :** by using ordered boosting as a new method to estimate the gradient, it can overcome gradient bias using unbiased boosting techniques with categorical features and enhance the generalization ability of the model [43].

- **Random permutation :** by implementation of the gradient boosting, CatBoost uses a random permutation of the training examples to prevent over-fitting problems. The set of input values are first randomly organised to create the random permutations. The average sample values are then calculated for the same category. If a permutation is $\sigma = (\sigma_1, \cdots, \sigma_n)$, it is substituted with:

$$\frac{\sum_{j=1}^{p-1}[x_{\sigma_{j,k}} = x_{\sigma_{p,k}}]Y_{\sigma_j} + \beta \cdot P}{\sum_{j=1}^{p-1}[x_{\sigma_{j,k}} = x_{\sigma_{p,k}}] + \beta} \tag{1}$$

  where P and β are a prior value and the weight of the prior, respectively.

- **One-hot encoding :** Because GBM cannot handle categorical features, in this case, one-hot encoding process is demanded. However, CatBoost directly contains the one-hot encoding process to convert categorical features to target statistics.

- **Optimized parameter tuning :** it provides a flexible interface for hyper-parameter tuning, which includes tuning the learning rate and the depth of the trees. These parameters guarantee to get the highest accuracy of the analysis [44].

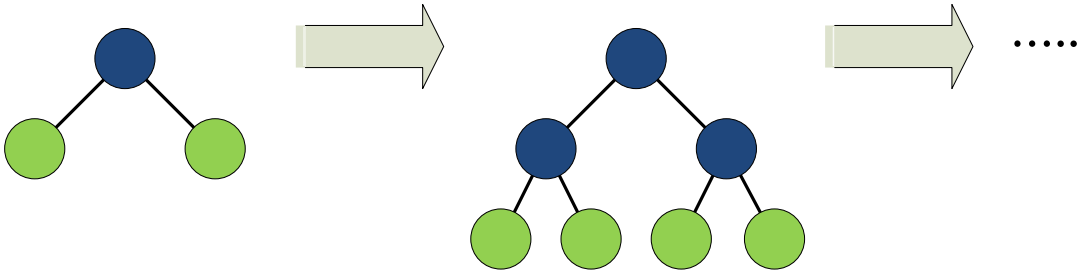


Figure 3: Level-wise tree growth in CatBoost.

3. Test database and design guidelines

3.1. Test database

Fig. 4 depicts cross sections of two typical CFST columns, where the concrete is filled in a circular or rectangular hollow section. The notations D, B, H, and t denote the outside diameter of a circular section, width and height of rectangular section and thickness of the steel sections, respectively. The results of 3103 tests collected from 173 studies were used as test database [1], which is summarised in Table 1. The compressive strength is the output variable for whole datasets.

Table 1: Summary of experimental specimens.

Dataset No.	Section type	Loading type	No. of specimens	Violin plot	Heat-map
1	Circular	Concentric*	1245	Fig. 5(a)	Fig. 5(b)
2		Eccentric$	485	Fig. 6(a)	Fig. 6(b)
3	Rectangular	Concentric*	979	Fig. 7(a)	Fig. 7(b)
4		Eccentric$	394	Fig. 8(a)	Fig. 8(b)

* Short (L/D (or B) $\leq$ 4) and slender column (L/D (or B) $>$ 4).

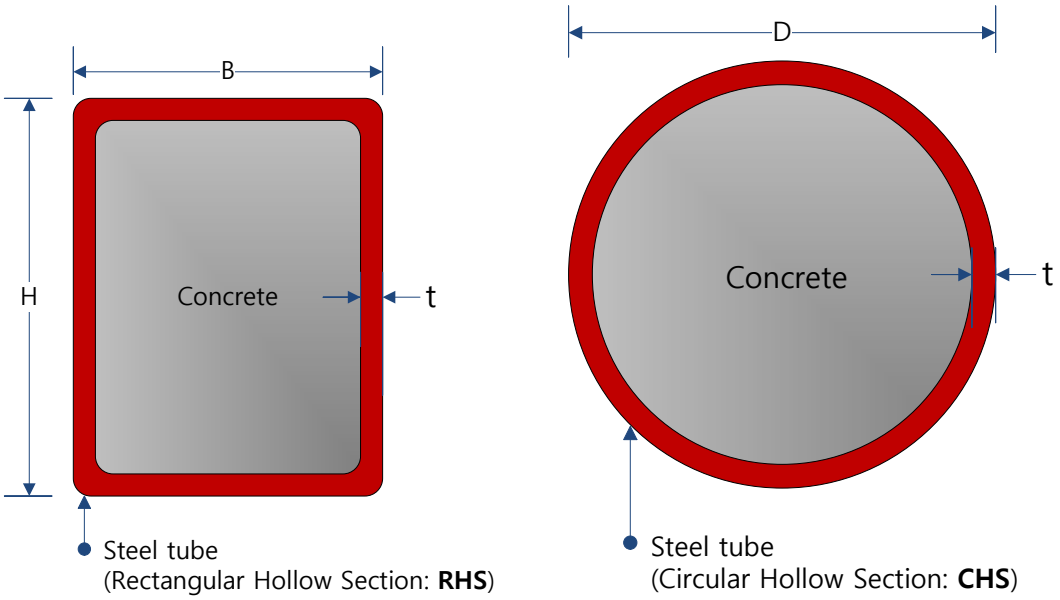
$ Beam-column.

Dataset 1 and 2 consist of 1245 and 485 testing results for concentric and eccentric circular columns. There are five input variables including steel and concrete strengths (f_y, f'_c) and geometry (D, t, L) for Dataset 1 and two extra variables related to eccentric parameters (e_t, e_b) for Dataset 2. The distribution and correlation matrix heat-map of features are shown in Figs. 5 and 6. The correlation between the compressive strength (N) and diameter (D) as well as thickness (t) is very strong with the coefficients of 0.91, 0.54 for Dataset 1 and 0.79 and 0.54 for Dataset 2, respectively.

Dataset 3 and 4 have 979 and 394 samples for concentric and eccentric rectangular columns. The compressive strength can be predicted by six input variables, namely, material strengths (f_y, f'_c) and geometry (B, H, t, L) for Dataset 3 and three extra variables including eccentric parameters (e_t, e_b, X) for Dataset 4. The violin plots show the distribution and correlation matrix heat-map of features are depicted in Figs. 7 and 8. A strong correlation is observed between N and three attributes (B, H, t) of samples.

3.2. Design guidelines

Detailed guidelines for calculating the strength of CFST columns from AISC 360-16 [15], Eurocode 4 [17] and AS/NZS 2327 [20] are given in Thai et al. [1]. For concentrically loaded columns, the section capacity is contributed from the structural steel tube and concrete infill by considering the confining effect which is significant for circular sections. In the AISC 360-16 code, this effect in circular section is considered by increasing the concrete compressive strength from $0.85f'_c$ to $0.95f'_c$. Meanwhile, both Eurocode 4 and AS/NZS 2327 account for the confining effect by using the same manner. In this method, the yield stress of the steel tube is reduced

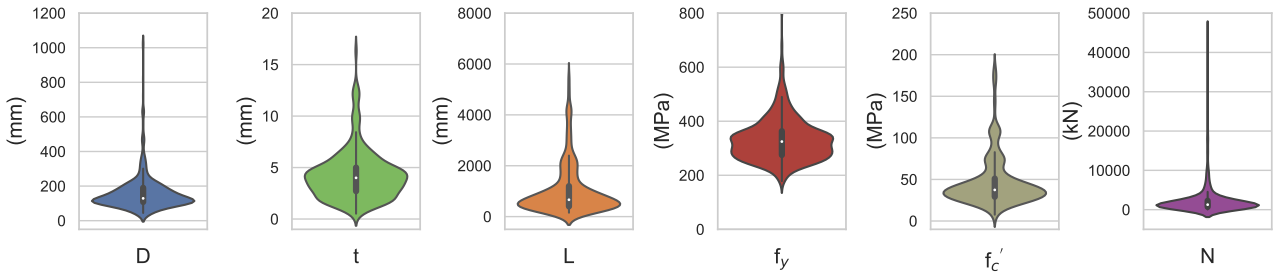


(a) RHS (b) CHS

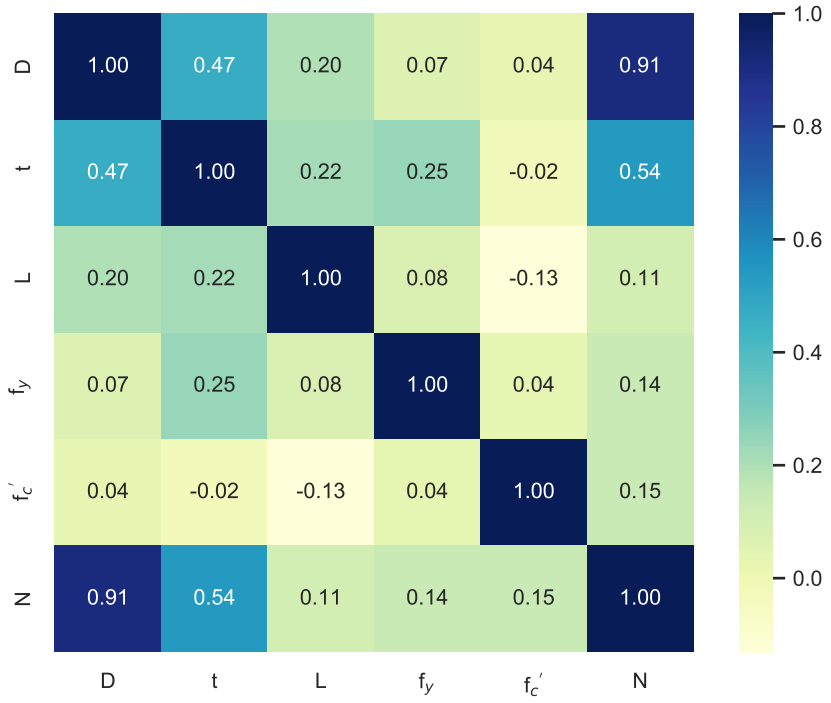
Figure 4: Rectangular hollow section (RHS) and circular hollow section (CHS).

by $\eta_s f_y$ due to the lateral pressure from the expansion of the concrete core, whilst the concrete compressive strength is increased by $(1+\eta_c \frac{t}{D}\frac{f_y}{f'_c})f'_c$ (with t and D being the thickness and outside diameter of the steel tube) due to the confining effect. The coefficients η_s and η_c are expressed as the functions of the relative member slenderness $[\bar{\lambda} = \sqrt{N_{us}/N_{cr}}]$ (with N_{us} and N_{cr} being the section strength and Euler buckling load of the column, respectively) as $\eta_s = 0.25\left(3 + 2\bar{\lambda}\right)$ and $\eta_c = 4.9 - 18.5\bar{\lambda} + 17\bar{\lambda}^2$. It can be seen from these equations that the confining effect is maximised in short columns at the smallest value of $\bar{\lambda}$ and is negligibly small in long columns when $\bar{\lambda}$ is greater than 0.5. This is a rational approach which represents exactly the mechanical behaviour of the confining effect in CFST circular columns under axial compression. Therefore, both Eurocode 4 and AS/NZS 2327 codes give the prediction of the section capacity of CFST circular columns more accurate than AISC 360-16 [1]. For eccentrically loaded columns, the AISC 360-16 code simplifies the axial force-moment interaction of CFST columns by using bi-linear curves for compact and slender sections as shown in Fig. 9a. Whereas, both Eurocode 4 and ASNZS 2327 adopt the four-point interaction curve as shown in Fig. 9b, and use a similar procedure to account for the effect of initial imperfections (the shaded part in Fig. 9b).

Regarding to the local buckling effect of the steel tube, Eurocode 4 does not allow the use of slender section due to ignoring the local buckling effect. Whereas, both AISC 360-16 and AS/NZS 2327 do allow for the use of non-compact and slender sections due to considering

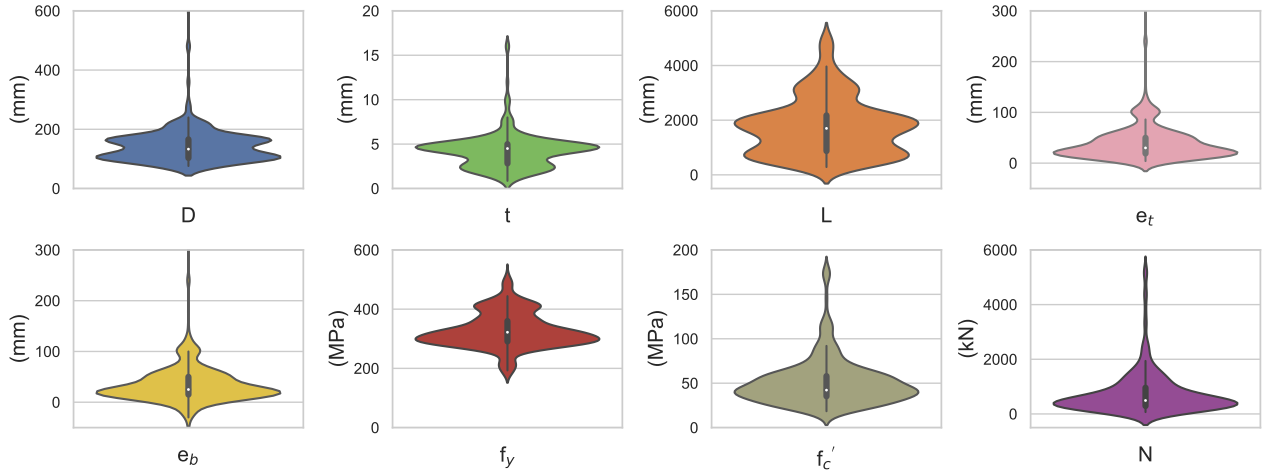


(a) Violin plots with distribution of features

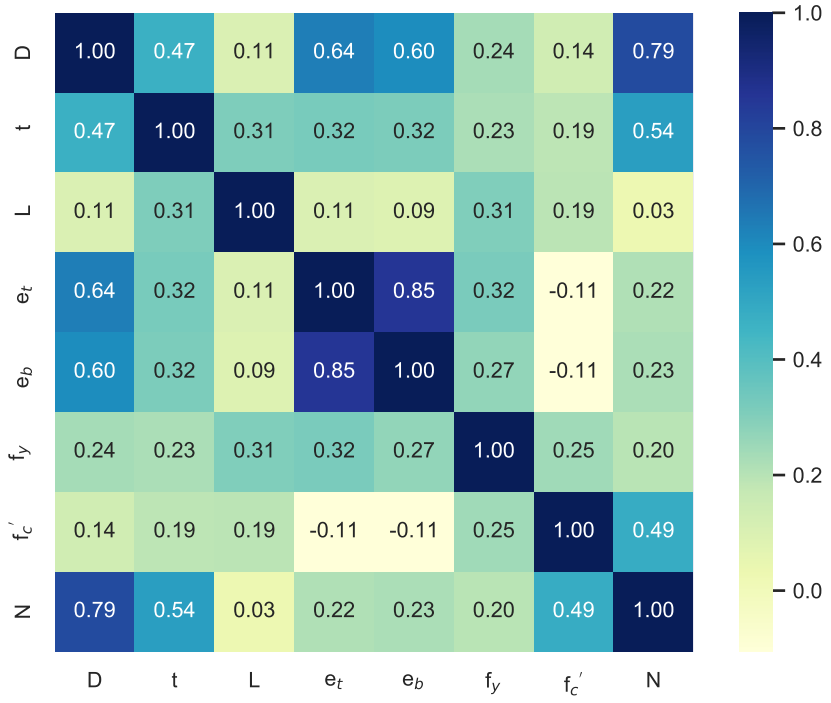


(b) Correlation matrix heat-map of features

Figure 5: Violin plots and correlation matrix heat-map of features for Dataset 1.

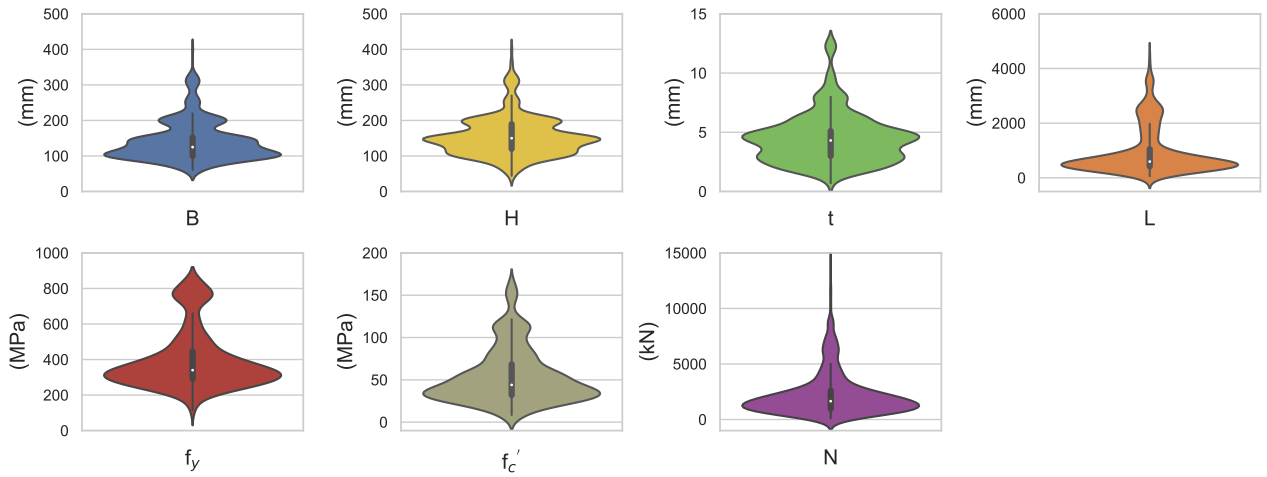


(a) Violin plots with distribution of features



(b) Correlation matrix heat-map of features

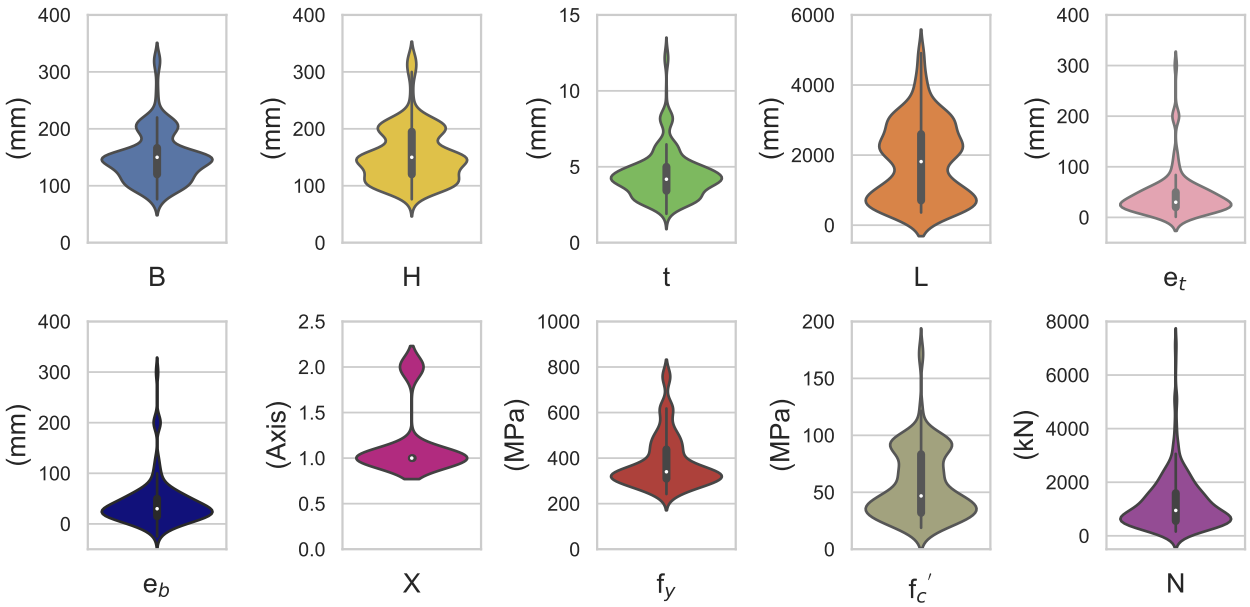
Figure 6: Violin plots and correlation matrix heat-map of features for Dataset 2.



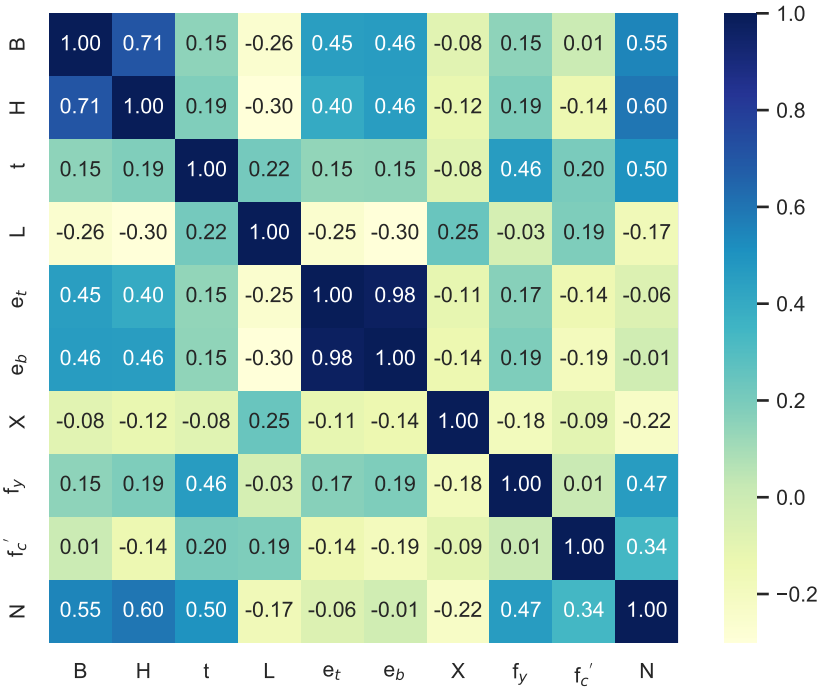
(a) Violin plots with distribution of features

(b) Correlation matrix heat-map of features

Figure 7: Violin plots and correlation matrix heat-map of features for Dataset 3.

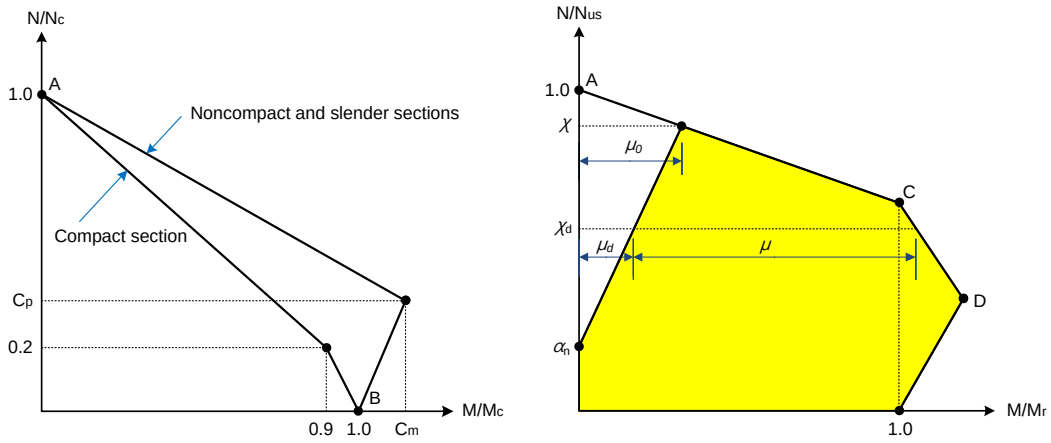


(a) Violin plots with distribution of features



(b) Correlation matrix heat-map of features

Figure 8: Violin plots and correlation matrix heat-map of features for Dataset 4.



(a) AISC 360-16 (b) Eurocode 4 and AS/NZS 2327

Figure 9: Non-dimensional N-M interaction diagram.

the local buckling effect. In the AISC 360-16 code, the local buckling effect is considered by reducing the concrete compressive strength from $0.95f'_c$ to $0.7f'_c$ and the strength of the steel tube from the yield stress f_y to the buckling stress f_{cr} (a function of the section slenderness ratio D/t). Meanwhile, the AS/NZS 2327 code accounts for the local buckling effect of the steel tube by means of the effective width method with a form factor k_f. The global buckling effect of columns is considered through a reduction factor χ as shown in Fig. 10. It can be seen that the column curves of Eurocode 4 and AS/NZS 2327 (without local buckling effect, i.e. $k_f = 1$) are almost identical.

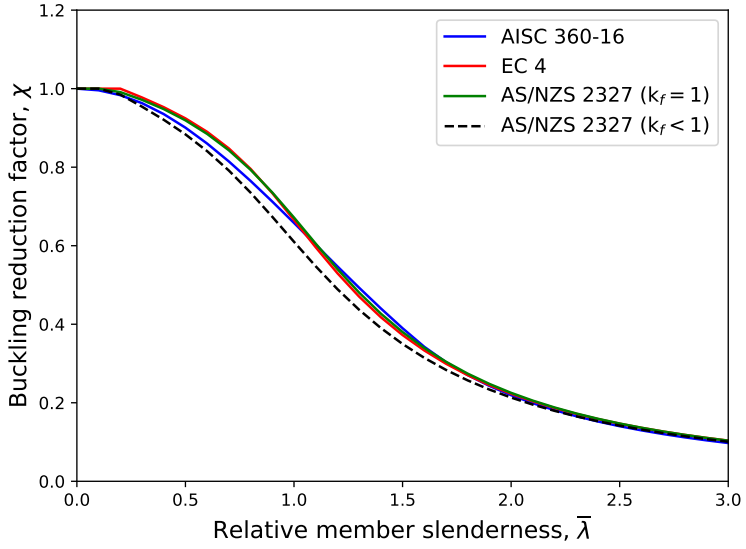


Figure 10: Column curves from the considered design codes.

4. Parameters setting and selection of training dataset amount

4.1. Parameters setting

A scikit-learn Python package `CatBoostRegressor` and a stratified 10-fold cross-validation are used to set the training model. Two hyper-parameters, maximum depth (`max_depth`) and learning rate (`learning_rate`) are chosen before the learning process begins. They should be appropriately selected to reach the best performance in the machine learning models. A grid search is the most popular method to find the best optimal combination. The distributed grid search using `GridSearchCV` of scikit-learn in Python ([45]) is implemented in this study. Fig. 11 shows the scores of tuning hyper-parameters heat, where their R^2 scores can be identified. The best combination of two hyper-parameters can be found by the brightest colour. `max_depth`=4 is used for whole dataset but `learning_rate` is different and depends on each dataset, which is 0.3, 0.05, 0.1 and 0.05 for Datasets 1-4, respectively.

4.2. Selection of training dataset amount

The accuracy of the machine learning model relies on the training data amount. In order to evaluate the test accuracy of CatBoost, the following performance measures are used.

- Coefficient of determination (R^2)

$$R^2 = \left(\frac{n \sum ty - (\sum t)(\sum y)}{\sqrt{n(\sum t^2) - (\sum t)^2}\sqrt{n(\sum y^2) - (\sum y)^2}} \right)^2 \tag{2}$$

  where t and y are the target and predicted values and n indicates the number of data samples.

- Mean Square Error (MSE)

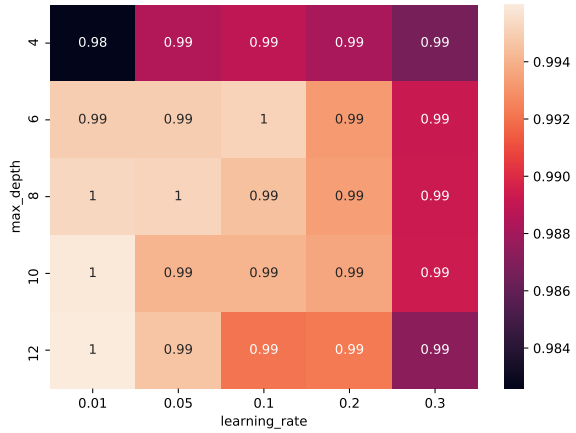
$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(t_i - y_i)^2 \tag{3}$$

- Root Mean Squared Error (RMSE)

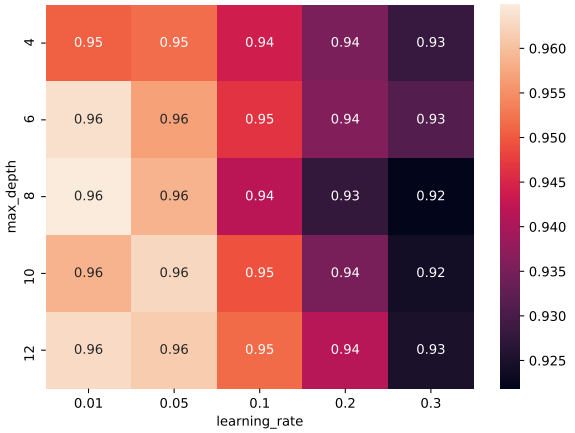
$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(t_i - y_i)^2} \tag{4}$$

- Mean Absolute Error (MAE)

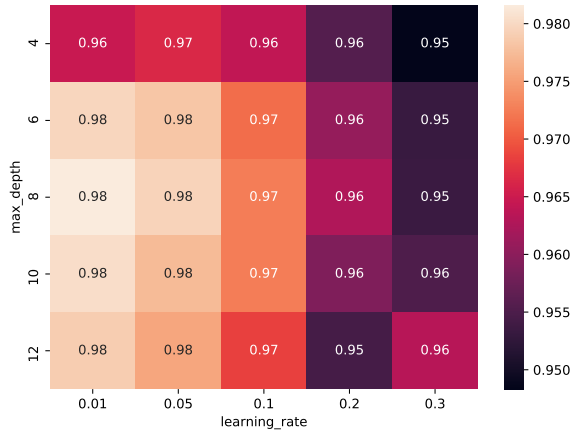
$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|t_i - y_i| \tag{5}$$



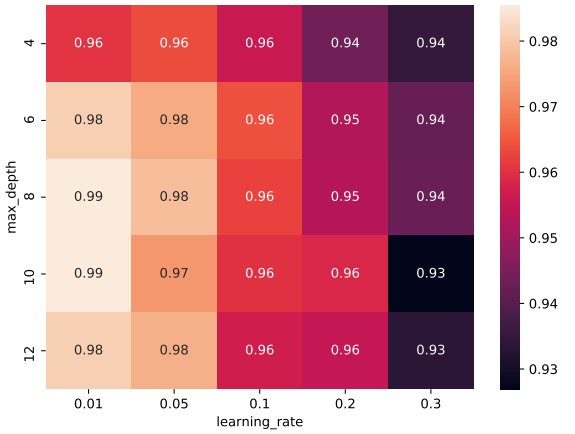
(a) Dataset 1 (`max_depth`= 4; `learning_rate`= 0.3)



(b) Dataset 2 (`max_depth`= 4; `learning_rate`= 0.05)



(c) Dataset 3 (`max_depth`= 4; `learning_rate`= 0.1)



(d) Dataset 4 (`max_depth`= 4; `learning_rate`= 0.05)

Figure 11: The best combination of hyper-parameters (`max_depth` and `learning_rate`) in each dataset.

- Coefficient of Variation (CoV)

$$\text{CoV} = \frac{\text{RMSE}}{\text{Mean}} \tag{6}$$

The results of three cases, in which the dataset is divided for 9–1, 8.75-1.25 and 8–2 as training and testing, are given in Table 2. Dataset is divided equally into 10 subsets, in which 9, 8.75 and 8 subsets are used to build the strong learner while the rest ones (1, 1.25 and 2) are used to validate the model. It is clear to see that all results have very high accuracy in which R^2 are very close to unity. As the training data amount increases, RMSE and MAE reduce. For the test of Dataset 1, when the training data increases from 80% to 90%, R^2 increases from 0.995 to 0.996 and RMSE and MAPE decrease from 280.77, 119.142 to 231.152, 109.345. The results from 9-1 test give the best prediction and are used in the following sections.

5. Results and discussion

Several machine learning models are used to train four datasets, which are given in Table 1, to predict the strength of the CFST columns. The results are validated against a comprehensive database of 3103 tests covering a wide range of material and geometric properties as well as loading configurations. The ratio of the strength obtained from experiments to that from CatBoost model and design codes are given in tables and plotted in figures.

5.1. Comparison of CatBoost with several machine learning models

The performance of CatBoost is demonstrated by comparing with several machine learning models including Classification And Regression Tree (CART), Adaptive Boosting (AdaBoost), Gradient Boosting (GB), Random Forest (RF), Extreme Gradient Boosting (XGBoost), Light Gradient Boosting (LightGB), Neural Network (NN) and Support Vector Machine (SVM). It can be observed from Table 3 that five boosting ensemble methods (AdaBoost, GB, XGBoost, LightGB and CatBoost) as well as bagging ensemble method (RF) give very high accuracy in prediction. Among them, CatBoost shows an obvious superior over single learning methods (NN and SVM) for all the four measures. For example, R^2 increase from 0.962 (NN) to 0.9996 for Dataset 1, while the MAE decrease from 298.015 (NN) to 109.345. The performance of CatBoost is improved significantly. Based on the above observation, it can be concluded that CatBoost provides the best performance with the smallest error for these datasets.

Table 2: Different training dataset amount results using CatBoost.

Dataset	Split	Performance indicators			
		R^2	MSE	RMSE	MAE
1	9-1 Train	1.000	4080.250	63.801	33.277
	9-1 Test	0.996	57894.073	231.152	109.345
	8.75-1.25 Train	1.000	4035.832	63.437	32.526
	8.75-1.25 Test	0.996	61694.425	243.242	113.089
	8-2 Train	1.000	3728.278	60.921	31.085
	8-2 Test	0.995	87313.519	280.770	119.142
2	9-1 Train	0.999	911.164	30.173	22.135
	9-1 Test	0.964	24319.791	137.579	60.895
	8.75-1.25 Train	0.999	913.452	30.200	22.039
	8.75-1.25 Test	0.963	25222.807	143.864	60.976
	8-2 Train	0.999	867.673	29.439	21.392
	8-2 Test	0.965	23680.972	144.530	64.009
3	9-1 Train	0.997	10702.576	103.309	63.287
	9-1 Test	0.982	70021.877	259.090	141.312
	8.75-1.25 Train	0.997	10423.678	101.905	62.609
	8.75-1.25 Test	0.979	80173.306	274.818	144.565
	8-2 Train	0.997	9608.625	97.796	60.993
	8-2 Test	0.978	81685.599	283.340	152.710
4	9-1 Train	0.999	1479.532	38.426	29.519
	9-1 Test	0.981	21477.168	132.306	80.922
	8.75-1.25 Train	0.999	1480.014	38.432	29.392
	8.75-1.25 Test	0.976	22171.287	139.674	84.575
	8-2 Train	0.999	1463.175	38.163	29.362
	8-2 Test	0.973	28452.692	163.148	93.044

Table 3: Comparison between the performance of CatBoost with several machine learning models.

Dataset	Reference	R^2	MSE	RMSE	MAE
1	Present (CatBoost)	0.996	57894.073	231.152	109.345
	CART	0.978	225875.636	444.333	154.754
	AdaBoost	0.984	146579.192	372.080	249.915
	GB	0.994	79287.530	267.877	109.291
	RF	0.982	293742.469	501.507	209.865
	XGBoost	0.995	66962.798	243.538	121.858
	LightGB	0.990	89441.261	287.131	125.247
	NN	0.962	547184.312	687.431	298.015
	SVM	0.976	247092.407	488.957	286.563
2	Present (CatBoost)	0.964	24319.791	137.579	60.895
	CART	0.896	48660.790	207.621	89.571
	AdaBoost	0.939	37859.434	182.122	86.809
	GB	0.921	40081.254	182.728	68.753
	RF	0.934	33981.481	178.523	85.407
	XGBoost	0.949	32933.484	160.048	71.955
	LightGB	0.952	34176.014	164.268	77.721
	NN	0.929	37601.985	184.895	101.777
	SVM	0.907	55596.804	201.024	89.400
3	Present (CatBoost)	0.982	70021.877	259.090	141.312
	CART	0.933	237713.990	473.991	200.162
	AdaBoost	0.926	268370.394	510.002	386.042
	GB	0.975	97241.874	297.450	137.456
	RF	0.966	129475.199	343.567	188.431
	XGBoost	0.981	72756.248	264.612	145.719
	LightGB	0.980	78150.591	271.573	136.580
	NN	0.954	177596.574	396.922	263.501
	SVM	0.967	120339.029	341.573	213.207
4	Present (CatBoost)	0.981	21477.168	132.306	80.922
	CART	0.871	137417.275	324.806	165.827
	AdaBoost	0.950	53706.034	203.213	110.783
	GB	0.973	30498.199	158.949	85.764
	RF	0.946	58662.214	225.695	151.225
	XGBoost	0.977	23580.041	140.923	82.247
	LightGB	0.926	65191.574	238.238	135.531
	NN	0.952	49963.143	176.497	97.809
	SVM	0.795	207787.040	324.207	155.860

5.2. Comparison of predicted capacity

Table 4 shows the comparison of the present results from 10-fold cross validation and those from the code predictions (AISC 360-16, Eurocode 4 and AS/NZS 2327) as well as previous study [24] using genetic algorithms (GA) for four datasets. All present results show very high prediction accuracy in terms of coefficient of determination (R^2), which is the lowest value (R^2=0.964) for Dataset 2 and the highest one (R^2=0.996) for Dataset 1. It should be noted that R^2 is close to unity indicating high accuracy of model. Evidently, compared with the results performed by previous researchers [24], the present model has a much better performance for all the five measures. Comparing error metrics (MAE and RMSE) also illustrates well performance with a remarkable improvement of the present model. For example, with the nearly same R^2=0.996 and 0.993 with Eurocode 4, CatBoost obtains a lower RMSE, MAE (231.152, 109.345) compared to that (347.536, 193.663), which means that their improvement is 33.48% and 43.53 %, respectively. The relation between the measured output and predicted one is almost linear as can be seen in Fig. 12. It implies that the present predictions are close to the experiment ones. Overall, present model accurately captures the strength of concentrically and eccentrically loaded CFST columns with a better overall performance compared with the code predictions.

Other key statistics in Table 5 and Fig. 13 show the high predictive capability of present model compared with the code predictions. The present percentage in the error range of 5% is 91.75%, 84.74%, 86.93% and 88.32% for Datasets 1-4 as well as 91.08% for whole dataset, respectively. This number depends on the code predictions and is varied in the range of 12.62%-54.03%. It should be mentioned that the number of predictions that exceeded 5% the actual value in experiment tests shows how design codes overestimate the strength of the CFST columns by a large margin. This shows the need for machine learning algorithms in order to better predictive capability and accuracy. It can overcome many limitations associated with those adopted in design codes which do not account for specific material strengths and slenderness ratio.

The effect of the material strengths and slenderness ratio on the strength predictions of the present model and code predictions are shown in Figs. 14-16. It is clear to see that the results from three codes underestimate the column capacity. Among them, Eurocode 4 and AS/NZS 2327 give the best predictions, whereas, the most conservative ones are for AISC 360-16. For example, their mean values are 1.094, 1.091 and 1.272 for Dataset 1 (Table 4). While the

predictions from three codes beyond material limit and slenderness are less conservative than those within it [1], CatBoost provides nearly similar experiment results with the mean values as unity without any limits. It highlights the novelty of machine learning in general and CatBoost in particular to predict the strength of the CFST columns.

5.3. Effects of input variable sensitivities and input variable number

The sensitivity of each input variable and effect of the number of input variable on the strength of the CFST columns are studied in Fig. 17 and Table 6. In the feature selection methods, the feature importance ranking is calculated on the training data using the relationship between input and output variables [46]. Here, a Gini variable importance measure [47], which is commonly used as the splitting criterion in classification tress, has been selected. Three geometric features (D, H, B) are the most important factors that have strong effect on the final prediction since they are the main contributor to the cross-sectional area of CFST columns. For circular section, the relative feature important of D is 69.32% and 63.962% for Datasets 1 and 2. For rectangular section, it is much lower and depends on concentric or eccentric loading. The most critical one for Dataset 3 is B with 36.575% and for Dataset 4 is H with 25.301%. Steel yield strength (f_y) and concrete compressive strength (f'_c) have higher relative important for rectangular section than circular one.

Finally, effect of the input variable number on the predictions is analysed. The intention is to investigate how the model performs if some input variables are missing. Table 6 presents the results with various combinations, which include original data (combination 1) and neglect some input variables (combination 2-6). As expected, the accuracy is dropped in combination 2 for Dataset 1 and 2 when the D is not considered. The lowest accuracy (R^2=0.789) is found for combination 3 for Dataset 2 when both D and f'_c, which are the top two relative feature important (Fig. 17), are not available. The same trend can be observed for Dataset 3 and 4. In summary, three geometric features (D, H, B) should be included in the model to get a high accuracy prediction.

6. Conclusion

In this paper, CatBoost is used to predict the strength of the CFST columns. Four datasets collected from the literatures, which dealt with different attributes, are trained. Violin plots with distribution and correlation matrix heat map of features are plotted to see their relationship in these datasets. The comparison of the CatBoost from 10-fold cross validation and those

from the code predictions (AISC 360-16, Eurocode 4 and AS/NZS 2327) shows very high prediction accuracy. The coefficient of determination (R^2) from CatBoost, which is the lowest value (R^2=0.964) for Dataset 2 and the highest one (R^2=0.996) for Dataset 1. The present percentage in the error range of 5% is 91.75%, 84.74%, 86.93% and 88.32% for Datasets 1-4 as well as 91.08% for whole dataset, respectively. While the predictions from three codes beyond material limit and slenderness are less conservative than those within it, CatBoost provides nearly similar experiment results with the mean values as unity without any limits. This algorithm can overcome any limitations associated within design codes such as material strengths and slenderness ratio to provide an accurate solution.

Table 4: Comparison of the performance of CatBoost in four datasets with 10-fold cross validation.

Dataset	Reference	Mean	CoV	R^2	RMSE	MAE
1	Present	1.000	0.043	0.996	231.152	109.345
	GA ([24])	0.980	0.160	0.991	384.730	232.440
	AISC 360-16	1.272	0.179	0.986	1032.678	493.790
	Eurocode 4	1.094	0.149	0.993	347.536	193.663
	AS/NZS 2327	1.091	0.150	0.992	380.268	202.628
2	Present	1.000	0.051	0.964	137.579	60.895
	GA ([24])	1.200	0.170	0.976	126.860	93.560
	AISC 360-16	1.229	0.243	0.942	252.031	152.576
	Eurocode 4	1.083	0.175	0.959	173.366	92.387
	AS/NZS 2327	1.170	0.186	0.929	239.399	113.223
3	Present	1.000	0.038	0.982	259.090	141.312
	GA ([24])	1.020	0.130	0.977	126.860	91.560
	AISC 360-16	1.181	0.178	0.959	481.465	330.077
	Eurocode 4	1.069	0.182	0.954	442.438	270.814
	AS/NZS 2327	1.075	0.171	0.960	403.150	255.102
4	Present	1.000	0.022	0.981	132.306	80.922
	GA ([24])	1.260	0.180	0.954	218.760	168.420
	AISC 360-16	1.222	0.230	0.942	356.817	252.027
	Eurocode 4	1.056	0.192	0.938	251.648	155.515
	AS/NZS 2327	1.110	0.203	0.939	285.749	178.965

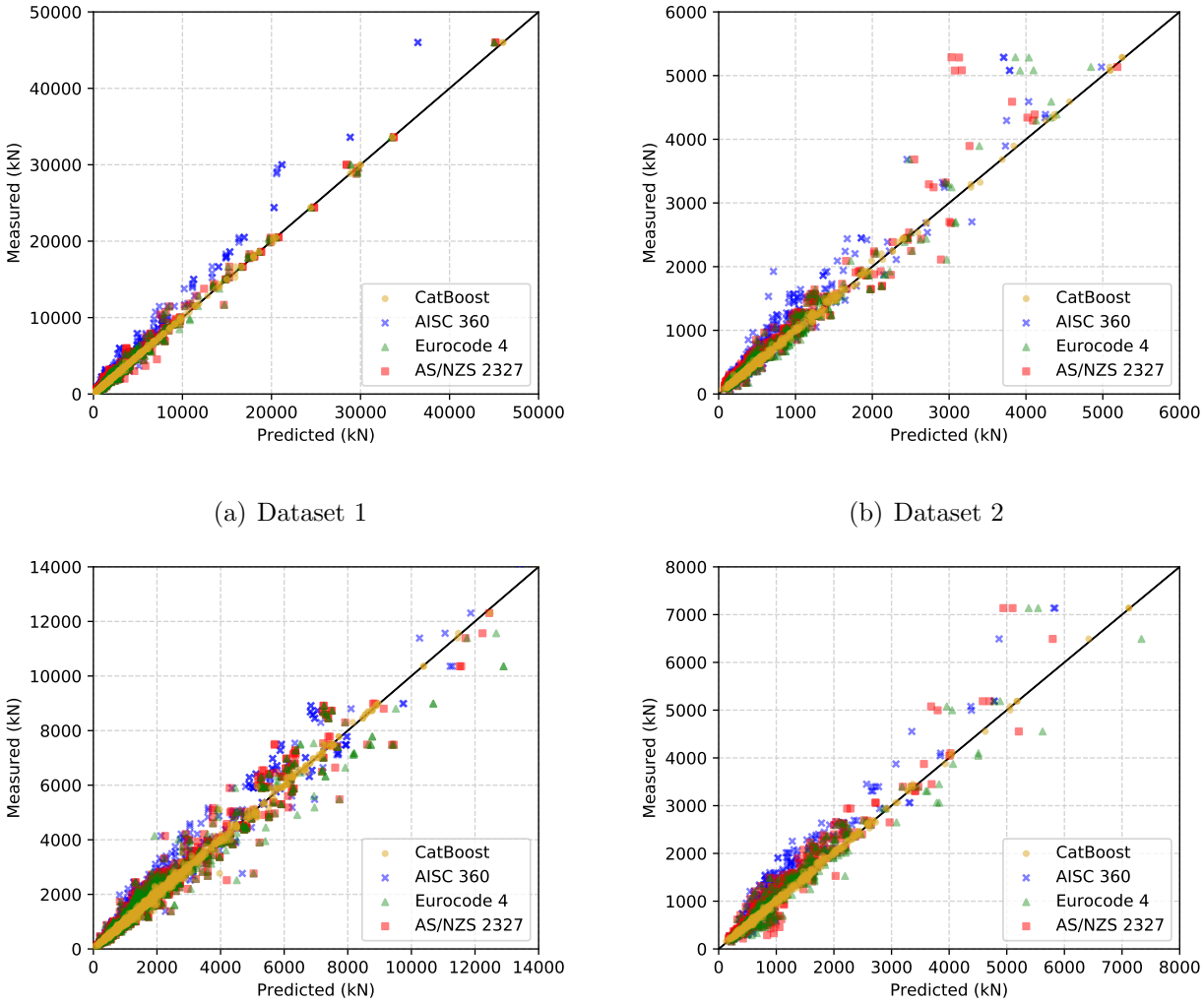
Acknowledgement

This research was supported by a grant (NRF-2020R1A4A2002855) from NRF (National Research Foundation of Korea) funded by MEST (Ministry of Education and Science Technology) of Korean government.

7. References

References

[1] S. Thai, H.-T. Thai, B. Uy, T. Ngo, Concrete-filled steel tubular columns: Test database, design and calibration, Journal of Constructional Steel Research 157 (2019) 161–181.



(a) Dataset 1

(b) Dataset 2

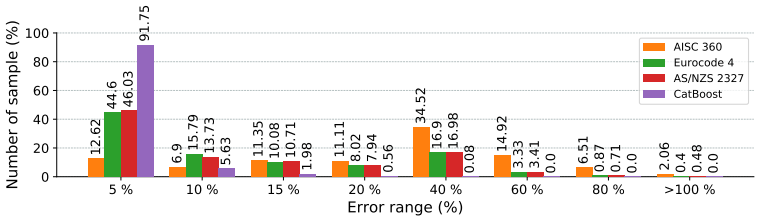
(c) Dataset 3

(d) Dataset 4

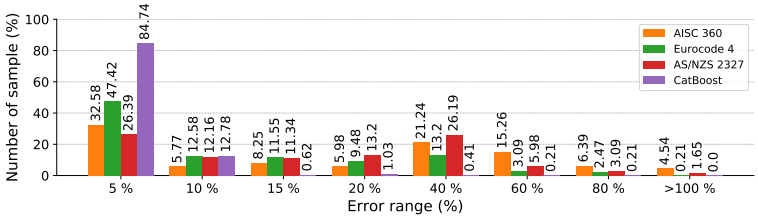
Figure 12: The correlation between actual and predicted output using the CatBoost algorithm.

Table 5: Number of test and corresponding percentage in the error range from CatBoost and code predictions

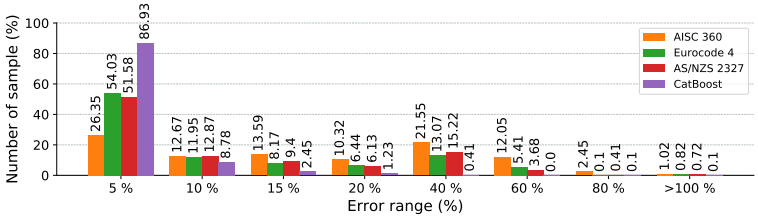
Error range	Dataset	AISC 360-16		Eurocode 4		AS/NZS 2327		Present	
(%)		No of test	(%)	No of test	(%)	No of test	(%)	No of test	(%)
5	1	159	12.62	562	44.60	580	46.03	1156	91.75
	2	158	32.58	230	47.42	128	26.39	411	84.74
	3	258	26.35	529	54.03	505	51.58	851	86.93
	4	105	26.65	207	52.54	160	40.61	348	88.32
10	1	87	6.90	199	15.79	173	13.73	71	5.63
	2	28	5.77	61	12.58	59	12.16	62	12.78
	3	124	12.67	117	11.95	126	12.87	86	8.78
	4	36	9.14	42	10.66	34	8.63	32	8.12
15	1	143	11.35	127	10.08	135	10.71	25	1.98
	2	40	8.25	56	11.55	55	11.34	3	0.62
	3	133	13.59	80	8.17	92	9.40	24	2.45
	4	33	8.38	42	10.66	40	10.15	9	2.28
20	1	140	11.11	101	8.02	100	7.94	7	0.56
	2	29	5.98	46	9.48	64	13.20	5	1.03
	3	101	10.32	63	6.44	60	6.13	12	1.23
	4	29	7.36	33	8.38	39	9.90	3	0.76



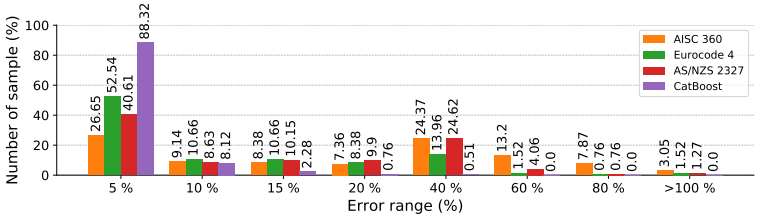
(a) Dataset 1



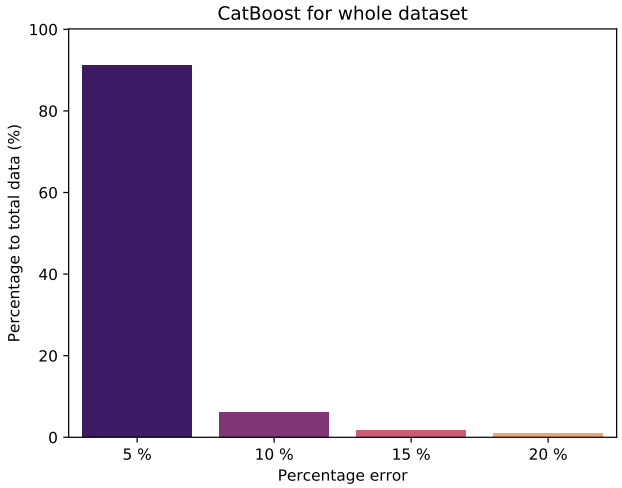
(b) Dataset 2



(c) Dataset 3

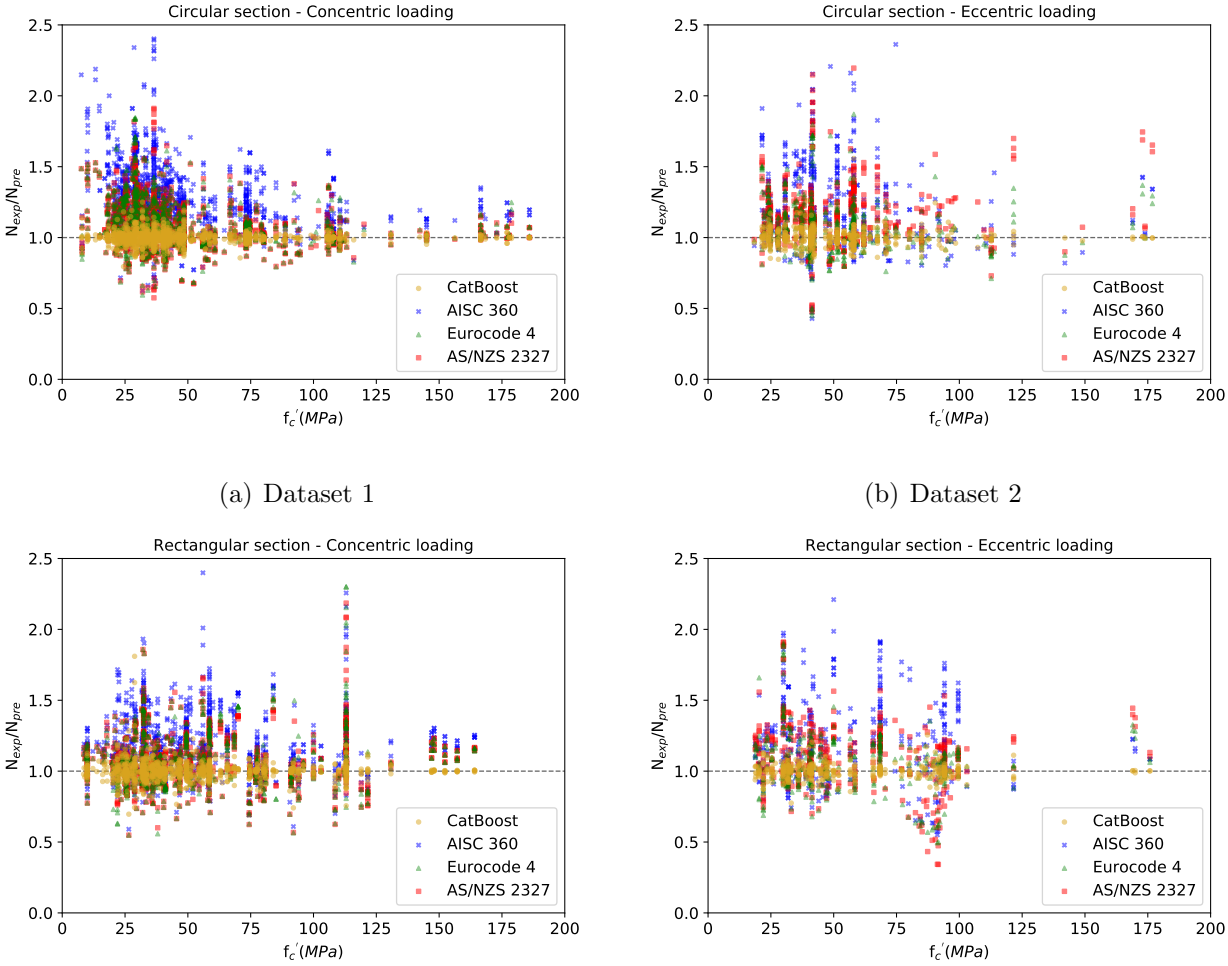


(d) Dataset 4



(e) Whole dataset

Figure 13: Distribution of the errors from CatBoost and code predictions.



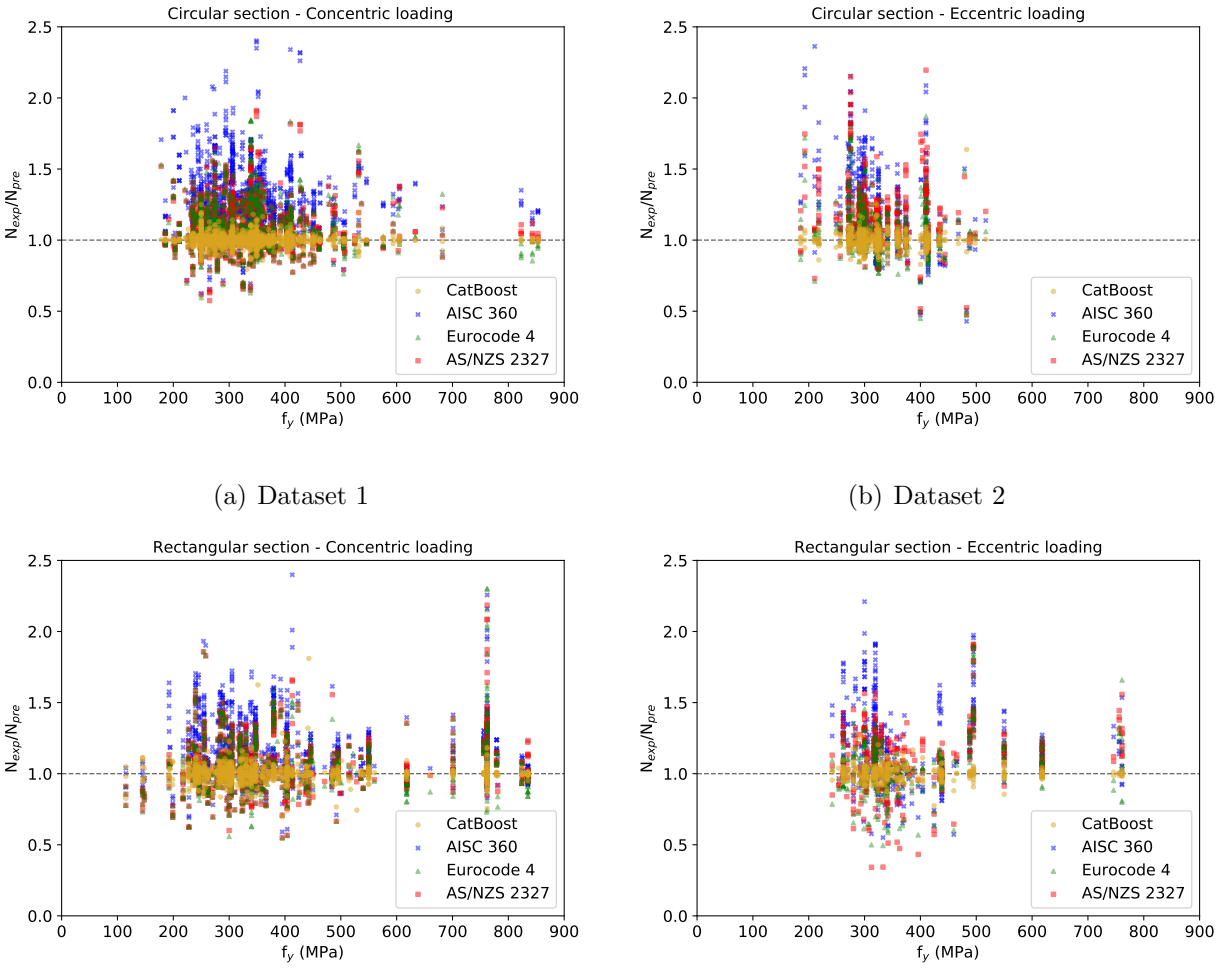
(a) Dataset 1

(b) Dataset 2

(c) Dataset 3

(d) Dataset 4

Figure 14: Effect of concrete compressive strength on CatBoost and code predictions for circular and rectangular columns.



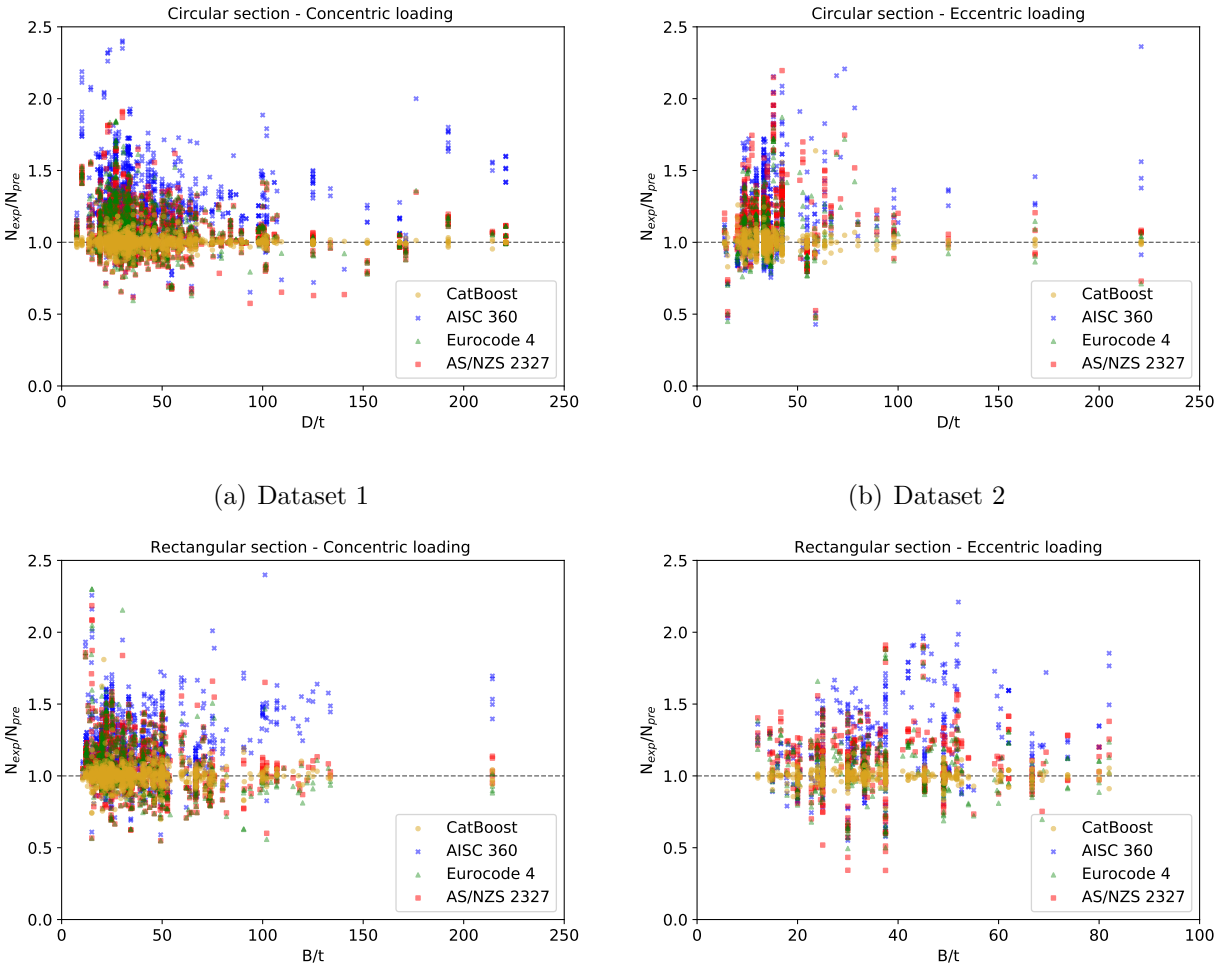
(a) Dataset 1

(b) Dataset 2

(c) Dataset 3

(d) Dataset 4

Figure 15: Effect of steel yield strength on CatBoost and code predictions for circular and rectangular columns.



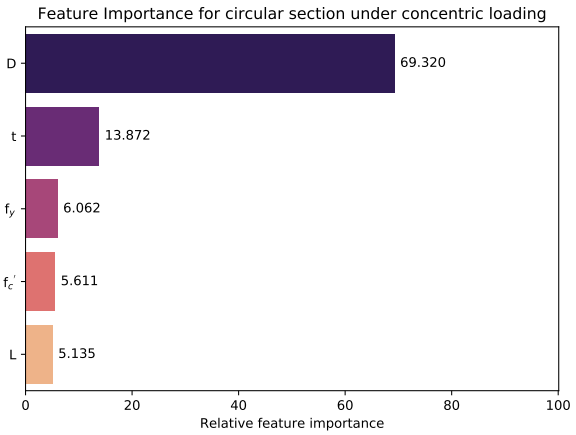
(a) Dataset 1

(b) Dataset 2

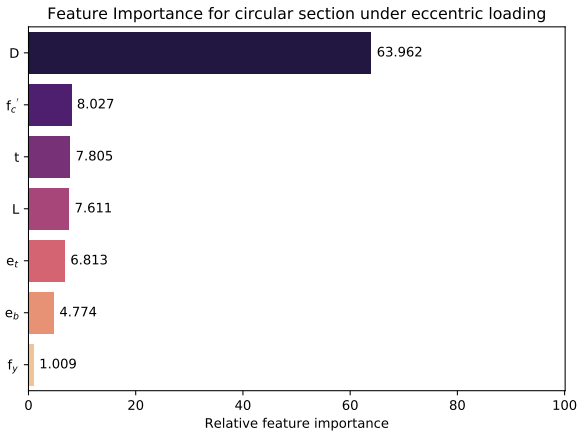
(c) Dataset 3

(d) Dataset 4

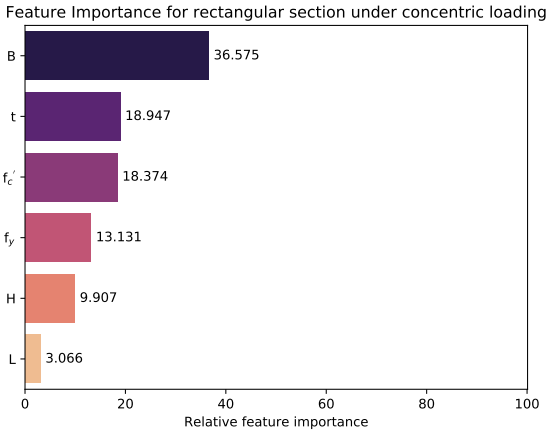
Figure 16: Effect of slenderness on CatBoost and code predictions for circular and rectangular columns.



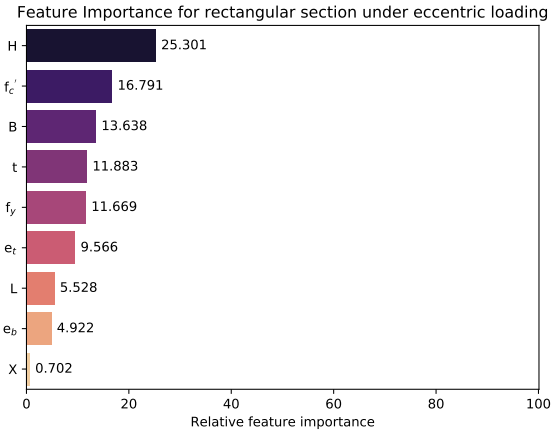
(a) Dataset 1



(b) Dataset 2



(c) Dataset 3



(d) Dataset 4

Figure 17: Relative importance of different input variables for circular and rectangular columns.

Table 6: Performance measures for analysis of different input variable combinations with 10-fold cross validation.

Dataset	Combinations	Performance indicators			
		R^2	MSE	RMSE	MAE
1	1: D, t, f_y, $f_c^{'}$, L	0.996	57894.073	231.152	109.345
	2: t, f_y, $f_c^{'}$, L	0.947	637816.116	760.461	291.023
	3: D, t, f_y, $f_c^{'}$	0.993	88872.197	292.030	156.155
	4: D, t, f_y	0.959	372518.992	602.872	282.293
2	1: D, $f_c^{'}$, t, L, e_t, e_b, f_y	0.964	24319.791	137.579	60.895
	2: $f_c^{'}$, t, L, e_t, e_b, f_y	0.864	76908.03	256.893	123.595
	3: t, L, e_t, e_b, f_y	0.789	113505.028	311.995	155.115
	4: D, $f_c^{'}$, t, L, e_t, e_b	0.957	30157.339	153.315	63.816
	5: D, $f_c^{'}$, t, L, e_t	0.960	28473.703	148.125	64.271
	6: D, $f_c^{'}$, t, L	0.847	81733.416	281.241	165.814
3	1: B, t, $f_c^{'}$, f_y, H, L	0.982	70021.877	259.090	141.312
	2: t, $f_c^{'}$, f_y, H, L	0.978	84389.639	284.052	164.144
	3: t, f_y, H, L	0.921	296603.150	535.334	307.185
	4: B, t, $f_c^{'}$, f_y, H	0.969	115826.524	332.788	181.346
	5: B, t, $f_c^{'}$, f_y	0.966	128715.266	353.863	208.358
	6: B, t, $f_c^{'}$	0.956	158630.305	396.024	241.177
4	1: H, $f_c^{'}$, B, t, f_y, e_t, L, e_b, X	0.981	21477.168	132.306	80.922
	2: $f_c^{'}$, B, t, f_y, e_t, L, e_b, X	0.967	34102.685	168.556	104.994
	3: B, t, f_y, e_t, L, e_b, X	0.944	52376.630	216.121	134.610
	4: H, $f_c^{'}$, B, t, f_y, e_t, L, e_b	0.982	20415.825	129.284	79.542
	5: H, $f_c^{'}$, B, t, f_y, e_t, L	0.981	20481.847	134.018	87.571
	6: H, $f_c^{'}$, B, t, f_y, e_t	0.969	28858.254	162.289	108.525

[2] K. Klöppel, W. Goder, Traglastversuche mit ausbetonierten stahlrohen und aufstellung einer bemessungsformel, Der Stahlbau 26 (1) (1957) 1–10.

[3] R. B. Knowles, R. Park, Strength of concrete filled steel columns, Journal of the structural division 95 (1969) 2565–2588.

[4] M. Denavit, Steel-concrete composite column database.

[5] Z. Tao, U. Brian, L.-H. Han, S.-H. He, Design of concrete-filled steel tubular members according to the australian standard as 5100 model and calibration, Australian Journal of Structural Engineering 8 (3) (2008) 197–214.

[6] C. D. Goode, Composite columns-1819 tests on concrete-filled steel tube columns compared with eurocode 4, The Structural Engineer 86 (16) (2008) 33–38.

[7] R. Leon, T. Perea, J. Hajjar, M. Denavit, Concrete-filled tubes columns and beam-columns: a database for the aisc 2005 and 2010 specifications, Festschrift Gerhard Hanswille 20 (2011) 203–212.

[8] J. Hajjar, B. Gourley, C. Tort, M. Denavit, P. Schiller, N. L. Mundis, Steel-concrete composite structural systems, Department of Civil and Environmental Engineering, Northeastern University.

[9] J. R. Liew, M. Xiong, D. Xiong, Design of concrete filled tubular beam-columns with high strength steel and concrete, in: Structures, Vol. 8, Elsevier, 2016, pp. 213–226.

[10] M.-X. Xiong, D.-X. Xiong, J. R. Liew, Axial performance of short concrete filled steel tubes with high-and ultra-high-strength materials, Engineering Structures 136 (2017) 494–510.

[11] M. Khan, B. Uy, Z. Tao, F. Mashiri, Behaviour and design of short high-strength steel welded box and concrete-filled tube (CFT) sections, Engineering Structures 147 (2017) 458–472.

[12] B. Uy, Strength of short concrete filled high strength steel box columns, Journal of Constructional Steel Research 57 (2) (2001) 113–134.

[13] M. Mursi, B. Uy, Strength of slender concrete filled high strength steel box columns, Journal of Constructional Steel Research 60 (12) (2004) 1825–1848.

[14] K. Sakino, H. Nakahara, S. Morino, I. Nishiyama, Behavior of centrally loaded concrete-filled steel-tube short columns, Journal of structural engineering 130 (2) (2004) 180–188.

[15] AISC 360-16. Specification for structural steel buildings; 2016.

[16] BS5400-5. Steel, concrete and composite bridges. Code of practice for design of composite bridges. British Standards Institute (BSI); 2005.

[17] EN1994-1-1. Eurocode 4: Design of composite steel and concrete structures - Part 1-1: General rules and rules for buildings; 2004.

[18] Architectural Institute of Japan (AIJ). Recommendations for design and construction of concrete filled steel tubular structures, Japan; 1997.

[19] GB 50936. Technical code for concrete-filled steel tubular structures. China National Standards; 2014.

[20] Standards Australia. AS/NZS 2327 Composite structures - Composite steel-concrete construction in buildings; 2017.

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

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

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

[24] M. Naser, S. Thai, H.-T. Thai, Evaluating structural response of concrete-filled steel tubular columns through machine learning, Journal of Building Engineering (2020) 101888doi:https://doi.org/10.1016/j.jobe.2020.101888.

[25] S. Jegadesh, S. Jayalekshmi, Application of artificial neural network for calculation of axial capacity of circular concrete filled steel tubular columns, Int. J. Earth Sci. Eng 8 (2015) 35–42.

[26] M. Ahmadi, H. Naderpour, A. Kheyroddin, ANN Model for Predicting the Compressive Strength of Circular Steel-Confined Concrete, International Journal of Civil Engineering 15 (2) (2017) 213–221.

[27] Y. Du, Z. Chen, C. Zhang, X. Cao, Research on axial bearing capacity of rectangular concrete-filled steel tubular columns based on artificial neural networks, Frontiers of Computer Science 11 (5) (2017) 863–873.

[28] V.-L. Tran, S.-E. Kim, Efficiency of three advanced data-driven models for predicting axial compression capacity of CFDST columns, Thin-Walled Structures 152 (2020) 106744. doi:https://doi.org/10.1016/j.tws.2020.106744.

[29] V.-L. Tran, D.-K. Thai, D.-D. Nguyen, 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 (2020) 106720. doi:https://doi.org/10.1016/j.tws.2020.106720.

[30] Q. Ren, M. Li, M. Zhang, Y. Shen, W. Si, Prediction of ultimate axial capacity of square concrete-filled steel tubular short columns using a hybrid intelligent algorithm, Applied Sciences 9 (14) (2019) 2802.

[31] S. H. Mai, M. E. A. Ben Seghier, P. L. Nguyen, J. Jafari-Asl, D.-K. Thai, A hybrid model for predicting the axial compression capacity of square concrete-filled steel tubular columns, Engineering with Computersdoi:https://doi.org/10.1007/s00366-020-01104-w.

[32] Z.-H. Zhou, Ensemble methods: foundations and algorithms, CRC press, 2012.

[33] A. V. Dorogush, V. Ershov, A. Gulin, Catboost: gradient boosting with categorical features support, ArXiv abs/1810.11363.

[34] F. Stulp, O. Sigaud, Many regression algorithms, one unified model: A review, Neural Networks 69 (2015) 60–79.

[35] D.-C. Feng, Z.-T. Liu, X.-D. Wang, Y. Chen, J.-Q. Chang, D.-F. Wei, Z.-M. Jiang, Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach, Construction and Building Materials 230 (2020) 117000. doi:https://doi.org/10.1016/j.conbuildmat.2019.117000.

[36] T. G. Dietterich, Ensemble methods in machine learning, in: Multiple Classifier Systems, Springer Berlin Heidelberg, Berlin, Heidelberg, 2000, pp. 1–15.

[37] L. Breiman, Bagging predictors, Machine Learning 24 (2) (1996) 123–140.

[38] R. E. Schapire, The strength of weak learnability, Machine learning 5 (2) (1990) 197–227.

[39] D. H. Wolpert, Stacked generalization, Neural networks 5 (2) (1992) 241–259.

[40] J. H. Friedman, Greedy function approximation: a gradient boosting machine, Annals of statistics (2001) 1189–1232.

[41] M. H. D. M. Ribeiro, L. dos Santos Coelho, Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series, Applied Soft Computing 86 (2020) 105837.

[42] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, A. Gulin, Catboost: unbiased boosting with categorical features, in: Advances in neural information processing systems, 2018, pp. 6638–6648.

[43] A. V. Dorogush, V. Ershov, A. Gulin, Catboost: gradient boosting with categorical features support, arXiv preprint arXiv:1810.11363.

[44] C. Bentéjac, A. Csörgő, G. Martínez-Muñoz, A comparative analysis of gradient boosting algorithms, Artificial Intelligence Review (2020) 1–31.

[45] M. Feurer, K. Eggensperger, S. Falkner, M. Lindauer, F. Hutter, Auto-sklearn 2.0: The next generation, arXiv preprint arXiv:2007.04074.

[46] Hanchuan Peng, Fuhui Long, C. Ding, Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (8) (2005) 1226–1238. doi:10.1109/TPAMI.2005.159.

[47] A.-L. Boulesteix, A. Bender, J. Lorenzo Bermejo, C. Strobl, Random forest Gini importance favours SNPs with large minor allele frequency: impact, sources and recommendations, Briefings in Bioinformatics 13 (3) (2011) 292–304. arXiv:https://academic.oup.com/bib/article-pdf/13/3/292/679494/bbr053.pdf, doi:10.1093/bib/bbr053.