СТРОИТЕЛЬНАЯ МЕХАНИКА ИНЖЕНЕРНЫХ КОНСТРУКЦИЙ И СООРУЖЕНИЙ
STRUCTURAL MECHANICS OF ENGINEERING CONSTRUCTIONS AND BUILDINGS

2025. 21(6). 509–523
ISSN 1815-5235 (Print), 2587-8700 (Online)
HTTP://JOURNALS.RUDN.RU/STRUCTURAL-MECHANICS

DOI: 10.22363/1815-5235-2025-21-6-509-523
EDN: ECJWUG

Research article / Научная статья

Rectangular Concrete-Filled Steel Tube Rational Dimensions under Uniaxial Eccentric Compression

Anton S. Chepurnenko✉, **Samir H. Al-Zgul**, **Batyr M. Yazyev**

Don State Technical University, *Rostov-on-Don, Russian Federation*
✉ anton_chepurnenk@mail.ru

Received: July 20, 2025
Revised: August 27, 2025
Accepted: September 5, 2025

Abstract. An algorithm for generating the training dataset and the machine learning model for selecting the cross-sectional dimensions of eccentrically compressed concrete filled steel tubular (CFST) columns have been developed. The paper presents a predictive model based on the CatBoost algorithm for determining the optimal geometric parameters (width b and height h) of the cross-section of rectangular CFST columns in compliance with regulatory strength requirements. The input parameters used were the concrete compressive strength class B according to Russian standards, the magnitude of the longitudinal force F, the wall thickness of the steel section t and the eccentricity of load application e. The model was trained on a synthetic sample formed taking into account the conditions of limit equilibrium under the combined action of the axial force and bending moment, restrictions on the cross-sectional dimensions in the range from 100 to 500 mm, strength conditions, as well as the requirements for minimizing the cost of the structure. The application of the CatBoost algorithm allowed achieving high forecasting accuracy with an average of two target variable metrics: the determination coefficient $R^2 = 0.999122$ and the average error in determining the section dimensions of 2.485 mm. The obtained results demonstrate the significant potential for using the developed model in the practical activities of design organizations, ensuring the accuracy of calculations while simultaneously optimizing material costs and reducing the time for implementing design solutions.

Keywords: bearing capacity, limit equilibrium, cross-section optimization, machine learning, Catboost

Conflicts of interest. The authors declare that there is no conflict of interest.

Authors' contribution: *Chepurnenko A.S.* — software, writing, review and editing; *Al-Zgul S.H.* — conducting research, software, data processing, graphic design, text writing; *Yazyev B.M.* — supervision, conceptualization. All of the authors read and approved the final version of the article.

Data Availability Statement. The training dataset is available for download at: https://doi.org/10.13140/RG.2.2.10073.99683

For citation: Chepurnenko A.S., Al-Zgul S.H., Yazyev B.M. Rectangular concrete-filled steel tube rational dimensions under uniaxial eccentric compression. *Structural Mechanics of Engineering Constructions and Buildings.* 2025;21(6):509–523. http://doi.org/10.22363/1815-5235-2025-21-6-509-523 EDN: ECJWUG

Anton S. Chepurnenko, Doctor of Technical Sciences, Professor of the Department of Structural Mechanics and Theory of Structures, Don State Technical University, 1 Gagarin Sq., Rostov-on-Don, 344003, Russian Federation; eLIBRARY SPIN-code: 7149-7981, ORCID: 0000-0002-9133-8546; e-mail: anton_chepurnenk@mail.ru

Samir H. Al-Zgul, Postgraduate student of the Department of Structural Mechanics and Theory of Structures, Don State Technical University, 1 Gagarin Sq., Rostov-on-Don, 344003, Russian Federation; eLIBRARY SPIN-code: 4483-8340, ORCID: 0000-0001-6182-786X; e-mail: samiralzgulfx@gmail.com

Batyr M. Yazyev, Doctor of Technical Sciences, Professor of the Department of Structural Mechanics and Theory of Structures, Don State Technical University, 1 Gagarin Sq., Rostov-on-Don, 344003, Russian Federation; eLIBRARY SPIN-code: 5970-5350, ORCID: 0000-0002-5205-1446; e-mail: ps62@yandex.ru

Рациональные размеры прямоугольной трубобетонной колонны при внецентренном сжатии

А.С. Чепурненко ✉, С.Х. Аль-Згуль, Б.М. Языев

Донской государственный технический университет, *Ростов-на-Дону, Российская Федерация*
✉ anton_chepurnenk@mail.ru

Поступила в редакцию: 20 июля 2025 г.
Доработана: 27 августа 2025 г.
Принята к публикации: 5 сентября 2025 г.

Аннотация. Разработан алгоритм формирования обучающего датасета, а также модель машинного обучения для подбора размеров поперечного сечения внецентренно сжатых трубобетонных колонн. Представлена прогнозная модель на основе алгоритма CatBoost для определения оптимальных геометрических параметров (ширины b и высоты h) поперечного сечения прямоугольных трубобетонных колонн с соблюдением нормативных требований по прочности. В качестве входных параметров использованы класс бетона по прочности на сжатие B согласно российским стандартам, величина продольной силы F, толщина стенки стального профиля t и эксцентриситет приложения нагрузки e. Обучение модели проводилось на синтетической выборке, сформированной с учетом условий предельного равновесия при комбинированном действии продольной силы и изгибающего момента, ограничений на габариты сечения в диапазоне от 100 до 500 мм, условия прочности, а также требования минимизации стоимости конструкции. Применение алгоритма CatBoost позволило достичь высокой точности прогнозирования с усредненным по двум целевым переменным метрикам: коэффициентом детерминации $R^2 = 0{,}999122$ и средней ошибкой определения размеров сечения 2,485 мм. Полученные результаты демонстрируют значительный потенциал использования разработанной модели в практической деятельности проектных организаций, обеспечивая точность расчетов при одновременной оптимизации материальных затрат и сокращении времени выполнения проектных решений.

Ключевые слова: внецентренное сжатие, трубобетонная колонна, несущая способность, предельное равновесие, оптимизация размеров поперечного сечения, машинное обучение, Catboost

Заявление о конфликте интересов. Авторы заявляют об отсутствии конфликта интересов.

Заявление о доступности данных. Обучающий датасет доступен для скачивания по ссылке: https://doi.org/10.13140/RG.2.2.10073.99683

Вклад авторов: *Чепурненко А.С.* — программное обеспечение, подготовка текста статьи, рецензирование и редактирование; *Аль-Згуль С.Х.* — проведение исследования, программное обеспечение, обработка данных, графическое оформление, написание текста статьи; *Языев Б.М.* — общее научное руководство, формулировка концепции исследования. Все авторы ознакомлены с окончательной версией статьи и одобрили ее.

Для цитирования: *Чепурненко А.С., Аль-Згуль С.Х.., Языев Б.М.* Рациональные размеры прямоугольной трубобетонной колонны при внецентренном сжатии // Строительная механика инженерных конструкций и сооружений. 2025. Т. 21. № 6. С. 509–523. http://doi.org/10.22363/1815-5235-2025-21-6-509-523 EDN: ECJWUG

1. Introduction

The paper investigates rectangular concrete filled steel tube (CFST) columns under uniaxial eccentric compression. The widespread use of such structures in modern construction is explained by a number of key advantages: the synergistic effect of the combined resistance of the steel tube and concrete filling [1], increased fire resistance [2], high energy capacity under dynamic loads and economic efficiency at all stages of construction and operation of the structure. Due to their high spatial rigidity and ability to resist combined loads, these structures are widely used in construction as load-bearing elements both in Russia [3] and abroad [4]. However, their widespread implementation faces serious difficulties due to the lack of reliable calculation methods that adequately take into account the complex nature of the interaction between the steel shell and concrete filling under various types of loading.

The foundation of modern research in the field of CFST column analysis is experimental data obtained during field and laboratory tests [5; 6]. This data serves as the basis for the development of more accurate analytical methods that take into account the nonlinear behavior of materials [7]. For a detailed stress analysis, researchers widely use numerical modeling, such as the finite element method. To work effectively with growing amounts of information, specialized databases are created to systematize experimental results [8–10]. Since the start of the 21st century, the use of artificial intelligence and machine learning has been actively developing. These technologies are applied to predict the load-bearing capacity of structures and optimize their parameters. Taken together, all these areas allow to comprehensively study the behavior of composite structures and develop improved calculation methods.

The development of machine learning methods for designing CFST columns is largely driven by fundamental limitations of both normative and numerical calculation methods. On one hand, traditional normative approaches (Eurocode 4[1], AISC 360-16[2], SP 266.1325800.2016[3]), despite their widespread use in design practice, have significant drawbacks: they are only applicable to a narrow range of material characteristics, do not allow for reverse design tasks, are limited to certain types of cross-sections, and do not take into account complex loading cases.

On the other hand, numerical methods implemented in existing finite element software (ABAQUS, ANSYS, SCAD, LIRA, etc.), although they provide more accurate modeling of nonlinear behavior of structures [11], require significant computing resources, labor-intensive model calibration, and are insufficiently effective for practical use in design work [12; 13].

These limitations of the traditional approaches have created the conditions for the introduction of innovative analysis technologies, in particular artificial neural networks and other artificial intelligence methods capable of adequately accounting for complex nonlinear interactions in structures.

An artificial neural network for calculating the load-bearing capacity of square CFST columns under axial compression was first used in a study by H. Gao [14], where a three-layer feedforward network trained on experimental data showed high prediction accuracy when validated on an independent sample. This study confirmed the fundamental possibility of using neural network models as an effective auxiliary tool for engineering calculations.

Subsequent studies [15–19] significantly expanded the scope of application of artificial neural networks for CFST column analysis. These papers present improved neural network architectures trained on both experimental data and numerical simulation results.

Alongside the development of neural network approaches, significant progress has been made in the field of gradient boosting algorithms, which demonstrate comparable accuracy with greater computational efficiency and interpretability of results.

In [20], a comparative study was conducted on the accuracy of predicting the load-bearing capacity of reinforced concrete columns using five machine learning algorithms: AdaBoost, GBR, XGBoost, LightGBM, and CatBoost. The analysis used a database of experimental data, including, in particular, 401 tests of rectangular eccentrically compressed columns, containing the following parameters: geometric characteristics (cross-sectional dimensions, wall thickness, element length), physical and mechanical properties of materials, eccentricity values, and ultimate load-bearing capacity values. Statistical analysis confirmed the representativeness of the data for a wide range of parameters. This database was previously validated by Thai et al. [21] for the evaluation of normative methods (AISC 360, Eurocode 4, AS/NZS 2327[4]), which attests to its reliability.

A comparative analysis showed that CatBoost provides the highest prediction accuracy for eccentrically loaded rectangular columns, while LightGBM demonstrates slightly lower accuracy, but outperforms it in training speed by 1.5–2 times. Although the LightGBM algorithm is characterized by high training speed when working with large amounts of data due to the optimization of the tree construction process and supports distributed computing, CatBoost was preferred in this study. The key factor in the choice was the

[1] EN 1994-1-1 (2004) (English): Eurocode 4: Design of composite steel and concrete structures. Available from: https://www.phd.eng.br/wp-content/uploads/2015/12/en.1994.1.1.2004.pdf (accessed: 13.06.2025).

[2] ANSI/AISC 360-22. An American National Standard Specification for Structural Steel Buildings. Chicago: American Institute of Steel Construction; 2022. Available from: https://www.aisc.org/publications/steel-standards/aisc-360/ (accessed: 13.06.2025).

[3] SP 266.1325800.2016 Composite steel and concrete structures. Design rules. Moscow: Stroitel’stvo Publ.; 2017.

[4] AS/NZS 2327:2017. Composite structures — Composite steel-concrete construction in buildings. AS/NZS 2327:2017. Composite structures — Composite steel-concrete construction in buildings. Available from: https://www.standards.org.au/standards-catalogue/standard-details?designation=as-nzs-2327-2017 (accessed: 13.06.2025).

need to simultaneously predict two interrelated geometric parameters: the height and width of the cross-section of CFST columns. Unlike LightGBM, CatBoost has built-in mechanisms for multidimensional regression, which eliminates the need to create complex user-defined loss functions. An important advantage of CatBoost is its resistance to overfitting, provided by a combination of ordered boosting and automatic L2 regularization, which is especially important when working with synthetic data. Thus, despite the absence of categorical features in the analyzed data and the potential advantages of LightGBM in processing speed, the choice of CatBoost proved to be methodologically reasonable, as it best suits the specifics of the problem at hand, given that for the dataset under consideration, the gain in processing speed is not a determining factor, while the advantages of CatBoost in terms of accuracy and prediction stability become of primary importance.

It is also worth noting that despite the significant number of studies devoted to CFST columns, most of them focus on verification calculations in determining the load-bearing capacity [22; 23], while issues related to determining the required cross-sectional dimensions remain insufficiently studied. Moreover, while a few studies do address design tasks, they are mainly limited to cases of central compression of round columns, while the issues of optimal design of rectangular CFST columns under eccentric loading are practically not considered, despite the importance of such calculations.

In this regard, it seems particularly relevant to develop new approaches to solving these problems for rectangular cross-sections under eccentric compression. This article is a continuation of the research [24–28] aimed at obtaining the most convenient and reliable tools for performing verification and design calculations for CFST columns.

The model proposed in this article can be used for automated design of CFST columns, reducing calculation time and increasing the economic efficiency of structural solutions. The results of the study are of interest to structural designers involved in the development and implementation of CFST columns, as well as for further scientific research in the field of optimization design of these structures.

2. Methods

Short columns are considered, for which deformations are small and do not lead to a significant change in the eccentricity of the longitudinal force. The following values were taken as input parameters for the machine learning model:

1) Concrete compressive strength class B, MPa according to the Interstate Standard in force in Russia GOST 18105-2018[5];

2) Compressive force magnitude F, kN;

3) Tube wall thickness t, mm;

4) Compressive force eccentricity e, mm;

The output parameters of the model are width b, mm and height h, mm of the cross-section.

Synthetic data was generated to train the model. The concrete class ranged from B15 to B80. The compressive load varied from 500 to 10000 kN in increments of 100 kN, the wall thickness varied from 3 mm to 22 mm in increments of 1 mm, and the eccentricity of the longitudinal force varied from 10 to 250 mm in increments of 10 mm. The selection of these concrete class ranges was determined by the possibility of purchasing concrete mixes with the declared strength from mass producers (ultra-high-strength concrete and low-grade concrete were not considered). The selected wall thickness range is determined by the rolled steel sections available in the Russian Federation. The minimum eccentricity value of 10 mm in the training set corresponds to the minimum value of random eccentricity according to SP 63.13330.2018[6]. The upper eccentricity limit of 250 mm is justified by the fact that at large eccentricities of

[5] GOST 18105-2018. Concrete. Rules for control and strength assessment. A.A. Gvozdev Concrete and Reinforced Concrete Research Institute. Moscow: Standardinform Publ.; 2019.

[6] SP 63.13330.2018. Concrete and reinforced concrete structures. General provisions. Available from: https://docs.cntd.ru/document/554403082 (accessed: 13.06.2025).

the longitudinal force, when the structure resists mainly bending rather than compression, the use of CFST structures becomes irrational. For each set of values $[B, F, t, e]$ the problem of determining the optimal values of dimensions b and h was solved under minimum cost of the structure taking into account the requirement to satisfy the strength condition.

Since most rectangular tube sections manufactured for the Russian market are made from grade 09G2S steel, yield strength R_y of the tube material is assumed to be constant. For this grade of steel, it averages 345 MPa.

The cost of 1 linear meter of the structure S is determined by the formula:

$$S = S_b bh + S_s \rho_s 2(b+h)t \to \min, \tag{1}$$

where S_b is the cost of 1 m[3] of concrete; S_s is the cost of 1 tonne of steel; $\rho_s = 7.85 \text{t/m}^3$ is the density of steel.

The value of S_s was assumed to be 86800 rubles per tonne as an average value for tube sections of grade 09G2S steel according to the data of Metallotorg JSC[7] as on 27.03.2025 for the city of Rostov-on-Don, Russian Federation. The value of S_b depends on the concrete class. Table 1 contains the values of S_b for the considered concrete classes in this study according to the data of Astragal company[8]. This table also shows the values of the design strength of concrete depending on its class based on the Russian code SP 63.13330.2018[9].

Table 1. **The cost of 1 m³ of concrete depending on its compressive strength class, as well as the design compressive strength**

Concrete class B	The cost of 1 m³ of concrete, S_b, rub	Design compressive strength R_b, MPa
B15	5300	8.5
B20	5700	11.5
B25	6250	14.5
B30	6650	17
B35	6900	19.5
B40	7700	22
B45	8700	25
B50	9200	27.5
B55	10200	30
B60	10800	33
B70	12700	37
B80	13400	41

Source: made by S.H. Al-Zgul.

For designing the cross-section, it is necessary to write the strength condition of the CFST element under eccentric compression. This article considers the case when bending moment acts only in one plane. The effect of strength increase due to the resistance of concrete in confined conditions is not taken into account. The wall thickness of the tube is assumed to be small compared to the cross-sectional dimensions b and h. The external dimensions of the column are considered approximately equal to the dimensions of the concrete core.

---

[7] Metallotorg. Available from: https://www.metallotorg.ru/ (accessed: 13.06.2025).

[8] Astragal. Available from: https://www.astragal.su/ (accessed: 13.06.2025).

[9] SP 63.13330.2018. Concrete and reinforced concrete structures. General provisions. Available from: https://docs.cntd.ru/document/554403082 (accessed: 13.06.2025).

The limit equilibrium method is used to determine the ultimate load under eccentric compression. Previously, this method for CFST columns was validated using experimental data in [29]. The stress in the compressed zone of the concrete core is assumed to be equal to R_b. The stress in the tension zone of the concrete core is not taken into account. The stresses in the compression and tension zones of the steel tube are assumed to be equal to R_y with the corresponding signs (Figure 1).

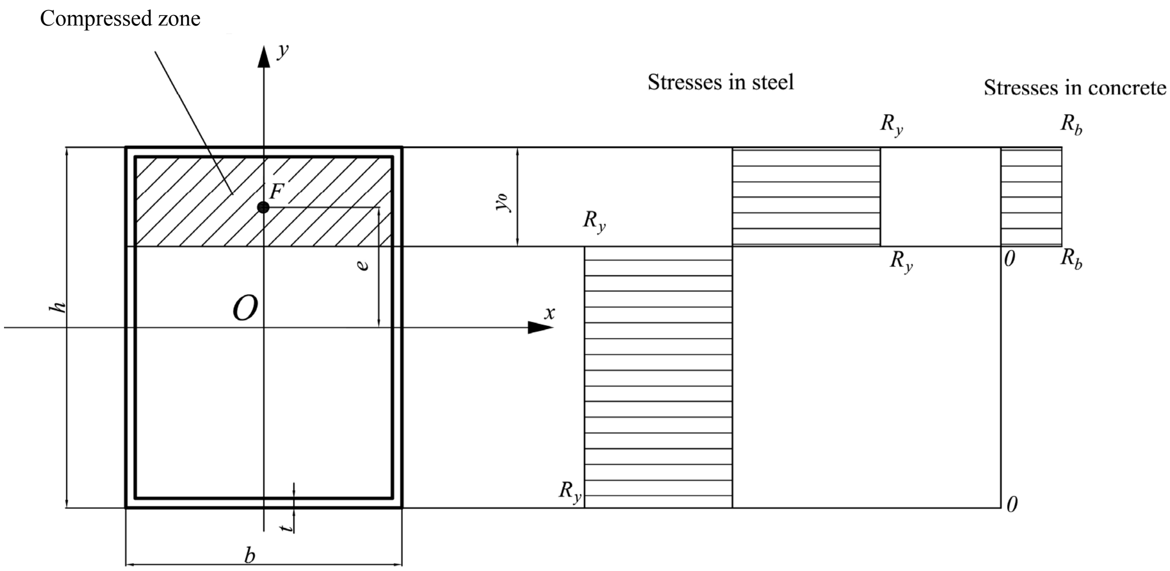

Figure 1. Diagram for determining the maximum load
S o u r c e: made by S.H. Al-Zgul.

The equation of equilibrium between internal and external forces projected onto the longitudinal axis of the column is as follows:

$$F = R_b b y_0 + 2R_y t y_0 + R_y tb - R_y tb - 2R_y t\left(h - y_0\right). \tag{2}$$

From (2), parameter y_0 is expressed as:

$$y_0 = \frac{F + 2R_y th}{R_b b + 4R_y t}. \tag{3}$$

The equation of equilibrium between internal and external moments with respect to the x-axis in the limit state is written as:

$$\begin{aligned} Fe = R_b b y_0\left(\frac{y_0}{2} - \left(y_0 - \frac{h}{2}\right)\right) + 2R_y tb\frac{h}{2} + 2R_y t y_0\left(\frac{y_0}{2} - \left(y_0 - \frac{h}{2}\right)\right) + \\ + 2R_y t\left(h - y_0\right)\left(y_0 - \frac{h}{2} + \frac{h - y_0}{2}\right). \end{aligned} \tag{4}$$

Or, after simplification:

$$Fe = \frac{1}{2}R_b b y_0\left(h - y_0\right) + R_y t\left(bh + 2y_0\left(h - y_0\right)\right). \tag{5}$$

Substituting (3) into (5) yields a quadratic equation with respect to F, from which ultimate load F_{ult} is determined as the minimum positive root of this equation satisfying the condition:

$$y_0 = \frac{F_{ult} + 2R_y th}{R_b b + 4R_y t} \le h. \tag{6}$$

If none of the positive roots of the quadratic equation satisfy condition (6), this indicates that the compressed zone covers the entire concrete core and the neutral line runs either at the boundary between the concrete core and the lower side of the tube (Figure 2) or directly at the lower side of the tube. These cases are possible when the longitudinal force has a small eccentricity.

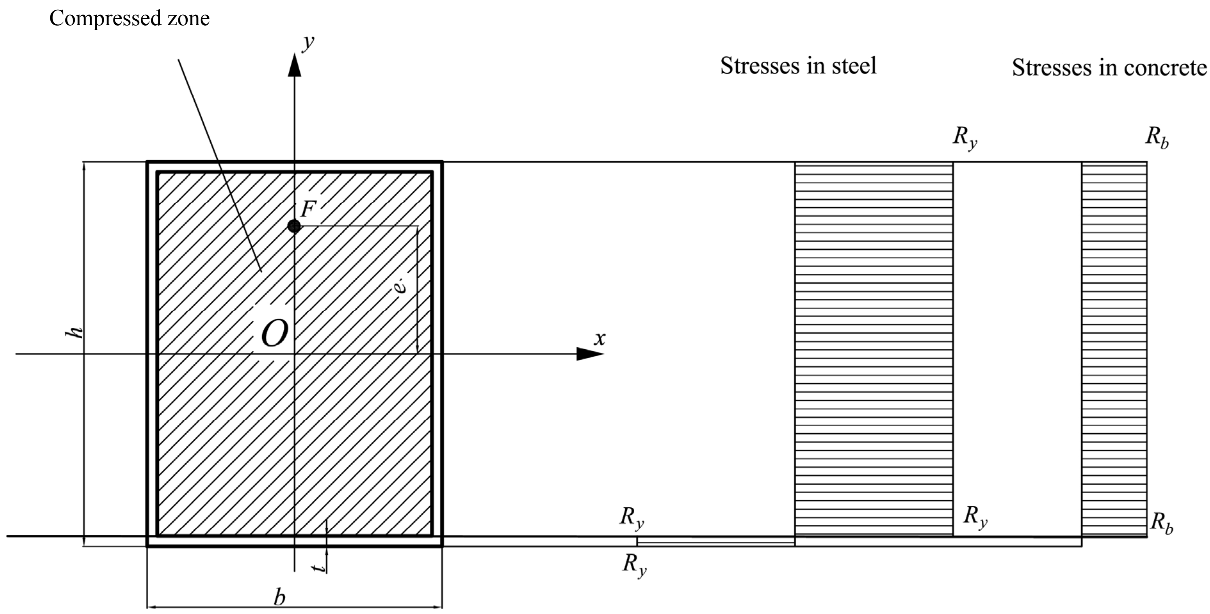

Figure 2. The case when the compressed zone covers the entire concrete core
S o u r c e: made by S.H. Al-Zgul.

For the case shown in Figure 2, the ultimate load is determined by the formula:

$$F_{ult2} = R_b bh + 2R_y ht. \tag{7}$$

If the neutral line is inside the lower side of the pipe, then F_{ult} will take an intermediate value between F_{ult2} and F_{ult0}, corresponding to the case of central compression:

$$F_{ult0} = R_b bh + 2R_y \left(h + b\right)t. \tag{8}$$

If when calculating y_0 it was obtained that $y_0 > h$, the ultimate load was determined using formula (7), which was included in the factor of safety.

The problem of finding the values of b and h from the minimum condition of cost function S limited by $F \le F_{ult}$ was solved as a nonlinear optimization problem using the interior point method [30] in the MATLAB R2024b environment (Optimization Toolbox package, fmincon function). Besides the limitation of $F \le F_{ult}$, the minimum and maximum values of dimensions b and h were also constrained. These constraints are dictated by the possibility of laying concrete inside the tube (limit on minimum dimensions) and the current range of rectangular tube sections (limit on maximum dimensions):

$$\begin{aligned}&100 \text{ mm} \le b \le 500 \text{ mm};\\&100 \text{ mm} \le h \le 500 \text{ mm}.\end{aligned} \tag{9}$$

Only the data for which a solution to the optimization problem could be found under the specified constraints were included in the training dataset. The total volume of the training data set ultimately amounted to 504841 rows. A fragment of the training dataset is shown in Table 2.

Table 2. **A fragment of the training dataset**

No.	*B*, MPa	*F*, kN	*t*, mm	*e*, mm	*b*, mm	*h*, mm
1	15	500	3	10	100	171
2	15	500	3	20	100	171
3	15	500	3	30	100	171
4	15	500	3	40	100	178
5	15	500	3	50	100	191
6	15	500	3	60	100	204
7	15	500	3	70	100	217
8	15	500	3	80	100	229
9	15	500	3	90	100	240
10	15	500	3	100	100	252
…	…	…	…	…	…	…
504832	80	10000	22	160	259	500
504833	80	10000	22	170	272	500
504834	80	10000	22	180	285	500
504835	80	10000	22	190	299	500
504836	80	10000	22	200	312	500
504837	80	10000	22	210	326	500
504838	80	10000	22	220	341	500
504839	80	10000	22	230	355	500
504840	80	10000	22	240	369	500
504841	80	10000	22	250	384	500

S o u r c e: made by S.H. Al-Zgul.

In this study, a program was developed in Python language with support of version 3.11+, implementing the gradient boosting algorithm (Catboost) for predicting two physically interrelated geometric characteristics (width and height) of CFST column cross-sections based on four initial characteristics (design compressive strength of concrete, magnitude of compressive force, tube wall thickness, eccentricity of compressive force). Since the presented model has a small number of parameters, it can be trained and used on most modern (as of 2025) consumer GPUs or even in cloud environments with free quotas (e.g., Google Colab, Kaggle). When setting the regression problem, the MultiRMSE (Composite RMSE) metric was used — a generalized case of the standard RMSE for simultaneous optimization of the loss function for two target variables — the width (b) and height (h) of the cross-section (10):

$$\text{MultiRMSE} = \sqrt{\frac{1}{2N}\sum\nolimits_{i=1}^{N}\left[\left(b_i - \hat{b}_i\right)^2 + \left(h_i - \hat{h}_i\right)^2\right]}, \tag{10}$$

where N is the number of records (columns) in the sample; 2 is the number of target variables; b_i is the true width of the i-th column section, mm; h_i is the true height of the i-th column section, mm; $\hat{b}_i$ is the predicted width of the i-th column section, mm; $\hat{h}_i$ is the predicted height of the i-th column section, mm.

As part of the study, the initial dataset was divided into three independent subsamples. The initial split was carried out in a 80% to 20% ratio, with 20% of the data being set aside as an isolated test sample for the final evaluation of the model quality. The remaining 80% of the data, constituting the training pool, was further divided: 25% was allocated for the validation sample (corresponding to 20% of the total data volume), and the remaining 75% (60% of the total volume) formed the final training sample.

The following control metrics were selected: the composite MultiRMSE metric on the validation dataset to evaluate the overall predictive power of the model, as well as individual metrics for each variable: MAE to evaluate the average accuracy of the model (in understandable, interpretable units), RMSE — allows detecting the influence of outliers when selecting the final model, and the coefficient of determination — R^2 demonstrates the advantage of the current model compared to the average prediction.

The split was performed by fixing the random seed parameter using the train_test_split function from the scikit-learn library. This makes the procedure completely deterministic and reproducible.

To verify the quality of the split, a statistical test for the homogeneity of distributions (Kruskal — Wallis for categorical/continuous features) between samples was performed based on key features. The test results (p_value > 0.05) did not reveal any statistically significant differences, which confirms the correctness of the random split.

3. Results and Discussion

The model training graph (Figure 3) demonstrates virtually identical behavior of the training and validation curves. At the initial stage (before ~500 iterations), there is a sharp decrease in RMSE from ~1.2 to ~0.1, after which the process enters a phase of smooth optimization, reaching a stable plateau after ~2000 iterations.

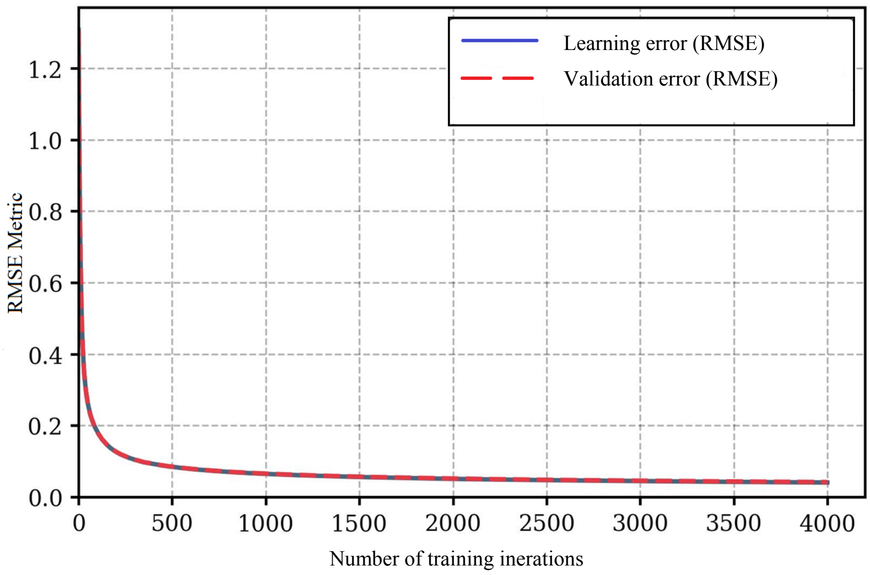

Figure 3. Graph of the learning curve for training and validation samples

S o u r c e: made by S.H. Al-Zgul.

The analysis of the model accuracy (Figures 4, 5) showed that the coefficient of determination reaches R^2 = 0.999033 for the cross-section width (b) and R^2 = 0.999211 the cross-section height (h), which indicates that the predicted values correspond almost perfectly to the target values. It is worth to note the uniform distribution of errors across the entire range of the studied dimensions (100–500 mm). This fact, along with the small dispersion, confirms the high reliability of the algorithm. It should be emphasized that the difference in prediction accuracy between the parameters does not exceed 2%, which indicates the balance of the predictive power of the model for the two target variables. Since the application of the Catboost algorithm is essentially a solution to the problem of multidimensional nonlinear interpolation, the limits of the model application are determined by the range of input parameters in the training dataset.

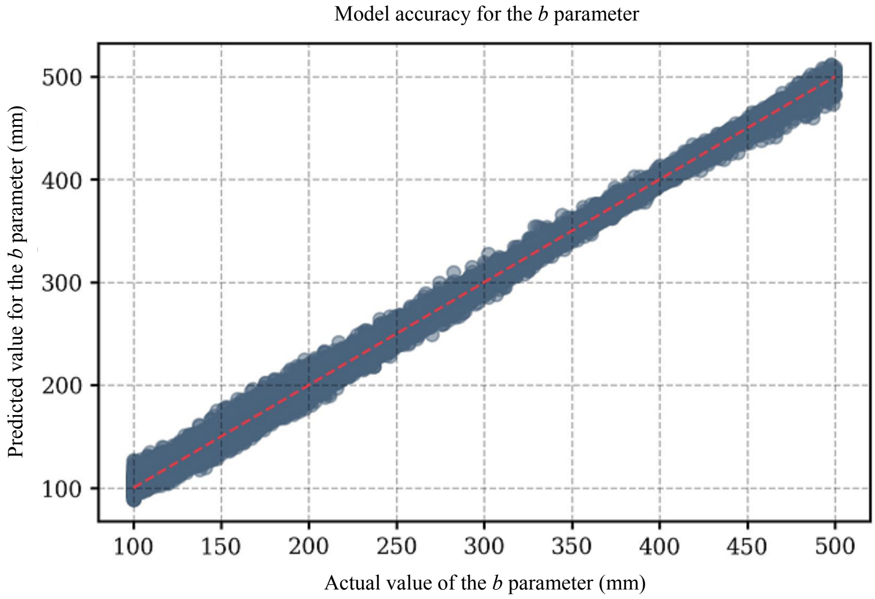

Figure 4. Accuracy of prediction of section width (*b*)
S o u r c e: made by A.S. Chepurnenko, S.H. Al-Zgul.

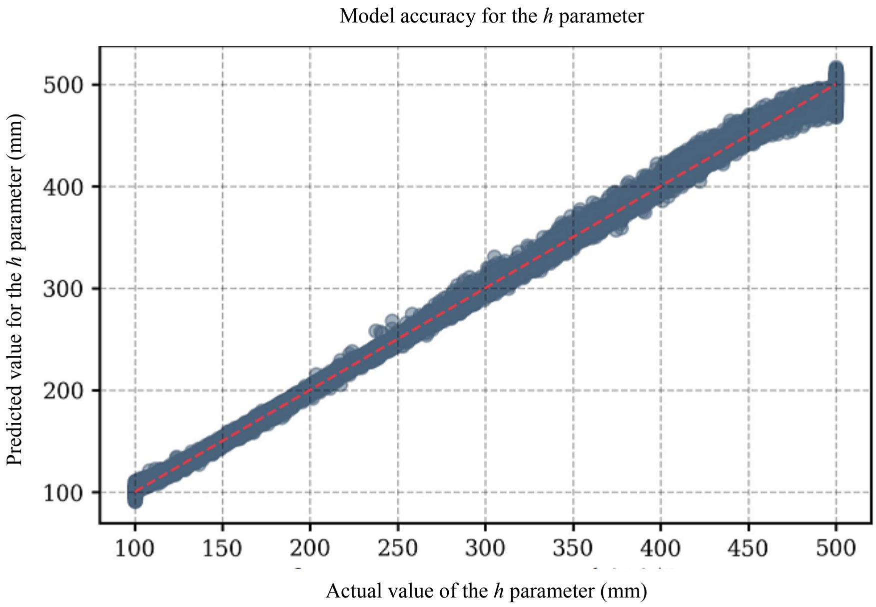

Figure 5. Accuracy of prediction of section height (*h*)
S o u r c e: made by A.S. Chepurnenko, S.H. Al-Zgul.

To evaluate the contribution of individual features to the final result, it was decided to use SHAP analysis. This method, based on the principles of game theory, allows to accurately determine the independent contribution of each feature, unlike the standard feature importances, which can give distorted results due to the possible uneven distribution of Gain when splitting initially balanced subsamples.

SHAP-value calculation requires a trained model and a test dataset. SHAP values were calculated for the entire test sample. TreeExplainer was used, which employs the TreeSHAP algorithm for efficient and accurate calculation. As the background distribution, which is required to determine the basic mathematical expected value E[f(x)] of the model, 100 randomly selected objects from the training sample were used (100 objects is the optimal compromise between accuracy and computational efficiency). Feature importance was estimated as the mean absolute SHAP-value (mean |SHAP|) across all objects in the test sample.

The results of the feature importance analysis using SHAP-values demonstrate a clear, physically justified hierarchy of the influence of input parameters (or, more accurately, features) on the geometric characteristics of the cross-section (Figures 6, 7). At the same time, the SHAP dependence graphs demonstrate the nature of the influence of individual variables on the target value (Figures 8, 9).

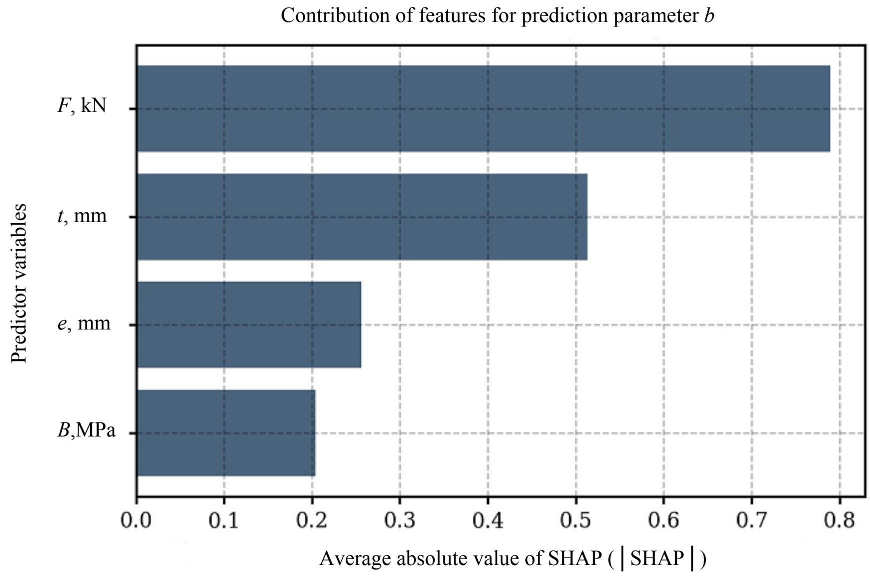

Figure 6. Analysis of the importance of features for predicting the section width (*b*)

S o u r c e: made by A.S. Chepurnenko, S.H. Al-Zgul.

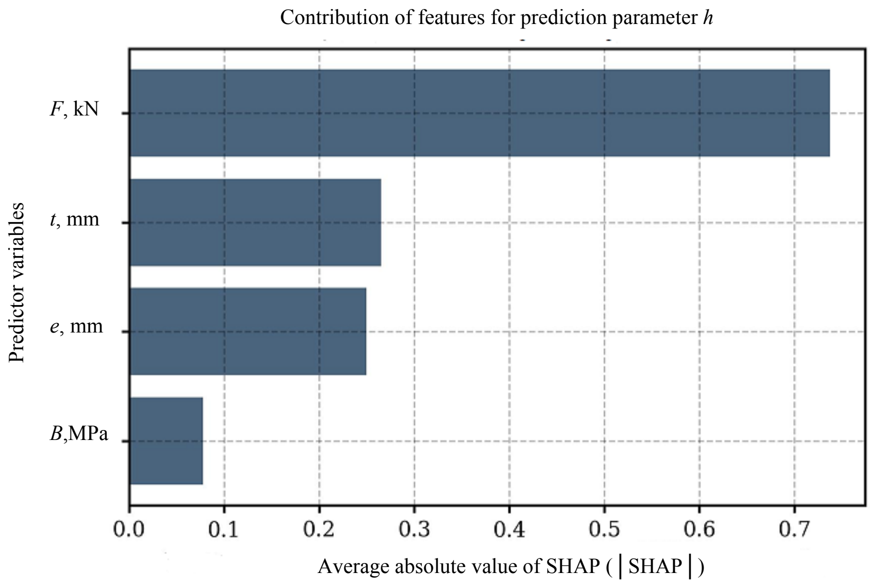

Figure 7. Analysis of the importance of features for predicting the section height (*h*)

S o u r c e: made by A.S. Chepurnenko, S.H. Al-Zgul.

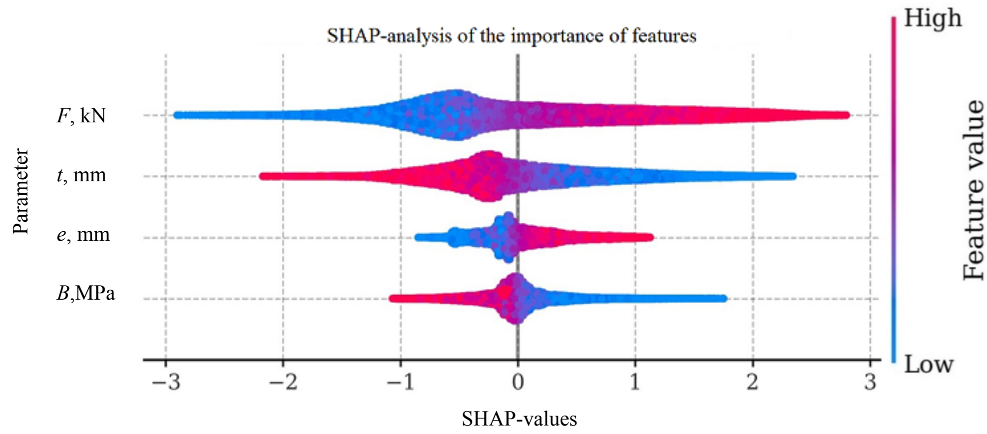

Figure 8. SHAP-analysis of the influence of factors on the width of the section
S o u r c e: made by A.S. Chepurnenko, S.H. Al-Zgul.

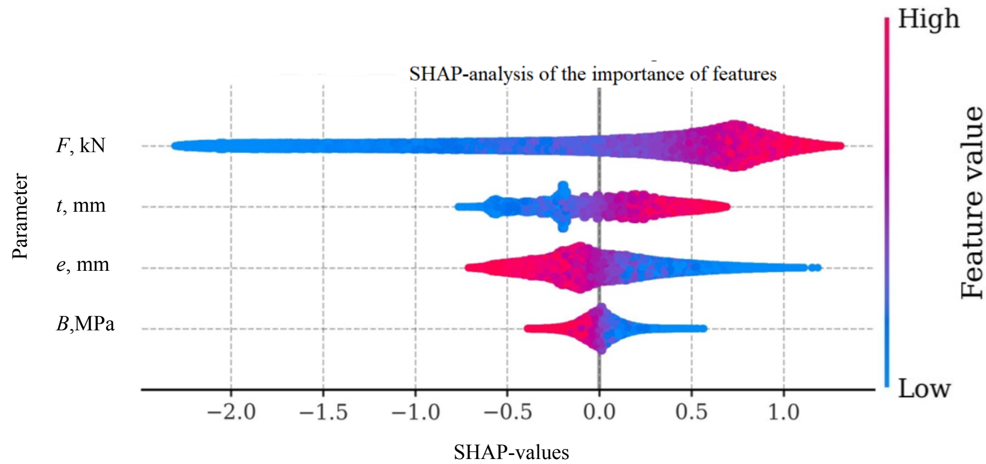

Figure 9. SHAP-analysis of the influence of factors on the height of the section
S o u r c e: made by A.S. Chepurnenko, S.H. Al-Zgul.

Data analysis revealed the dominant influence of the longitudinal force (F) on the geometric parameters of the cross-section, which is confirmed by the maximum values of SHAP indicators (0.79 for the width and 0.75 for the height). This pattern is fully consistent with the fundamental principles of calculating the load-bearing capacity of compressed elements. It is worth noting that the character of the influence of the load differs for the parameters under consideration: the section width demonstrates a more pronounced dependence with the SHAP-value in the range ≈ ±3 compared to the height in the range ≈ ±2, which is explained by the aspects of stress distribution in a rectangular section under eccentric compression. The results show a positive dependence between the increase in load and the increase in the geometric parameters of the cross-section (width b and height h), which is confirmed by the data in Figures 8 and 9.

The analysis reveals a significant difference in the degree of influence of the eccentricity (e) on the section parameters. Although eccentricity is significant for both dimensions, its effect on height is more pronounced (2nd in importance) than on width (3rd in importance). This is explained by the quadratic dependence of the moment resistance on the height of the cross-section, which makes an increase in h more effective in counteracting eccentric loads (Figures 6, 7). Graphical visualization confirms the need for |a proportional increase of both dimensions as eccentricity increases, which corresponds to the mechanics of the structure resistance to the combined action of bending moments and longitudinal forces.

The study using SHAP-value revealed a greater influence of wall thickness (t) on the cross-sectional width (SHAP~0.51) compared to the height (SHAP~0.25) (Figures 6, 7). According to the results, an increase in wall thickness leads to a decrease in both geometric parameters (b and h), which is expected, since an increase in the wall thickness increases the proportion of forces absorbed by the steel shell.

Concrete class (B) demonstrates a statistically significant but relatively weak influence on the geometric parameters of the cross-section, which is particularly pronounced for the height (Figures 6, 7). This behavior is fully consistent with the mechanics of composite CFST structures, where the concrete core is mainly involved in the resistance at the ultimate stages of loading. The analysis revealed a clear inverse relationship between the concrete class and the cross-sectional dimensions: as the strength characteristics of the concrete increased, there was a regular decrease in both the width and height of the cross-section (Figures 8, 9).

The results of the analysis reveal significant dependencies between the structural parameters (wall thickness), mechanical properties of materials (concrete class, steel strength), and loading conditions (longitudinal force, eccentricity) on one hand, and the geometric characteristics of CFST columns on the other.

4. Conclusion

The following scientific results were obtained based on the findings of the study:

1. A regression model based on the CatBoost algorithm has been developed and validated for predicting the optimal geometric parameters (width b and height h) of rectangular concrete filled steel tube columns under eccentric compression. The model demonstrates high prediction accuracy with a total composite metric MultiRMSE ≈ 3.6 mm (on the validation dataset), as well as MAE = 2.503 mm for width b, MAE = 2.467 mm for height h (on the test dataset) with a coefficient of determination $R^2 > 0.999$ for both target parameters. The design values of the height and width of the cross-section obtained using the algorithm are for reference only. The final selection of the geometric parameters must comply with the requirements of GOST, SP, or other regulatory documents and is agreed upon by the structural designer based on detailed calculations and a technical and economic justification.
2. A multi-criteria optimization method has been developed that takes into account strength conditions, design constraints, and economic efficiency by minimizing the cost function S.
3. The interior point method has proven its effectiveness in solving this nonlinear problem, providing an optimal balance between load-bearing capacity, overall dimensions, and economic parameters of the structure, as well as allowing design requirements and technological constraints to be taken into account at the stage of forming the training sample.
4. SHAP-analysis allowed to establish a clear hierarchy of importance of the input parameters, where the longitudinal force demonstrates dominant influence (SHAP-values in the range of 0.75-0.79), which quantitatively confirms its key role in determining the dimensions of the structure. At the same time, a differentiated influence of eccentricity on the geometric parameters was revealed — its influence on section height h was more pronounced compared to width b. The obtained distribution of feature importance fully corresponds to the fundamental theoretical principles of structural mechanics, which confirms the physical validity of the developed model.

The developed approach ensures optimal design of rectangular concrete filled steel tube structures under eccentric compression while meeting reliability and cost-effectiveness requirements within the range of input parameters on which the model was trained.

The obtained results form the scientific and methodological basis for the automation of concrete filled steel tube structure design processes, opening up promising areas for further research, including: expanding the model to account for long-term loads and complex loading conditions, integration with BIM technologies, and the development of regulatory recommendations for the implementation of machine learning methods in design practice. These areas will significantly improve the efficiency of structural design while ensuring the required reliability and cost-effectiveness.

References

1. Rimshin V.I., Semenova M.N., Shubin I.L., Krishan A.L., Astafieva M.A. Studies of the bearing capacity of non-centrally compressed steel-tube concrete columns. *Construction Materials*. 2022;(6):8–14. (In Russ.) https://doi.org/10.31659/0585-430X-2022-803-6-8-14 EDN: YHDXCL

2. Pan Y.C., Wang G.H., Xiang K. Overview of research progress for concrete-filled steel tubular columns after exposure to fire. *Applied Mechanics and Materials.* 2014;638–640:197–201. https://doi.org/10.4028/www.scientific.net/AMM.638-640.197

3. Krishan A.L., Krishan M.A., Sabirov R.R. Perspectives to apply concrete filled steel tube columns in construction projects of Russia. *Vestnik of Nosov Magnitogorsk State Technical University.* 2014;1(45):137–140. (In Russ.) EDN: RZPTUD

4. Hirakawa K. Saburi K., Kushima S., Kojima K. Performance-based design of 300 m vertical city "Abeno Harukas." *International Journal of High-Rise Buildings*. 2014;3(1):35–48. Available from: https://civil808.com/sites/default/files/performance_based_design_of_300.pdf (accessed: 10.03.2025).

5. Knowles R.B., Park R. Strength of concrete filled steel tubular columns. *Journal of the Structural Division*. 1969;95:2565–2588. Available from: https://trid.trb.org/view.aspx?id=105608 (accessed: 12.06.2025).

6. Sakino K., Nakahara H., Morino S., & Nishiyama I. Behavior of centrally loaded concrete-filled steel-tube short columns. *Journal of Structural Engineering*. 2004;130(2):180–188. https://doi.org/10.1061/(ASCE)0733-9445(2004)130:2(180

7. Tran V.-L., Thai D.K. A new empirical formula for prediction of the axial compression capacity of CCFT columns. *Steel and Composite Structures*. 2019;33(2):181–194. https://doi.org/10.12989/scs.2019.33.2.181

8. Tao Z., Uy B., Han L.-H., He S.-H. Design of Concrete-Filled Steel Tubular Members According to the Australian Standard AS 5100 Model and Calibration. *Australian Journal of Structural Engineering.* 2008;8(3):197–214. https://doi.org/10.1080/13287982.2008.11464998

9. Goode C.D. Composite columns-1819 tests on concrete-filled steel tube columns compared with Eurocode 4. *The Structural Engineer.* 2008; 86(16):33–38.

10. Gao P., Zhou X., Liu J., Lin X., Wang X., Chen Y.F. Experimental Assessment on the Size Effects of Square Concrete-Filled Steel Tubular Columns under Axial Compression. Engineering Structures. 2023;281:115706. https://doi.org/10.1016/j.engstruct.2023.115706

11. Chepurnenko A., Yazyev B., Khashkhozhev K., Chepurnenko V., Meskhi B., Beskopylny A. Simplified 2D finite element model for calculation of the bearing capacity of eccentrically compressed concrete-filled steel tubular columns. *Applied Sciences (Switzerland).* 2021;11(24):11645. https://doi.org/10.3390/app112411645 EDN: AKSSMM

12. Lu D., Chen Zh., Ding F., Chen Zh., Sun P. Prediction of mechanical properties of the stirrup-confined rectangular cfst stub columns using fem and machine learning. *Mathematics.* 2021;9(14):1643. https://doi.org/10.3390/math9141643 EDN: MXXJFT

13. Mustafa R., Ahmad M.T. Appraisal of numerous machine learning techniques for the prediction of axial load carrying capacity of rectangular concrete column. *Asian J Civ Eng.* 2024;25:4471–4486. https://doi.org/10.1007/s42107-024-01060-6

14. Gao H. Calculation of load-carrying capacity of square concrete filled tube columns based on neural network. *Applied Mechanics and Materials*. 2011;71-78:847–850. https://doi.org/10.4028/www.scientific.net/AMM.71-78.847

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

16. Wang H.J., Zhu H.B., Wei H. Bearing capacity of concrete filled square steel tubular columns based on neural network. *Advanced Materials Research.* 2012;502:193–197. https://doi.org/10.4028/www.scientific.net/AMR.502.193

17. Du Y., Chen., Wang Y.-B., Liew J.Y.R. Ultimate resistance behavior of rectangular concrete-filled tubular beam-columns made of high-strength steel. *Journal of Constructional Steel Research.* 2017;133:418–433. https://doi.org/10.1016/j.jcsr.2017.02.024

18. Le T.T., Asteris P.G., Lemonis M.E. Prediction of axial load capacity of rectangular concrete-filled steel tube columns using machine learning techniques. *Engineering with Computers.* 2022;38:3283–3316. https://doi.org/10.1007/s00366-021-01461-0

19. Zarringol M., Patel V.I., Liang Q.Q. Artificial neural network model for strength predictions of CFST columns strengthened with CFRP. *Engineering Structures.* 2023;281:115784. https://doi.org/10.1016/j.engstruct.2023.115784 EDN: AEKDAT

20. Degtyarev V.V., Thai H.T. Design of concrete-filled steel tubular columns using data-driven methods. *Journal of Constructional Steel Research*. 2023;200:107653. https://doi.org/10.1016/j.jcsr.2022.107653 EDN: AUPVMJ

21. Thai H.T., Thai S., Ngo T., Uy B., Kang W.H., Hicks S.J. Reliability considerations of modern design codes for CFST columns. *Journal of Constructional Steel Research.* 2021;177:106482. https://doi.org/10.1016/j.jcsr.2020.106482 EDN: FYAXTF

22. Megahed K. Strength prediction of ECC-CES columns under eccentric compression using adaptive sampling and ML techniques. *Scientific Reports*. 2025;15:1202. https://doi.org/10.1038/s41598-024-83666-z

23. Zarringol M., Naser M.Z. Explainable machine learning model for prediction of axial capacity of strengthened CFST columns. In Book: *Interpretable Machine Learning for the Analysis, Design, Assessment, and Informed Decision Making for Civil Infrastructure.* Elsevie Publ.; 2024. P. 229–253. https://doi.org/10.1016/B978-0-12-824073-1.00016-2

24. Chepurnenko A., Yazyev B., Khashkhozhev K., Chepurnenko V., Meskhi B., Beskopylny A. Simplified 2D finite element model for calculation of the bearing capacity of eccentrically compressed concrete-filled steel tubular columns. *Applied Sciences (Switzerland).* 2021;11(24):11645. https://doi.org/10.3390/app112411645 EDN: AKSSMM

25. Chepurnenko A., Turina V., Akopyan V. Simplified method for calculating the bearing capacity of slender concrete-filled steel tubular columns. *Civil Eng*. 2023;4(3):1000–1015. https://doi.org/10.3390/civileng4030054 EDN: TDQHJD

26. Chepurnenko A., Al-Zgul S., Tyurina V. Machine learning for predicting required cross-sectional dimensions of circular concrete-filled steel tubular columns. *Buildings.* 2025;15(9):1438. https://doi.org/10.3390/buildings15091438

27. Al-Zgul S., Tyurina T., Chepurnenko A., Chepurnenko A., Akopyan V. Artificial neural network models for predicting required cross-section dimensions of concrete filled steel tubular columns. *The Open Civil Engineering Journal.* 2025;19(1). https://doi.org/10.2174/0118741495387193250411105201

28. Chepurnenko A.,Yazyev B., Al-Zgul S., Tyurina V. Simplified finite element model for rectangular CFST columns strength calculation under eccentric compression. *Magazine of Civil Engineering*. 2025;18(2):13406. https://doi.org/10.34910/MCE.134.6

29. Al-Zgul S., Tyurina V., Chepurnenko A. A simplified method for determining the bearing capacity of eccentrically compressed rectangular CFST columns with eccentricities in two planes. *The Open Construction & Building Technology Journal*. 2025;19:e18748368402845. https://doi.org/10.2174/0118748368402845250709060829

30. Chepurnenko A., Turina V., Akopyan V.F. Optimization rectangular and box sections in oblique bending and eccentric compression. *Construction Materials and Products.* 2023;6(5):1–14. https://doi.org/10.58224/2618-7183-2023-6-5-2