International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Predicting Compressive Strength of CFST Columns Using Machine Learning Algorithms

V Rashmi, B Yateeswara Rao, D Hemasundara Rao, U Charan Kumar, S Mohan

Department of Civil Engineering, GMR Institute of Technology

Abstract:

Concrete-Filled Steel Tubes (CFST) columns are extensively used in civil engineering due to their enhanced strength and stability. Accurately predicting the compressive strength of CFST columns is crucial for ensuring safe and efficient design. Traditional prediction methods often involve complex and time-consuming calculations that may lack precision. This study investigates the application of machine learning (ML) algorithms as an alternative approach for predicting the compressive strength of CFST columns. A comprehensive dataset, encompassing material properties, geometric dimensions, and loading conditions, is used to train and validate various ML models, including linear regression, decision trees, random forests, support vector machines, and neural networks. Model performance is assessed using metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2). Results demonstrate that ML models, particularly ensemble methods like random forests and advanced neural networks, offer significantly higher accuracy and reliability compared to traditional methods. The study also highlights the critical role of feature selection and hyperparameter tuning in optimizing model performance. These findings indicate that ML algorithms provide a robust and efficient alternative for predicting the compressive strength of CFST columns. Future research could expand the dataset, explore additional ML techniques, and integrate these models into practical engineering software applications.

Keywords: Compressive Strength Prediction, Concrete-Filled Steel Tubes (CFST), Machine Learning Algorithms, Structural Engineering Design.

Introduction:

Concrete-Filled Steel Tube (CFST) columns are a crucial feature in modern structural engineering. These columns are known for their exceptional performance and versatility. They utilize the strengths of both concrete and steel. The steel tube acts as formwork and provides confinement to the concrete. This confinement enhances the concrete's compressive strength. The concrete core, in turn, stabilizes the steel tube and prevents local buckling. The composite action between concrete and steel creates highly efficient structural elements. These elements offer high load-bearing capacity and excellent ductility. They also exhibit significant energy absorption capabilities. As a result, CFST columns are well-suited for high-rise buildings, bridges, and seismic applications.

The development of CFST columns began in Japan in the mid-20th century. They later gained widespread use across Europe and North America. Today, CFST columns are essential for ensuring the stability and resilience of modern infrastructure. Predicting the compressive strength of CFST columns accurately is critical. Traditional methods often struggle to capture the complex interactions between key factors. These factors include material properties, geometric parameters, and environmental conditions. As a result, traditional methods can lead to overly conservative or unsafe designs.

Machine Learning (ML) provides a promising alternative for this challenge. ML algorithms can learn from extensive data and identify intricate patterns. They use historical data from experiments and real-world projects to improve predictions. ML models can better account for nonlinear interactions between variables. This results in more accurate and efficient compressive strength estimates. The integration of ML can enhance the reliability and efficiency of structural designs. This study aims to develop and validate ML models for CFST strength prediction. It also compares these models with traditional methods and offers practical guidelines. Machine Learning models have the potential to revolutionize how CFST columns are designed. They enable engineers to make more informed decisions with greater confidence. Additionally, ML techniques can continuously improve as more data is collected over time. This makes them highly adaptable to evolving construction practices and standards. Overall, the adoption of ML in structural engineering marks a significant advancement, promoting innovative, safer, and more cost-effective construction methods.

Methodology:

The methodology for this project involves several critical steps, including data preprocessing and exploration, model training, model evaluation, and hyperparameter tuning. Each step is essential to ensure the accuracy and effectiveness of the machine learning models used for predicting compressive strength of CFST columns.

Data Preprocessing and Exploration

1. Data Collection: The dataset consists of 1305 samples, each with five predictor variables (Diameter of the steel tube, Thickness of the steel tube, Length of the CFST column, Yield strength of the steel, Compressive strength of the concrete) and one target variable (compressive strength).
2. Data Cleaning: Handling missing values, outliers, and inconsistencies in the dataset. This involves removing or imputing missing values and ensuring data integrity.
3. Feature Scaling: Normalizing the numerical features to ensure that they have a consistent scale, which is particularly important for algorithms sensitive to feature magnitude (e.g., SVM, KNN).
4. Exploratory Data Analysis (EDA): Visualizing data distributions, identifying patterns and correlations, and understanding the relationships between variables. Techniques such as histograms, box plots, scatter plots, and correlation matrices are used in this step.
5. Dimensionality Reduction: If necessary, applying techniques like Principal Component Analysis (PCA) to reduce the number of features while retaining most of the data's variance, thus simplifying the model training process.

Machine Learning Algorithms

A diverse array of machine learning algorithms will be employed in this project to capture different aspects of the data and identify the most effective model for crop prediction:

1. Linear Regression: A statistical method for predicting continuous outcomes, applied here to estimate the compressive strength of Concrete-Filled Steel Tubular (CFST) columns based on various influencing factors..
2. Support Vector Machine (SVM): A powerful regressor that finds the hyperplane which best separates the data into different classes.
3. K-Nearest Neighbours (KNN): A simple, instance-based learning algorithm that classifies samples based on the majority class among their nearest neighbours.
4. Decision Tree: A model that splits the data into branches to make predictions based on decision rules inferred from the features.
5. Random Forest: An ensemble method that builds multiple decision trees and merges their predictions for improved accuracy and robustness.
6. Gradient Boosting: An advanced ensemble technique that builds models sequentially, each correcting the errors of its predecessor.
7. Bagging : Another ensemble method that combines the predictions of multiple models to reduce variance and avoid overfitting.
8. AdaBoost: An ensemble technique that combines weak classifiers to form a strong classifier, focusing on samples that are difficult to classify.
9. Boosting: It is an ensemble technique that improves regression model performance by sequentially combining the predictions of multiple weak learners, typically decision trees, where each new model corrects the errors made by the previous ones, leading to more accurate and robust predictions.

Results & Discussion:

Exploratory Data Analysis (EDA)

Hist Plot

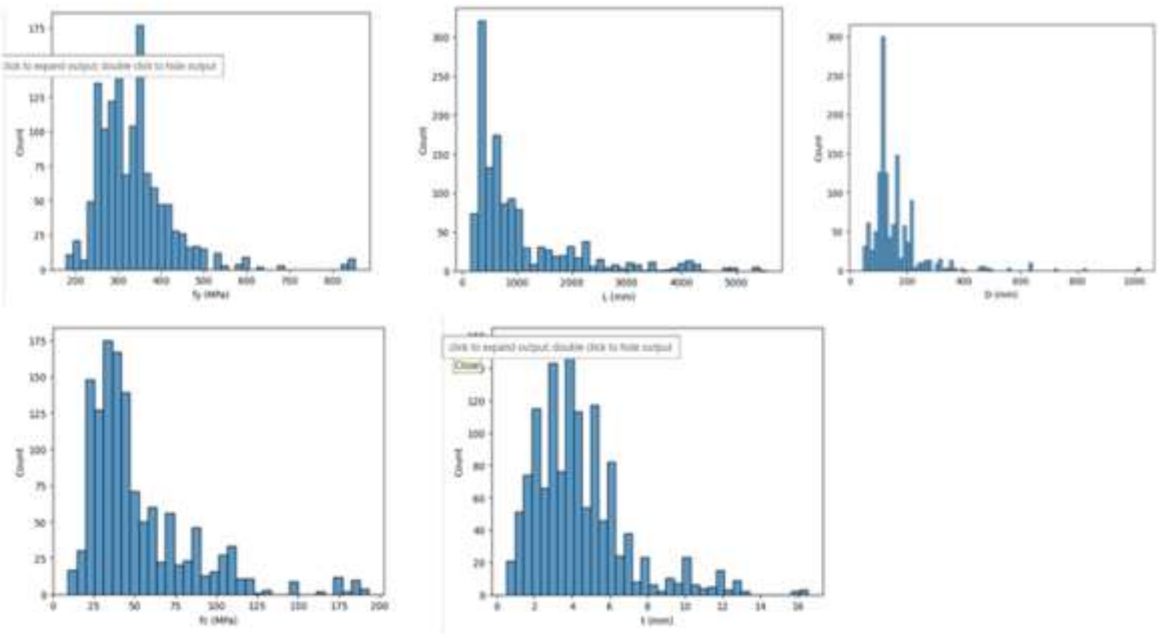


Figure 1 Hist Plot

Interpretations: The histograms for each parameter used in predicting the compressive strength of CFST columns reveal a clear concentration around specific ranges. The yield strength of steel (fy) predominantly falls between 300–400 MPa, while the concrete compressive strength (fc) is mostly around 25–50 MPa, indicating that moderate-strength materials are common. Column length (L) and outer diameter (D) both show a strong skew towards smaller sizes, with the majority of columns being under 1000 mm in length and 200 mm in diameter. Similarly, tube thickness (t) has a concentration between 2–6 mm, with thicker tubes being rare. Overall, the dataset is biased toward shorter, smaller-diameter columns with moderate material strengths and thin steel tubes, which might affect the model's performance for designs outside these typical ranges.

Boxplots

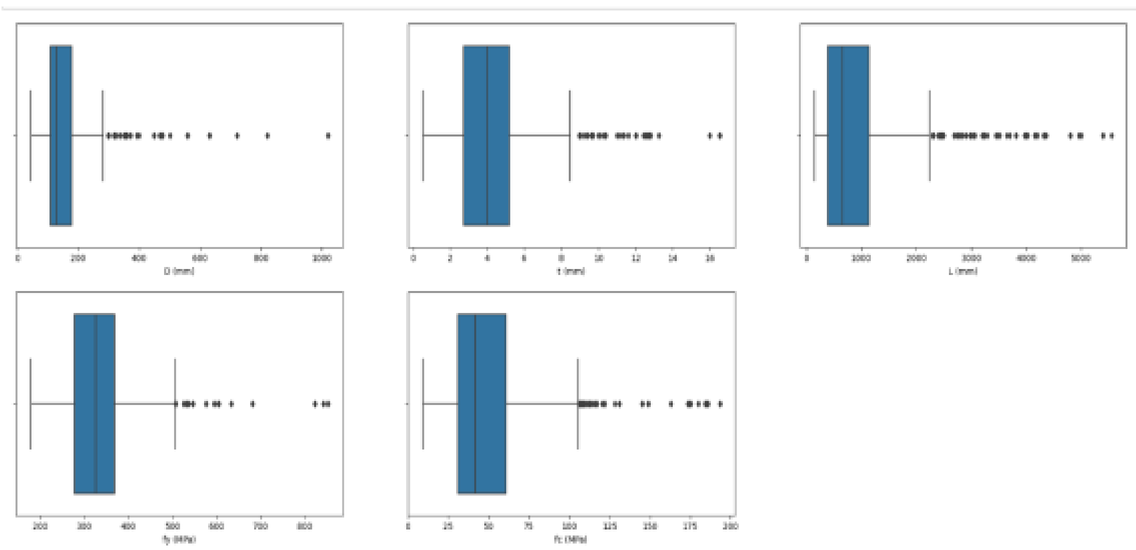


Figure 2 BoxPlots

Interpretation: The box plots reveal that all five parameters—diameter (D), thickness (t), length (L), yield strength of steel (fy), and compressive strength of concrete (fc)—have a central concentration of values within a well-defined interquartile range (IQR), but each also displays numerous outliers on the higher end. The diameter (D) and length (L) parameters show particularly wide spreads with significant outliers, indicating variability in geometric dimensions, which could influence buckling and load distribution. Thickness (t) has a narrower range but still exhibits outliers, suggesting some columns have notably thicker steel walls that may contribute to greater strength. Both material properties, yield strength of steel (fy) and compressive strength of concrete (fc), also display outliers at higher values, reflecting variability in material performance that could impact the overall strength of CFST columns. Collectively, the distribution patterns and presence of outliers across all parameters emphasize the importance of handling these variations carefully, as they could influence the accuracy and reliability of compressive strength predictions.

Heat Map

Figure 3 Heat Map

Interpretation:

The heat map reveals generally weak to moderate correlations among the parameters influencing the compressive strength of CFST columns. The diameter and thickness exhibit a moderate positive relationship (0.43), while other correlations, such as those between length and compressive strength , are weak or negligible. The yield strength is weakly correlated with all other variables, with the highest being its correlation with thickness (0.25). Overall, the heat map suggests limited linear relationships, indicating that compressive strength may not be strongly predicted by these basic geometric and material properties alone, necessitating more complex modeling approaches to capture the underlying behavior.

Pair Plot

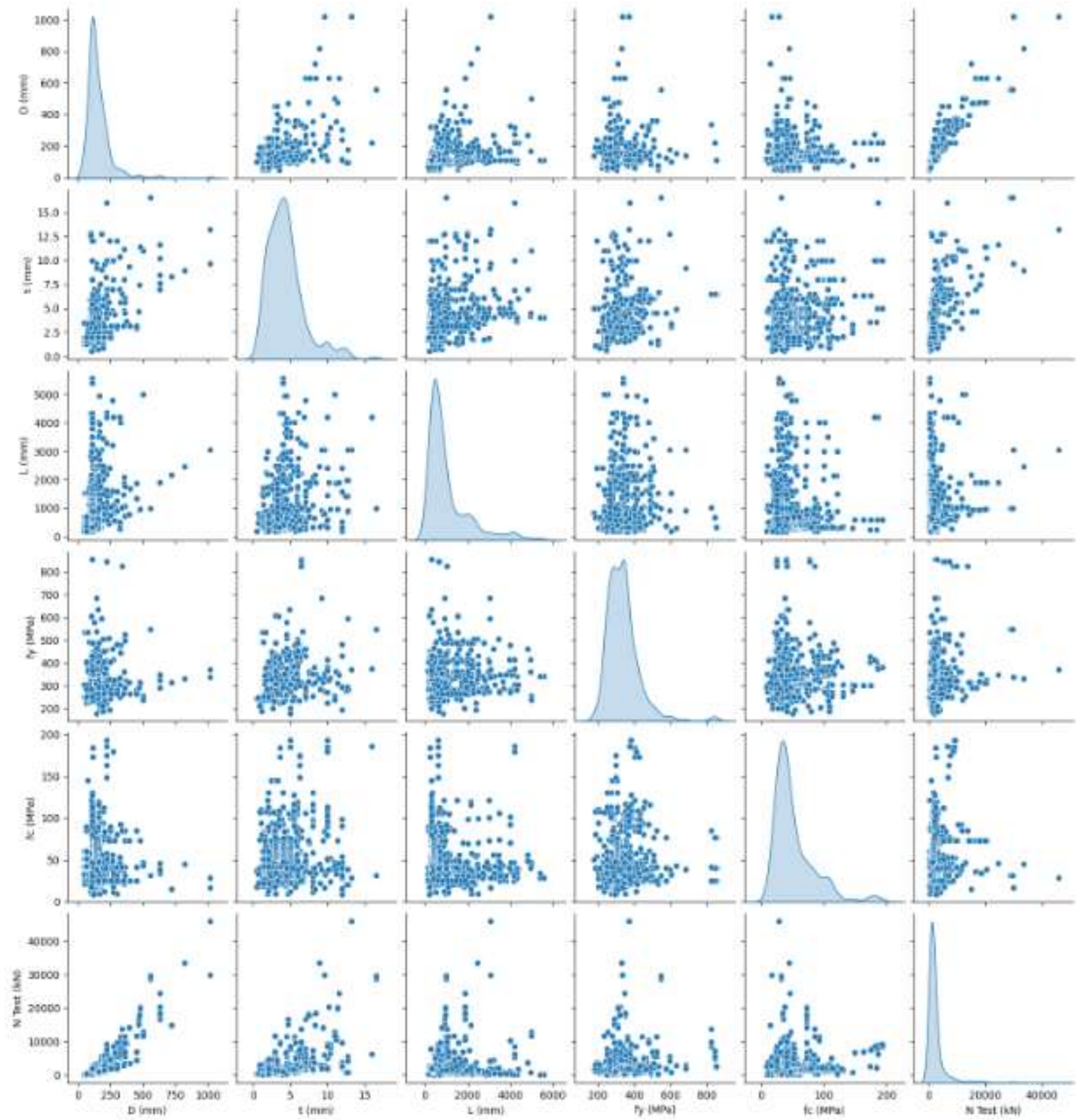
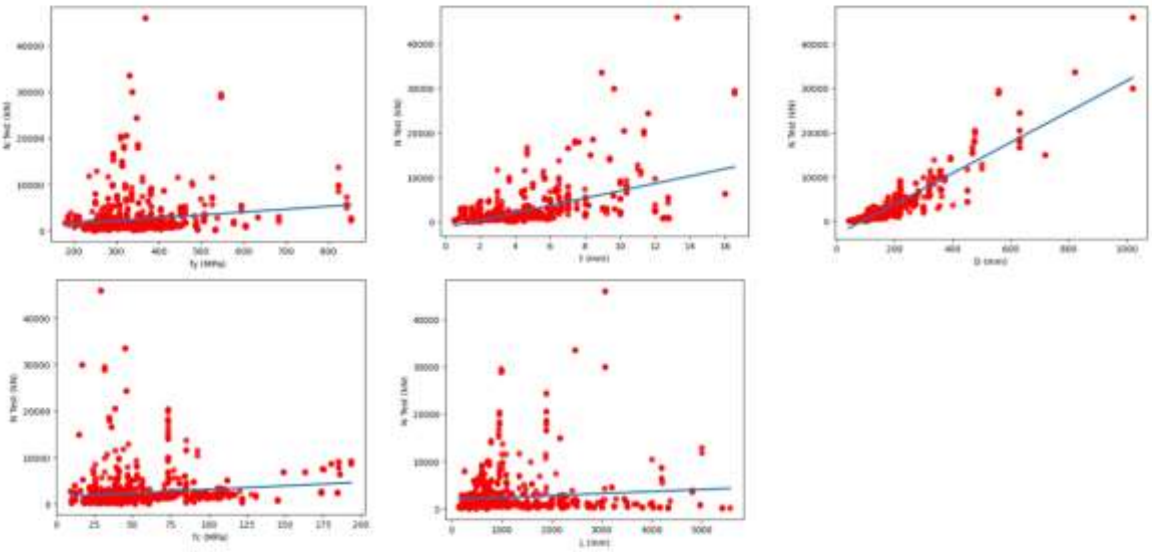


Figure 4 Pair Plot

Interpretation:

The pair plot reveals that most variables exhibit skewed distributions, with values clustered towards lower ranges, especially for diameter , length , yield strength , and compressive strength . The scatter plots generally show weak and diffuse relationships, with no strong linear trends between the variables. Notably, the relationship between compressive strength and total load suggests a potential positive correlation, indicating that higher compressive strength might be associated with higher load-bearing capacity. However, the other variable pairs, such as diameter with thickness or length with strength, show scattered patterns, implying limited linear predictability. Overall, the plot highlights the complexity of interactions among variables and suggests the need for more advanced modeling approaches to better predict compressive strength.

Residual Plots



Interpretation:

The scatter plots compare the relationship between compressive strength and five predictors: yield strength of steel , steel tube thickness , diameter , concrete compressive strength , and column length . Among these, diameter shows the strongest positive correlation with , with data points following the regression line closely, suggesting it is a key predictor. In contrast, , , , and exhibit weaker positive trends with wide residual spreads, indicating that they are less effective in predicting compressive strength independently. The overall high scatter in residuals for most variables suggests that a combination of these predictors might provide a more accurate model for predicting.

Model Accuracy Comparisons

Comparing a regression algorithm's test and training accuracy reveals its ability to generalize. A significant gap between these accuracies suggests overfitting, where the model memorizes the training data but performs poorly on unseen data. Linear Regression, Ridge Regression, and Lasso Regression are generally less prone to this issue, while models like Decision Trees and K-Nearest Neighbors (KNN) can be more susceptible to overfitting. Ensemble methods like Random Forest and Gradient Boosting can help mitigate this by combining multiple models, often leading to improved generalization performance in predicting continuous outcomes.

Table 1 Accuracy Comparison

Model	Training Score	Testing Score
Linear Regression	86.82%	88.18%
SVM	75.36%	84.43%
KNN	92.73%	85.09%
Decision Tree Regression	97.73%	97.37%
Bagging(Decision Tree)	98.23%	97.51%
Gradient Boosting	99.25%	98.99%
Adaptive Boosting	95.06%	94.98%
Boosting(Random Forest)	99.24%	97.83%

The table showcases the performance of various machine learning models in terms of training and testing accuracy. Linear Regression demonstrates a moderate training accuracy of 86.82% and a slightly higher testing accuracy of 88.18%, indicating good generalization but limited capacity to capture complex patterns. SVM has a lower training accuracy of 75.36% and a higher testing accuracy of 84.43%, suggesting underfitting during training but reasonable performance on unseen data. KNN shows a high training accuracy of 92.73% but a significant drop to 85.09% for testing, indicating overfitting and challenges in generalizing well. Decision Tree Regression performs exceptionally with near-equal training (97.73%) and testing (97.37%) accuracies, showing it can effectively model data without overfitting. Bagging (Decision Tree) enhances stability and accuracy, achieving 98.23% for training and 97.51% for testing, indicating robust performance and good generalization. Gradient Boosting achieves the highest training accuracy at 99.25% and nearly matches this with a testing accuracy of 98.99%, showcasing excellent modeling of data patterns with minimal overfitting. Adaptive Boosting has balanced training and testing accuracies (95.06% and 94.98%, respectively), highlighting its ability to generalize well while boosting weak learners. Boosting (Random Forest) has very high training (99.24%) and testing (97.83%) accuracies, with a slight indication of overfitting, though it remains highly effective due to its ensemble nature. Overall, simpler models like Linear Regression and SVM generalize well but may struggle with complexity, while ensemble methods like Bagging, Boosting, and Random Forest deliver outstanding performance with minor overfitting, emphasizing the strength of ensemble techniques for predictive accuracy.

Conclusions:

- The prediction of compressive strength in circular concrete-filled steel tubular (CFST) columns using advanced machine learning methods represents a significant improvement over traditional empirical and analytical models.
- Techniques such as artificial neural networks (ANNs), support vector machines (SVMs), and ensemble methods like random forests and gradient boosting provide more accurate and reliable predictions of structural performance by uncovering complex patterns in extensive datasets.
- Advanced machine learning methods efficiently handle large amounts of data, integrating multiple factors such as steel tube dimensions, concrete properties, and loading scenarios to provide holistic predictions of compressive strength.
- Methods like cross-validation and hyperparameter tuning enhance the robustness and generalization capability of machine learning models, making them applicable to a wide range of real-world scenarios. Their adaptability allows for continual updates with new data, improving predictive power over time.
- Among the evaluated techniques, Gradient Boosting exhibited the best performance, achieving an accuracy of 98.99% by combining predictions from multiple weak learners (usually decision trees) and correcting previous errors, thereby improving overall model accuracy.

References:

1. Shen QH, Wang JF, Wang JX, Ding ZD. Axial compressive performance of circular CFST columns partially wrapped by carbon FRP. J Constr Steel Res 2019; 155: 90-106
2. Zhang KK, Liao FY, Huang ZW. Axial compression behavior of CFRP reinforced concrete filled steel tubes with spherical-cap gap. Journal of Building Structures 2019; 40 (S1): 02
3. Tao Z, Han LH, Zhuang JP. Using CFRP to strengthen concrete-filled steel tubular columns: stub column tests. Adv Steel Struct 2005; 1: 701-706.9 [in Chinese].
4. Prabhu GG, Sundarraja MC. Behaviour of concrete filled steel tubular (CFST) short columns externally reinforced using CFRP strips composite. Constr Build Mater 2013; 47: 1362-1371.
5. Ding FX, Lu DR, Bai Y, Gong YZ, Yu ZW, Ni M, Li W. Behaviour of CFRP-confined concrete-filled circular steel tube stub columns under axial loading. Thin-Walled Struct 2018; 125: 107-118.
6. Dong CX, Kwan AKH, Ho JCM. Axial and lateral stress-strain model for concretefilled steel tubes with FRP jackets. Eng Struct 2016; 126: 365-378.