

# ƯỚC LƯỢNG KHẢ NĂNG CHỊU NÉN ĐÚNG TÂM CỦA CỘT ỐNG THÉP NHỒI BÊ TÔNG BẰNG THUẬT TOÁN MÁY HỌC

Phạm Thái Hoàn<sup>a,\*</sup>

<sup>a</sup>*Khoa Xây dựng dân dụng và công nghiệp, Trường Đại học Xây dựng,  
55 đường Giải Phóng, quận Hai Bà Trưng, Hà Nội, Việt Nam*

*Nhận ngày 30/05/2021, Sửa xong 22/06/2021, Chấp nhận đăng 23/06/2021*

## Tóm tắt

Cấu kiện cột ống thép nhồi bê tông (Concrete-Filled Steel Tube-CFST) với nhiều ưu điểm được quan tâm nghiên cứu và ứng dụng rộng rãi trong các công trình xây dựng khoảng 3 thập kỷ qua. Bên cạnh đó, sự phát triển mạnh mẽ của kỹ thuật máy tính đã mở ra nhiều cơ hội ứng dụng các kỹ thuật trí tuệ nhân tạo (AI) vào trong thiết kế công trình. Bài báo giới thiệu và ứng dụng XGBoost, một trong những kỹ thuật trí tuệ nhân tạo mới nhất và mạnh mẽ nhất để ước lượng khả năng chịu nén dọc trục của cột ống thép nhồi bê tông (CFST) tiết diện tròn. Kết quả từ 1017 thí nghiệm về loại cấu kiện này được sử dụng làm dữ liệu để huấn luyện mô hình. Các thông số bao gồm đường kính ngoài của tiết diện ( $D$ ), chiều dày của ống thép ( $t$ ), cường độ chảy của thép ( $f_y$ ), cường độ chịu nén của bê tông ( $f'_c$ ) và chiều dài cột ( $L$ ) được sử dụng như những thông số đầu vào, trong khi đó khả năng chịu lực nén đúng tâm dọc trục ( $N$ ) là thông số đầu ra. Kết quả huấn luyện từ mô hình cho thấy thuật toán XGBoost có thể ứng dụng hiệu quả trong việc ước lượng khả năng chịu nén dọc trục của cột CFST tiết diện tròn với độ chính xác cao với thời gian ngắn.

*Từ khóa:* XGBoost; trí tuệ nhân tạo; máy học; ống thép nhồi bê tông; CFST.

## ESTIMATION OF CONCENTRIC AXIAL COMPRESSIVE LOAD-CARRYING CAPACITY OF CONCRETE-FILLED STEEL TUBE USING MACHINE LEARNING ALGORITHMS

### Abstract

Concrete-Filled Steel Tube (Concrete-Filled Steel Tube-CFST) columns with many advantages have been studied and widely applied in construction projects for decades. Besides, the rapid development of computer engineering has opened up many opportunities to apply artificial intelligence (AI) techniques in building design. This paper introduces and applies XGBoost, one of the newest and most powerful artificial intelligence techniques to estimate the concentric axial compressive load-carrying capacity of circular section concrete-filled steel tube (CFST) columns. The results from 1017 experiments on this type of columns are used as dataset to train the model. The outer diameter of the section ( $D$ ), the thickness of the steel tube ( $t$ ), the yield strength of the steel ( $f_y$ ), the compressive strength of the concrete ( $f'_c$ ), and the length of the column ( $L$ ) are used as input parameters, while the axial compression force capacity ( $N$ ) is the output parameter. The training results from the model showed that the XGBoost algorithm can be effectively applied in estimating the axial compressive capacity of the circular section CFST columns with high accuracy and short time.

*Keywords:* XGBoost; artificial intelligence; machine learning; concrete-filled steel tube; CFST.

[https://doi.org/10.31814/stce.nuce2021-15\(3V\)-06](https://doi.org/10.31814/stce.nuce2021-15(3V)-06) © 2021 Trường Đại học Xây dựng (NUCE)

## 1. Giới thiệu

Cột ống thép nhồi bê tông (concrete-filled steel tube) (CFST) là kết cấu bao gồm ống thép được lấp đầy bên trong bằng bê tông. Ống thép lúc này đóng vai trò vừa là ván khuôn cho quá trình đổ bê

\*Tác giả đại diện. Địa chỉ e-mail: [hoanpt@nuce.edu.vn](mailto:hoanpt@nuce.edu.vn) (Hoàn, P. T.)

tông, vừa tham gia tăng cường khả năng chịu tải của cùng bê tông. Trong khi đó, việc nhồi bê tông giúp tăng sự ổn định của ống thép và độ dẻo của tiết diện. Ngoài ra, cột CFST còn có khả năng hấp thụ năng lượng cao, độ bền cao và khả năng chống cháy tốt hơn [1, 2]. Vì vậy, ứng xử của cột CFST được quan tâm nghiên cứu rộng rãi trong suốt ba thập kỷ qua. Bài báo này tập trung nghiên cứu cột CFST tiết diện tròn do dạng tiết diện này có độ cứng và độ bền chảy tốt hơn so với các dạng tiết diện chữ nhật hoặc hình vuông như đã được chỉ ra trong các nghiên cứu trước đây [3–6]. Trong suốt nhiều thập kỷ qua, khá nhiều thí nghiệm về cọc tròn CFST đã được thực hiện với nhiều trọng tâm nghiên cứu khác nhau như: cường độ của bê tông [7], tỉ lệ của đường kính và độ dày ống thép [8] hay ảnh hưởng của liên kết giữa bê tông và ống thép [9], v.v. Một số chương trình thí nghiệm lớn về cọc CFST có thể kể đến như sau. Chương trình thí nghiệm của Knowles và Park [10] được thực hiện vào cuối những năm 1960 nhằm đánh giá ứng xử của cột CFST khi chịu tải đúng tâm và tải lệch tâm. Các thí nghiệm của Liu và Goel [11] cuối những năm 1980 nhằm đánh giá ứng xử của cọc CFST khi chịu tải trọng động theo chu kỳ. Kilpatrick và Rangan [12] nghiên cứu ảnh hưởng của việc sử dụng bê tông cường độ cao trong cột CFST. Sakino và cs. [13] nghiên cứu ảnh hưởng của hình dạng và cường độ ống thép, tỷ lệ đường kính ống trên chiều dày và cường độ của bê tông. Tuy nhiên, số lượng thí nghiệm thực hiện được vẫn còn nhiều hạn chế do các chương trình thí nghiệm thường đòi hỏi tiêu tốn nhiều tiền bạc, hệ thống thiết bị thử nghiệm đắt tiền, cũng như tốn nhiều thời gian và công sức của con người.

Dựa trên các kết quả thực nghiệm, các công thức thiết kế cọc CFST được đề xuất. Một cách truyền thống, các công thức thiết kế này được xây dựng dựa trên các phương pháp lý thuyết và mô phỏng dựa trên ứng xử của sự làm việc chung giữa 2 loại vật liệu thép và bê tông và lấy kết quả thí nghiệm để đánh giá độ chính xác của công thức cũng như mô hình được đề xuất. Tuy nhiên, do sai số trong quá trình thí nghiệm, sai khác giữa các chương trình thực nghiệm cũng như ảnh hưởng của các yếu tố thời tiết, vật liệu đầu vào hay tay nghề của người thực hiện khiến cho các công thức được xây dựng giữa các nhà khoa học có sự khác nhau đáng kể. Ví dụ, công thức xác định khả năng chịu tải của cột CFST được cung cấp trong các tiêu chuẩn Eurocode 4 [14], ACI 318 [15] và AISC 360 [16], tuy nhiên, không có công thức nào trong số này có sự tương đồng với nhau. Các mô hình này được xây dựng dựa trên mối quan hệ ứng suất-biến dạng được giả định trước của ống thép hoặc bê tông trong lòng nên tính chính xác của các mô hình là một dấu hỏi lớn.

Gần đây, với sự phát triển mạnh mẽ của khoa học máy tính, các kỹ thuật máy học (machine learning-ML) được ứng dụng rộng rãi trong mọi lĩnh vực của cả khoa học và cuộc sống. Các kỹ thuật ML có thể hiểu một cách đơn giản là các phương pháp cho phép xây dựng mô hình toán học phức tạp có độ chính xác rất cao nhằm thể hiện mối quan hệ giữa thông số đầu vào và thông số đầu ra của một tập dữ liệu cho trước. Với quan điểm này, ML được nhiều nhà khoa học hiện nay sử dụng để xác định ứng xử của công trình [17–20]. Ứng dụng ML trong ước lượng ứng xử của cột CFST cũng được nghiên cứu rộng rãi [21, 22]. Các nghiên cứu này cho thấy rằng nếu như chúng ta có một bộ dữ liệu đủ lớn thì các kỹ thuật ML hoàn toàn có thể được sử dụng để xây dựng nên các mô hình hiệu quả để ước lượng ứng xử của công trình nhằm thay thế hoặc hỗ trợ trong việc chính xác hóa các công thức thiết kế truyền thống. Một số thuật toán ML mạnh mẽ hiện nay có thể kể đến như là: Gradient Tree Boosting (GTB), học sâu (Deep Learning) (DL), Support Vector Machine (SVM), rừng ngẫu nhiên (Random Forest) (RF), cây quyết định (Decision Tree) (DT), v.v. Hiệu quả của ứng dụng thuật toán ML trong thiết kế công trình phụ thuộc nhiều yếu tố trong đó việc lựa chọn tham số hệ thống phù hợp cho từng lớp bài toán cụ thể là rất quan trọng. Mỗi lớp bài toán sẽ cần được thực hiện, phân tích để tìm ra cách sử dụng các tham số sao cho mô hình huấn luyện hiệu quả và tối ưu về mặt thời gian.

Trên cơ sở phân tích ở trên, trong nghiên cứu này, tác giả trình bày chi tiết cách xây dựng mô hình huấn luyện cho bài toán ước lượng khả năng chịu tải của cột CFST sử dụng thuật toán Extreme

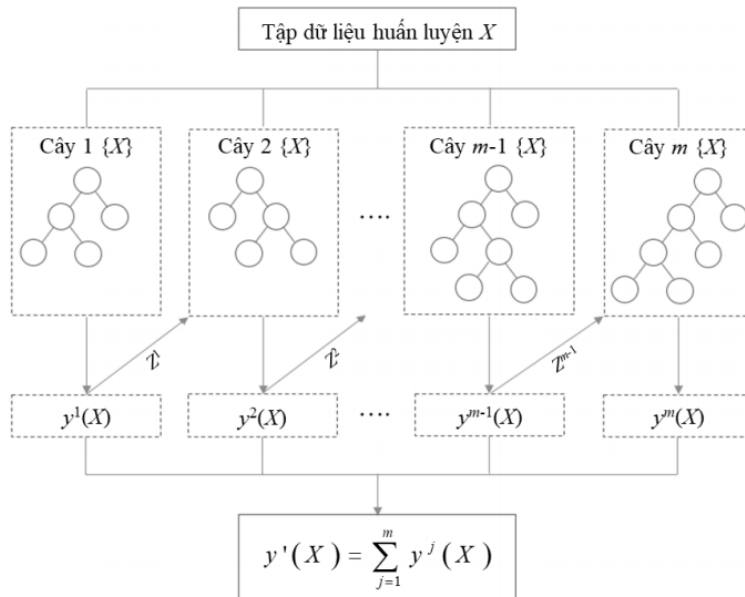
Gradient Boosting (XGBoost). XGBoost là một trong những thuật toán ML mới nhất được Chen và Guestin [23] đề xuất năm 2016 và tỏ ra rất hiệu quả trong việc giải quyết các bài toán phân tích hồi qui và phân loại [23–25]. Một tập dữ liệu các kết quả thí nghiệm bao gồm 1017 mẫu các cột CFST tròn được sử dụng để huấn luyện và kiểm tra chương trình được xây dựng. Ba mô hình ML khác bao gồm SVM, RF và Adaboost cũng được sử dụng để chứng minh tính hiệu quả của chương trình đề xuất.

## 2. Thuật toán XGBoost

XGBoost được Chen và Guestin [23] đề xuất vào năm 2016 dựa trên nguyên lý chung của các thuật toán cây tuần tự như GTB trên cơ sở kết hợp nhiều cây ra quyết định có độ sai số cao (còn gọi là các cây cơ sở hay cây học yếu) thành một cây ra quyết định tốt hơn. Nguyên lý cơ bản này được mô tả trong Hình 1. Đầu tiên, một cây ra quyết định được tạo ra một cách ngẫu nhiên. Tương ứng với dữ liệu đầu vào  $X_i$ , nó sẽ ước lượng được dữ liệu đầu ra là  $y_i^1$ . Lúc này sai số của dữ liệu ước lượng được,  $y_i^1$ , và dữ liệu chính xác,  $y_i$ , được gọi là  $z_i^1$ . Các giá trị  $z_i^1$  cũng chính là sai số của mô hình huấn luyện. Để giảm sai số này, cây ra quyết định thứ 2 được tạo ra. Khác với cây ra quyết định thứ nhất, dữ liệu đầu ra cho cây thứ 2 chính là các giá trị  $z_i^1$ . Thông số đầu ra của cây thứ 2 gọi là  $y_i^2$  và độ sai lệch của nó với  $z_i^1$  được gọi là  $z_i^2$ . Hiển nhiên rằng giá trị của  $z_i^2$  là bé hơn  $z_i^1$  hay nói cách khác là sai số của mô hình huấn luyện lúc này đã được giảm đi. Cứ như vậy, với việc tạo ra tuần tự nhiều cây ra quyết định, sai số của mô hình huấn luyện sẽ giảm dần hay độ chính xác của nó được tăng lên. Giá trị ước lượng cuối cùng của mô hình huấn luyện GTB sẽ là:

$$y'_i = \sum_{j=1}^m y_i^j \tag{1}$$

trong đó  $m$  là số cây ra quyết định được sử dụng.



Hình 1. Sơ đồ minh họa quá trình huấn luyện của thuật toán GTB

Quá trình huấn luyện của GTB chính là quá trình tối thiểu hóa tổng sai số giữa giá trị ước lượng cuối cùng  $y'_i$  và  $y_i$  ( $i = 1, 2, \dots, N$ ) được thể hiện qua hàm mất mát  $L(y_i, y'_i)$ . Có khá nhiều dạng hàm mất mát có thể được sử dụng, trong đó phổ biến nhất là hàm sai số toàn phương trung bình (mean squared error) (MSE) được viết dưới dạng như sau:

$$MSE = \sum_{i=1}^N (y_i - y'_i)^2 \quad (2)$$

Để nâng cao hiệu quả của thuật toán GTB, đặc biệt là tránh hiện tượng mô hình quá khớp, trong thuật toán XGBoost một hàm chính quy hóa (regularization)  $\Omega()$  được thêm vào và hàm mất mát mới ở vòng lặp thứ  $t$  được thể hiện như sau:

$$J^{(t)} = \sum_{i=1}^N L(y_i, y_i^{(t)}) + \sum_{n=1}^t \Omega(f_n) \quad (3)$$

trong đó các hàm  $f_n$  chứa thông tin về cấu trúc của cây ra quyết định và điểm của các lá. Lúc này, thông qua quá trình tối ưu hàm mục tiêu  $J$ , cấu trúc của các cây ra quyết định cũng sẽ được tối ưu. Tuy nhiên, việc huấn luyện cây ra quyết định là một bài toán phức tạp hơn rất nhiều so với bài toán tối ưu thông thường do rất khó để kiểm soát tất cả các cây cùng một lúc. Do vậy, giải pháp bổ sung thường được áp dụng trong đó những gì đã được học sẽ được giữ cố định lại và thêm cây một cây mới cùng thời điểm đó. Lúc này, thông số đầu ra  $y_i^{(t)}$  được xác định như sau:

$$\begin{aligned} y_i^{(0)} &= 0 \\ y_i^{(1)} &= f_1(\mathbf{X}_i) = y_i^{(0)} + f_1(\mathbf{X}_i) \\ y_i^{(2)} &= f_1(\mathbf{X}_i) + f_2(\mathbf{X}_i) = y_i^{(1)} + f_2(\mathbf{X}_i) \\ &\dots \\ y_i^{(t)} &= \sum_{n=1}^t f_n(\mathbf{X}_i) = y_i^{(t-1)} + f_t(\mathbf{X}_i) \end{aligned} \quad (4)$$

Nếu xem hàm tối ưu là hàm MSE thì phương trình (3) lúc này được viết thành:

$$\begin{aligned} J^{(t)} &= \sum_{i=1}^N L(y_i, y_i^{(t)}) + \sum_{n=1}^t \Omega(f_n) \\ &= \sum_{i=1}^N L(y_i, y_i^{(t-1)} + f_t(\mathbf{X}_i)) + \Omega(f_t) + \mathbf{constant} \end{aligned} \quad (5)$$

Về phía phương trình (5) được sử dụng biến đổi Taylor và giữ đến số hạng bậc 2. Lúc này ta có:

$$J^{(t)} = \sum_{i=1}^N \left[ L(y_i, y_i^{(t-1)}) + g_i f_t(\mathbf{X}_i) + \frac{1}{2} h_i f_t^2(\mathbf{X}_i) \right] + \Omega(f_t) + \mathbf{constant} \quad (6)$$

trong đó:

$$\begin{aligned} g_i &= \partial_{y_i^{(t-1)}} L(y_i, y_i^{(t-1)}) \\ h_i &= \partial_{y_i^{(t-1)}}^2 L(y_i, y_i^{(t-1)}) \end{aligned} \quad (7)$$

Loại bỏ các hằng số trong công thức (6), quá trình tối ưu sẽ là tối thiểu hóa hàm:

$$J^{(t)} = \sum_{i=1}^N \left[ g_i f_i(\mathbf{X}_i) + \frac{1}{2} h_i f_i^2(\mathbf{X}_i) \right] + \Omega(f_i) \quad (8)$$

Công thức (8) cho thấy rằng hàm mục tiêu lúc này chỉ phụ thuộc vào  $g_i$  và  $h_i$ . Do đó, XGBoost có thể tùy chọn hàm mất mát hiệu quả hơn và có thể áp dụng cho nhiều dạng hàm mất mát hơn với cùng một cách giải là xem  $g_i$  và  $h_i$  là biến đầu vào.

Hàm cấu trúc cây  $f_n$  được định nghĩa như sau:

$$f_t(\mathbf{X}) = w_q(\mathbf{X}) \quad (9)$$

trong đó  $w$  là véc-tơ điểm tại các lá,  $q$  là một hàm được gán cho mỗi dữ liệu tại lá tương ứng.

Hàm  $\Omega(f_n)$  có thể được chọn như sau [20]:

$$\Omega(f_n) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (10)$$

trong đó  $\gamma$  thể hiện cho độ phức tạp của tổ chức các lá trong cây ra quyết định;  $T$  là số lượng lá của một cây ra quyết định;  $\lambda$  là hệ số phạt của hàm phạt;  $w$  là véc-tơ điểm của các lá.

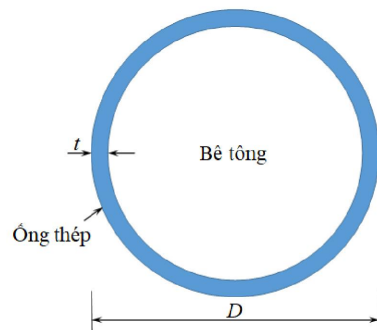
Lúc này hàm mục tiêu được viết thành:

$$\begin{aligned} J^{(t)} &\approx \sum_{i=1}^T \left[ \left( \sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left( \sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right] + \gamma T \\ &= \sum_{i=1}^T \left[ G_j w_j + \frac{1}{2} (H_j + \lambda) w_j^2 \right] + \gamma T \end{aligned} \quad (11)$$

trong đó  $I_j$  đại diện cho tất cả các mẫu trong lá thứ  $j$ . Bây giờ bài toán cực tiểu của hàm mất mát được chuyển thành bài toán tìm giá trị nhỏ nhất của hàm bậc hai. Nói cách khác, sau khi phân chia một nút nào đó trong cây quyết định, nếu hiệu suất của mô hình được cải thiện thì phân chia đó sẽ được chấp nhận. Nếu không, quá trình chia tách nút sẽ dừng lại. Nhờ đó, sự phân chia nút trong cây ra quyết định sẽ dừng sớm hơn giúp tăng hiệu suất của quá trình huấn luyện và tránh hiện tượng mô hình quá khớp.

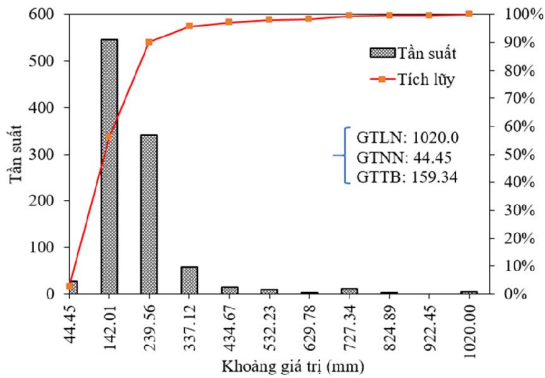
### 3. Xây dựng tập dữ liệu huấn luyện

Hình 2 thể hiện tiết diện ngang điển hình của cột CFST tiết diện tròn. Trong đó, đường kính ngoài của tiết diện và chiều dày của ống thép ký hiệu là  $D$  và  $t$ . Tập dữ liệu gồm 1017 kết quả thí nghiệm về cột CFST tiết diện tròn chịu nén đúng tâm do Thai và cs. [26] thu thập được sử dụng. Chi tiết bộ dữ liệu có thể xem tìm thấy trong tài liệu [26]. Các dữ liệu bao gồm 6 thông số:  $D$ ,  $t$ , cường độ chảy của thép ( $f_y$ ), cường độ chịu nén của bê tông ( $f'_c$ ), chiều dài cột ( $L$ ) và khả năng chịu nén

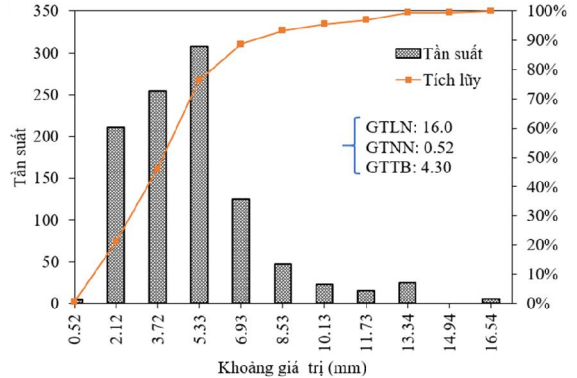


Hình 2. Minh họa tiết diện ống thép nhồi bê tông CFST

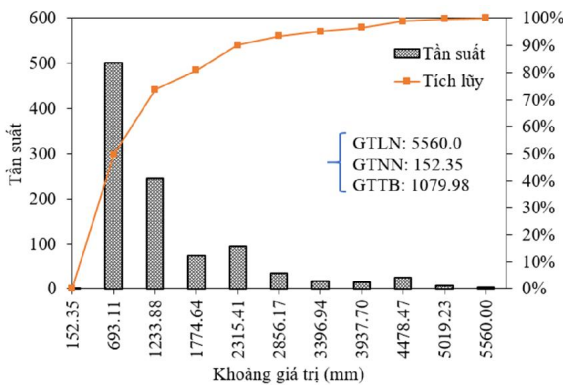
đúng tâm của cột ( $N$ ). Sự phân bố của các thông số này được thể hiện trong Hình 3. Các giá trị lớn nhất (GTLN), giá trị nhỏ nhất (GTNN) và giá trị trung bình (GTTB) của các thông số cũng được thể hiện. Kết quả cho thấy là phần lớn cột CFST được thí nghiệm với cường độ thép thông thường (trên 90% mẫu có  $f_y$  dưới 460 MPa và trên 75% mẫu có  $f'_c \leq 50$  MPa). Cùng với đó là phần lớn mẫu là dạng tiết diện chẵn và không chẵn, có rất ít tiết diện thuộc dạng lẻ. Sự tương quan giữa các thông



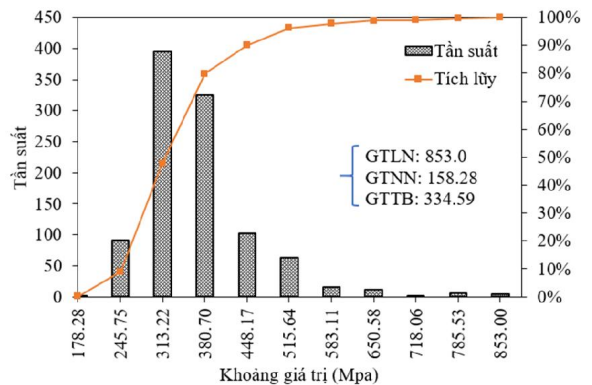
(a) Đường kính ngoài của cột ( $D$ )



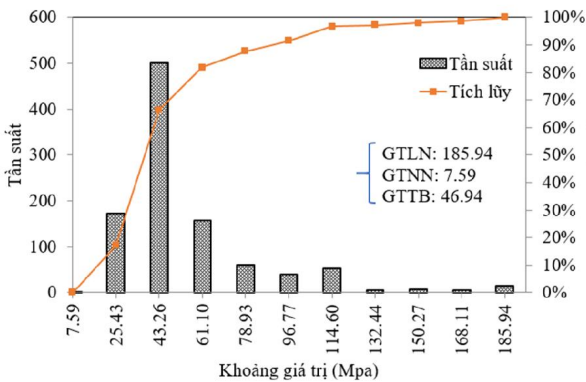
(b) Chiều dày của ống thép ( $t$ )



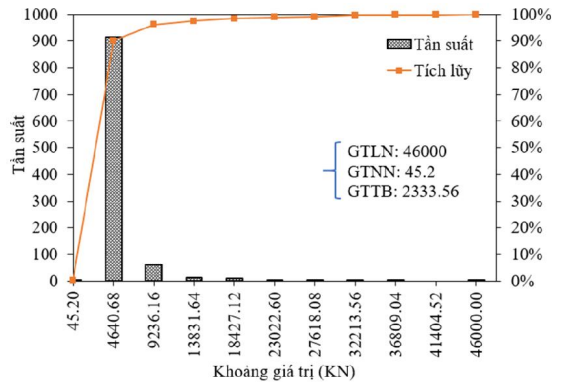
(c) Chiều dài cột ( $L$ )



(d) Cường độ chảy của thép ( $f_y$ )



(e) Cường độ chịu nén của bê tông ( $f'_c$ )

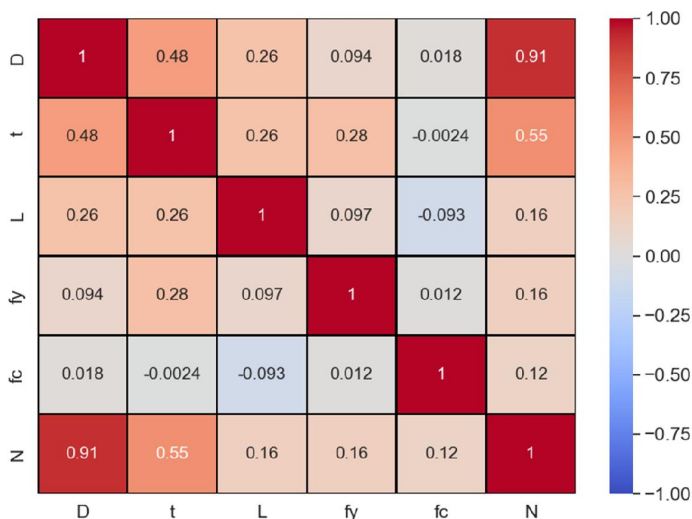


(f) Khả năng chịu nén đúng tâm của cột ( $N$ )

Hình 3. Phân bố của các thông số cột CFST



số được thể hiện trong Hình 4. Kết quả cho thấy có mối tương quan rất lớn giữa  $N$  và  $D$  cũng như là  $N$  và  $t$ .



Hình 4. Tương quan giữa các thông số khảo sát

Trong mô hình huấn luyện, các thông số đầu vào được lựa chọn là 5 thông số:  $D, t, f_y, f'_c$  và  $L$ . Thông số đầu ra chỉ có 1 thông số là  $N$ . Từ tập dữ liệu ban đầu, 2 tập dữ liệu riêng biệt được xây dựng là tập huấn luyện và tập kiểm tra. Các dữ liệu được chuẩn hóa về trong khoảng  $[0, 1]$  dựa trên công thức sau:

$$x'_i = \frac{x_i}{[x_i]} \tag{12}$$

trong đó  $[x_i]$  là giá trị lớn nhất của biến đầu vào  $x_i$ . Dạng mô hình sử dụng trong nghiên cứu này là mô hình hồi quy. Ngôn ngữ lập trình Python cùng thư viện Sklearn được sử dụng để xây dựng chương trình huấn luyện. Một số thông số chính của mô hình huấn luyện sử dụng thuật toán XGBoost được lấy như sau: số lượng cây ra quyết định  $n\_estimators = 10000$ , dạng cây ra quyết định là 'gbtree', tỉ lệ học được lấy bằng  $learning\_rate = 0,01$ , dạng hàm mất mát là MSE,  $\gamma = 0$  và  $\lambda = 1$ . Các tham số khác lấy theo giá trị hệ thống cho sẵn. Lưu ý rằng, các giá trị này được xác định dựa trên việc thử nhiều lần nhằm tìm ra kết quả tốt nhất cho mô hình huấn luyện.

#### 4. Kết quả

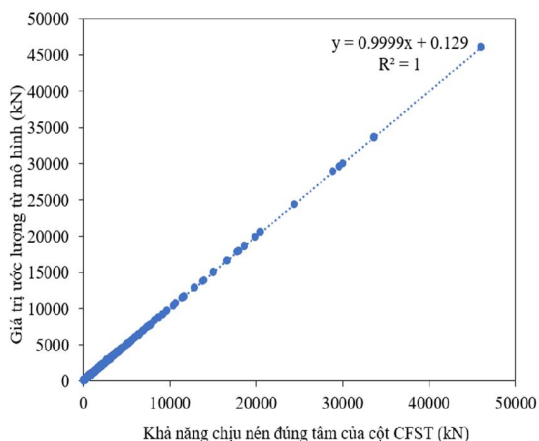
Trong phần này, hiệu suất của XGBoost được so sánh với 3 thuật toán ML phổ biến khác hiện nay là SVM, RF và Adaboost. Để tăng tính khách quan trong kết quả tính toán của các thuật toán, kỹ thuật Kfold được sử dụng với 10 tập con khác nhau. Điều này có nghĩa là tập dữ liệu sẽ được chia thành 10 tập con, và lần lượt chạy 10 lần trong đó cứ 1 lần chạy sẽ lấy 9 tập con tạo thành tập huấn luyện và tập còn lại là tập kiểm tra mô hình. Kết quả cuối cùng sẽ là giá trị trung bình của cả 10 lần chạy đó. Để giảm thiểu ảnh hưởng của tính ngẫu nhiên trong quá trình huấn luyện, mỗi thuật toán sẽ chạy 10 lần độc lập. Kết quả so sánh của các thuật toán được đánh giá dựa trên 3 giá trị là MSE, hệ số  $R^2$  và hệ số  $R^2$  hiệu chỉnh ( $\overline{R^2}$ ). Kết quả của mô hình huấn luyện đối với tập kiểm tra được thể hiện trong Bảng 1. Kết quả cho thấy rằng giá trị MSE trung bình khi sử dụng XGBoost chỉ là  $7,08E-5$ , nhỏ hơn khá nhiều so với kết quả thu được khi sử dụng SVM, RF và Adaboost. Mô hình huấn luyện sử dụng

XGBoost cũng cho kết quả  $R^2$  và  $\overline{R^2}$  là cao nhất trong 4 thuật toán ML được xem xét. Điều này có nghĩa rằng, đối với bài toán được xem xét, XGBoost có hiệu suất tốt hơn 3 thuật toán SVM, RF và Adaboost. Ngoài ra, thời gian chạy trung bình của XGBoost chỉ là 18,38 (s), nhỏ hơn khá nhiều so với 29,32 (s) của SVM, 68,73 (s) của RF và 34,67 (s) của Adaboost.

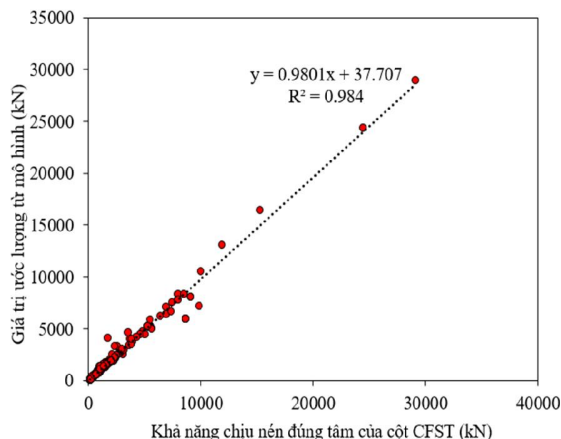
Bảng 1. So sánh các thuật toán ML đối với tập kiểm tra

Lần chạy	XGBoost			SVM			RF			Adaboost		
	MSE	$R^2$	$\overline{R^2}$	MSE	$R^2$	$\overline{R^2}$	MSE	$R^2$	$\overline{R^2}$	MSE	$R^2$	$\overline{R^2}$
1	3,21E-5	0,995	0,994	2,25E-4	0,968	0,966	2,15E-4	0,970	0,968	9,88E-5	0,985	0,985
2	4,35E-5	0,994	0,993	2,33E-4	0,963	0,961	2,24E-4	0,970	0,968	1,77E-4	0,982	0,981
3	4,67E-5	0,993	0,993	2,35E-4	0,960	0,958	2,33E-4	0,972	0,971	3,77E-4	0,967	0,965
4	1,78E-4	0,985	0,985	2,60E-4	0,962	0,960	2,22E-4	0,969	0,968	9,37E-5	0,982	0,981
5	5,07E-5	0,990	0,990	2,21E-4	0,972	0,970	4,43E-4	0,963	0,961	7,47E-5	0,987	0,986
6	3,56E-5	0,994	0,994	2,60E-4	0,965	0,963	1,73E-4	0,980	0,979	3,10E-4	0,977	0,975
7	3,55E-5	0,994	0,994	2,35E-4	0,967	0,965	3,78E-4	0,968	0,966	3,59E-4	0,957	0,955
8	1,93E-4	0,975	0,973	2,25E-4	0,967	0,965	1,89E-4	0,975	0,973	8,92E-5	0,984	0,983
9	3,12E-5	0,995	0,995	2,26E-4	0,968	0,966	1,90E-4	0,971	0,970	7,49E-5	0,988	0,987
10	6,11E-5	0,989	0,989	2,54E-4	0,967	0,965	2,74E-4	0,970	0,968	8,66E-5	0,989	0,988
Trung bình	7,08E-5	0,990	0,990	2,37E-4	0,966	0,964	2,54E-4	0,971	0,969	1,74E-4	0,980	0,979
$T_{ib}$ (s)	18,38			29,32			68,73			34,67		

Để đi sâu vào đánh giá độ chính xác của mô hình huấn luyện sử dụng XGBoost, kết quả ước lượng thu được từ việc chạy ngẫu nhiên 1 lần chương trình huấn luyện được so sánh với giá trị chính xác tương ứng. Tỷ lệ mẫu của tập huấn luyện và tập kiểm tra là 80/20. Kết quả được trình bày trong Hình 5. Các đường nét đứt trong hình thể hiện mối quan hệ giữa giá trị thí nghiệm và giá trị ước lượng từ mô hình. Các đường này có dạng phương trình là  $y = ax + b$ . Một mô hình tốt sẽ có giá trị  $a$  tiến gần đến 1 và  $b$  tiến gần đến 0. Ta có thể thấy rằng trong Hình 5, mô hình XGBoost có độ chính xác rất cao đối với cả tập huấn luyện và tập kiểm tra khi hệ số  $a > 0,98$  và  $b$  khá nhỏ. Hệ số tương quan  $R^2$  đạt được lên đến trên 98,4%.



(a) Xét với tập huấn luyện

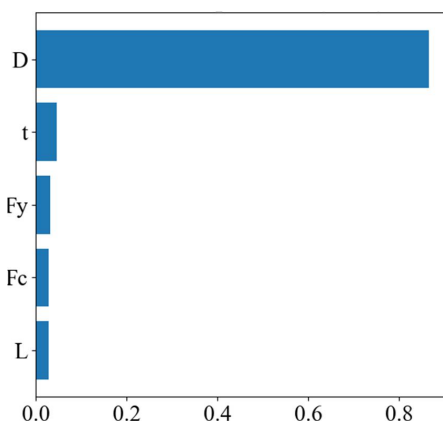


(b) Xét với tập kiểm tra

Hình 5. Giá trị ước lượng của mô hình huấn luyện so với kết quả thí nghiệm



Tầm quan trọng của các thông số đầu vào được thể hiện trong Hình 6. Có thể thấy rằng đường kính ngoài của cột ( $D$ ) có ảnh hưởng nhất đến mô hình huấn luyện và lớn hơn khá nhiều so với các thông số còn lại. Tiếp theo đó là chiều dày của của ống ( $t$ ). Ảnh hưởng của các thông số  $f_y$ ,  $f'_c$  và  $L$  đến mô hình huấn luyện là tương đồng với nhau.



Hình 6. Tầm quan trọng của các thông số đầu vào

## 5. Kết luận

Trong nghiên cứu này, một chương trình ước lượng khả năng chịu nén dọc trục của cột CFST tiết diện tròn được xây dựng. XGBoost, một trong những thuật toán máy học mạnh nhất hiện nay, được sử dụng. Một tập dữ liệu gồm 1017 kết quả thí nghiệm về cột CFST tiết diện tròn chịu nén đúng tâm được sử dụng làm dữ liệu huấn luyện. Thông số đầu vào của chương trình bao gồm: đường kính ngoài của tiết diện ( $D$ ), chiều dày của ống thép ( $t$ ), cường độ chảy của thép ( $f_y$ ), cường độ chịu nén tiêu chuẩn của bê tông ( $f'_c$ ) và chiều dài cột ( $L$ ). Thông số đầu ra là khả năng chịu nén đúng tâm của cột ( $N$ ). Kết quả huấn luyện của mô hình sử dụng XGBoost có MSE trung bình là  $7,08E-5$ , và giá trị trung bình của  $R^2$  và  $R^2$  hiệu chỉnh cùng bằng 99%. So với SVM, RF và Adaboost, hiệu suất của XGBoost là tốt nhất xét trên 3 chỉ số MSE,  $R^2$  và  $R^2$  hiệu chỉnh. Thời gian huấn luyện của thuật toán XGBoost cũng là nhanh nhất. Do đó, ta có thể kết luận rằng thuật toán XGBoost có thể ứng dụng hiệu quả trong việc ước lượng khả năng chịu nén dọc trục của cột CFST tiết diện tròn.

## Tài liệu tham khảo

- [1] Alatshan, F., Osman, S. A., Mashiri, F., Hamid, R. (2019). [Explicit Simulation of Circular CFST Stub Columns with External Steel Confinement under Axial Compression](#). *Materials*, 13(1):23.
- [2] Akbar, A., Farooq, F., Shafique, M., Aslam, F., Alyousef, R., Alabduljabbar, H. (2021). [Sugarcane bagasse ash-based engineered geopolymer mortar incorporating propylene fibers](#). *Journal of Building Engineering*, 33:101492.
- [3] Ren, Q., Li, M., Zhang, M., Shen, Y., Si, W. (2019). [Prediction of Ultimate Axial Capacity of Square Concrete-Filled Steel Tubular Short Columns Using a Hybrid Intelligent Algorithm](#). *Applied Sciences*, 9 (14):2802.
- [4] Zhang, Lyu, Yu (2019). [Prediction and Analysis of the Residual Capacity of Concrete-Filled Steel Tube Stub Columns under Axial Compression Subjected to Combined Freeze–Thaw Cycles and Acid Rain Corrosion](#). *Materials*, 12(19):3070.

- [5] Việt, V. Q., Hùng, T. V., Hoàn, P. T. (2019). Nghiên cứu khả năng chịu uốn của ống tròn hai lớp thép nhồi bê tông có liên kết mối nối bằng mô phỏng phần tử hữu hạn. *Tạp chí Khoa học Công nghệ Xây dựng (KHCN XD) - ĐHXD*, 13(4V):115–128.
- [6] Viet, V. Q., Ha, H., Hoan, P. T. (2019). Evaluation of ultimate bending moment of circular concrete-filled double skin steel tubes using finite element analysis. *Journal of Science and Technology in Civil Engineering (STCE) - NUCE*, 13(1):21–32.
- [7] Chen, J. B., Chan, T. M., Castro, J. M. (2017). Parametric study on the flexural behaviour of circular rubberized concrete-filled steel tubes. *Tubular Structures XVI*, CRC Press, 51–59.
- [8] Li, G., Chen, B., Yang, Z., Feng, Y. (2018). Experimental and numerical behaviour of eccentrically loaded high strength concrete filled high strength square steel tube stub columns. *Thin-Walled Structures*, 127: 483–499.
- [9] Javed, M. F., Sulong, N. H. R., Memon, S. A., Rehman, S. K. U., Khan, N. B. (2017). FE modelling of the flexural behaviour of square and rectangular steel tubes filled with normal and high strength concrete. *Thin-Walled Structures*, 119:470–481.
- [10] Knowles, R. B., Park, R. (1969). Strength of Concrete Filled Steel Tubular Columns. *Journal of the Structural Division*, 95(12):2565–2588.
- [11] Liu, Z., Goel, S. C. (1988). Cyclic Load Behavior of Concrete-Filled Tubular Braces. *Journal of Structural Engineering*, 114(7):1488–1506.
- [12] Kilpatrick, A. E., Rangan, B. V. (1999). Tests on High-Strength Concrete-Filled Steel Tubular Columns. *ACI Structural Journal*, 96(2).
- [13] Sakino, K., Nakahara, H., Morino, S., Nishiyama, I. (2004). Behavior of Centrally Loaded Concrete-Filled Steel-Tube Short Columns. *Journal of Structural Engineering*, 130(2):180–188.
- [14] Eurocode 4 EN 1994-1-1:2004. *Design of composite steel and concrete structures. Part 1-1: General rules and rules for buildings*. European Committee for Standardization.
- [15] ACI 318-14 (2014). *Building Code Requirements for Structural Concrete*. American Concrete Institute.
- [16] ANSI/AISC 360-16 (2016). *Specification for Structural Steel Buildings*. American Institute of Steel Construction.
- [17] Truong, V.-H., Vu, Q.-V., Thai, H.-T., Ha, M.-H. (2020). A robust method for safety evaluation of steel trusses using Gradient Tree Boosting algorithm. *Advances in Engineering Software*, 147:102825.
- [18] Kim, S.-E., Vu, Q.-V., Papazafeiropoulos, G., Kong, Z., Truong, V.-H. (2020). Comparison of machine learning algorithms for regression and classification of ultimate load-carrying capacity of steel frames. *Steel and Composite Structures*, 37(2):193–209.
- [19] Hung, T. V., Viet, V. Q., Thuat, D. V. (2019). A deep learning-based procedure for estimation of ultimate load carrying of steel trusses using advanced analysis. *Journal of Science and Technology in Civil Engineering (STCE) - NUCE*, 13(3):113–123.
- [20] Hùng, H. M., Hùng, T. V., Thuật, Đ. V., Việt, V. Q. (2020). Phương pháp xử lý hiện tượng mô hình quá khớp trong xây dựng mô hình học sâu để ước lượng khả năng chịu tải của giàn phi tuyến. *Tạp chí Khoa học Công nghệ Xây dựng (KHCN XD) - ĐHXD*, 14(1V):12–20.
- [21] Vu, Q.-V., Truong, V.-H., Thai, H.-T. (2021). Machine learning-based prediction of CFST columns using gradient tree boosting algorithm. *Composite Structures*, 259:113505.
- [22] Güneş, E. M., Gültekin, A., Mermerdaş, K. (2016). Ultimate capacity prediction of axially loaded CFST short columns. *International Journal of Steel Structures*, 16(1):99–114.
- [23] Chen, T., Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM.
- [24] Zhang, H., Qiu, D., Wu, R., Deng, Y., Ji, D., Li, T. (2019). Novel framework for image attribute annotation with gene selection XGBoost algorithm and relative attribute model. *Applied Soft Computing*, 80:57–79.
- [25] Nobre, J., Neves, R. F. (2019). Combining Principal Component Analysis, Discrete Wavelet Transform and XGBoost to trade in the financial markets. *Expert Systems with Applications*, 125:181–194.
- [26] Thai, S., Thai, H.-T., Uy, B., Ngo, T. (2019). Concrete-filled steel tubular columns: Test database, design and calibration. *Journal of Constructional Steel Research*, 157:161–181.