

XÂY DỰNG

TẠP CHÍ ĐIỆN TỬ CỦA BỘ XÂY DỰNG
JOURNAL OF CONSTRUCTION

TẠP CHÍ XÂY DỰNG - eISSN 3030-4482

Dự báo khả năng chịu lực dọc trục của cột ống thép bê tông composite hai lớp bằng mô hình Rừng ngẫu nhiên và giải thích thuật toán

Prediction of axial load capacity for hybrid FRP-concrete-steel double-skin tubular columns via Random Forest and model interpretability

➤ **PGS.TS Nguyễn Thùy Anh**

Nhóm nghiên cứu Ứng dụng công nghệ 4.0 trong Giao thông vận tải (I4T), Trường Đại học Công nghệ Giao thông vận tải

Email: anhnt@utt.edu.vn

THÔNG TIN BÀI BÁO

Chuyên mục: Khoa học công nghệ

Ngày nhận bài: 15/5/2026

Ngày sửa bài: 26/5/2026

Ngày chấp nhận đăng: 09/6/2026

Ngày xuất bản Online: 22/6/2026

Tác giả liên hệ:

Email: anhnt@utt.edu.vn

TÓM TẮT

Việc dự đoán khả năng chịu lực dọc trục của cột thép bê tông composite hai lớp FRP hình elip gặp nhiều thách thức do tính phi tuyến phức tạp của cơ chế tương tác giữa ba lớp vật liệu (FRP - bê tông - thép) và sự phân bố không đồng đều của hiệu ứng kiểm chế. Nghiên cứu này phát triển mô hình Rừng ngẫu nhiên dựa trên bộ dữ liệu gồm 112 mẫu và được tinh chỉnh thông qua kỹ thuật tìm kiếm lưới kết hợp kiểm chứng chéo. Độ ổn định của thuật toán được xác thực qua 100 lần mô phỏng Monte Carlo. Kết quả cho thấy mô hình RF đạt hiệu suất dự báo cao và khả năng tổng quát hóa tốt, với hệ số xác định trên tập kiểm tra đạt 0,972 và sai số phần trăm tuyệt đối trung bình là 4,8%. Bên cạnh đó, phân tích SHAP toàn cục xác định diện tích lõi bê tông, cường độ nén của bê tông và độ dày của vỏ FRP là ba thông số có ảnh hưởng lớn nhất. Đồng thời, kỹ thuật LIME được áp dụng để giải thích cục bộ, làm rõ sự khác biệt giữa tầm quan trọng toàn cục và cục bộ ở các mẫu có cấu hình đặc thù. Nghiên cứu cung cấp một mô hình dự báo chính xác, ổn định và có khả năng giải thích, góp phần hỗ trợ kỹ sư lựa chọn thông số thiết kế tối ưu cho cột composite hai lớp FRP hình elip.

Từ khóa: Cột composite FRP; khả năng chịu lực dọc trục; rừng ngẫu nhiên; SHAP; LIME.

ABSTRACT

Predicting the axial load capacity of elliptical hybrid FRP-concrete-steel double-skin tubular columns (DSTCs) presents significant challenges due to the complex nonlinear interaction mechanisms among the three material layers (FRP, concrete and steel) and the non-uniform distribution of the confinement effect. This study develops a Random Forest (RF) model trained on a dataset of 112 samples. The model is optimized using Grid Search combined with cross-validation and its stability is verified through 100 independent Monte Carlo simulations. The results demonstrate that the RF model achieves high predictive performance and strong generalization, with a coefficient of determination of 0.972 and a mean absolute percentage error of 4.8% on the test set. Global SHAP analysis identifies the concrete core area, concrete compressive strength, and FRP tube thickness as the three most influential parameters. In addition, the LIME technique is employed to provide local explanations, highlighting the differences between global and instance-specific feature importance, particularly for samples with distinctive configurations. Overall, the study offers an accurate, stable predictive model that supports structural engineers in selecting optimal design parameters for elliptical hybrid FRP-concrete-steel double-skin tubular columns.

Keywords: Hybrid FRP composite columns; axial load capacity; random forest; SHAP; LIME.

1. ĐẶT VẤN ĐỀ

Trong những năm gần đây, các cấu kiện composite kết hợp bê tông, thép và vật liệu sợi gia cường polymer (Fiber Reinforced Polymer - FRP) đã trở thành giải pháp hiệu quả nhằm nâng cao khả năng chịu lực, độ dẻo và khả năng chống ăn mòn của cột trong công trình dân dụng và cầu đường. Trong đó, cột ống thép bê tông composite hai lớp (Double-Skin Tubular Columns - DSTC) với lớp vỏ FRP bên ngoài, ống thép bên trong và bê tông nhồi ở giữa thể hiện nhiều ưu điểm vượt trội [1, 2]. Trong kết cấu này, ống FRP đóng vai trò như hệ ván khuôn vĩnh cửu trong quá trình đổ bê tông, giúp giảm chi phí thi công, đồng thời cung cấp khả năng chống ăn mòn cao, phù hợp với môi trường khắc nghiệt. Bên cạnh đó, lõi bê tông còn giúp ngăn chặn hiện tượng mất ổn định cục bộ sớm của ống thép trong, từ đó nâng cao đáng kể khả năng chịu lực của toàn bộ cấu kiện [3].

Mặc dù phần lớn các nghiên cứu trước đây tập trung vào DSTC có tiết diện tròn hoặc chữ nhật [4, 5], tiết diện elip đang ngày càng được quan tâm. Cột DSTC hình elip cho phép điều chỉnh độ cứng uốn khác biệt theo hai trục chính mà không làm suy giảm đáng kể hiệu ứng kiểm chế của lớp FRP, rất phù hợp với các trường hợp chịu tải trọng không đồng đều theo hai phương ngang [6]. Tuy nhiên, sự phức tạp về hình học cùng với cơ chế tương tác phi tuyến giữa FRP, bê tông và thép khiến việc dự đoán chính xác khả năng chịu lực dọc trục của cột elip DSTC trở nên khó khăn.

Các công thức thiết kế hiện hành và phương pháp giải tích truyền thống chủ yếu dựa trên mô hình bán thực nghiệm, thường thiếu độ chính xác khi áp dụng cho tiết diện elip do không định lượng được đầy đủ hiệu ứng kiểm chế không đồng đều. Để khắc phục những hạn chế này, học máy (Machine Learning - ML) đã được ứng dụng ngày càng rộng rãi trong lĩnh vực kỹ thuật kết cấu [7, 8]. Trong số các thuật toán ML, Rừng ngẫu nhiên (Random Forest - RF) nổi bật nhờ khả năng xử lý quan hệ phi tuyến phức tạp, giảm thiểu hiện tượng quá khớp và cung cấp thông tin về tầm quan trọng của các biến. Việc kết hợp RF với các kỹ thuật giải thích mô hình như SHAP và LIME còn giúp làm rõ đóng góp của từng tham số thiết kế đến kết quả dự đoán, nâng cao tính minh bạch và khả năng áp dụng thực tiễn. Do đó, nghiên cứu này tập trung phát triển mô hình RF được tối ưu siêu tham số bằng kỹ thuật tìm kiếm lưới để dự đoán khả năng chịu lực dọc trục

của cột DSTC hình elip chịu nén đúng tâm. Đồng thời, phân tích SHAP và LIME được thực hiện nhằm làm rõ cơ chế ảnh hưởng của các thông số đầu vào, cung cấp cơ sở khoa học trực quan hỗ trợ kỹ sư trong thiết kế.

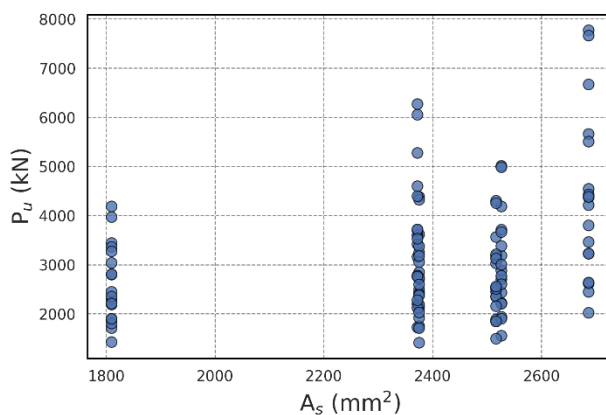
2. CƠ SỞ DỮ LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Thu thập và mô tả dữ liệu

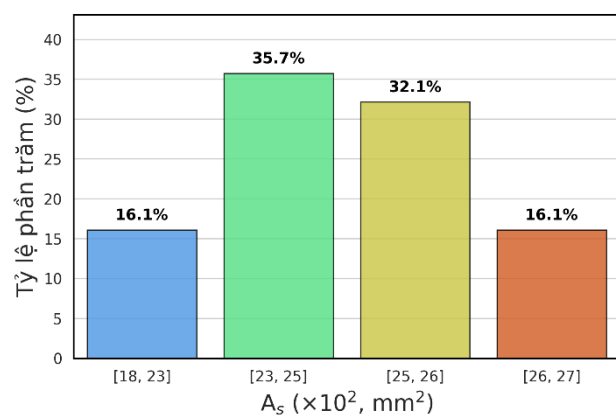
Để xây dựng và huấn luyện mô hình học máy, nghiên cứu này sử dụng bộ dữ liệu gồm 112 mẫu được trích xuất trong nghiên cứu của Yu và cộng sự [9]. Các thông số đầu vào được lựa chọn dựa trên cấu tạo của cột DSTC hình elip, bao gồm 5 thông số độc lập: Diện tích tiết diện ống thép trong (A_s), cường độ chảy của ống thép (f_y), diện tích lõi bê tông (A_c), cường độ chịu nén của bê tông (f_c), chiều dày ống FRP (t_{FRP}). Thông số đầu ra là khả năng chịu lực dọc trục của cột DSTC (P_u). Bảng 1 trình bày các giá trị thống kê mô tả chi tiết (bao gồm giá trị trung bình, độ lệch chuẩn, giá trị nhỏ nhất, trung vị và lớn nhất) của các thông số đầu vào và đầu ra. Bên cạnh đó, Hình 1 minh họa trực quan sự phân bố của từng thông số đầu vào cũng như mối quan hệ tương quan của chúng với khả năng chịu lực dọc trục P_u . Cột biểu đồ bên trái (biểu đồ phân tán) cho thấy xu hướng biến thiên của P_u khi các thông số vật liệu và hình học thay đổi. Có thể quan sát thấy các xu hướng tương quan dương, ví dụ như khi diện tích lõi bê tông hoặc cường độ vật liệu tăng, giá trị P_u nhìn chung cũng có xu hướng dịch chuyển lên các mức cao hơn. Cột biểu đồ bên phải (biểu đồ cột) thể hiện tỷ lệ phần trăm phân bố mật độ dữ liệu theo các dải giá trị cụ thể.

Bảng 1. Các thông số đầu vào và đầu ra của bộ dữ liệu

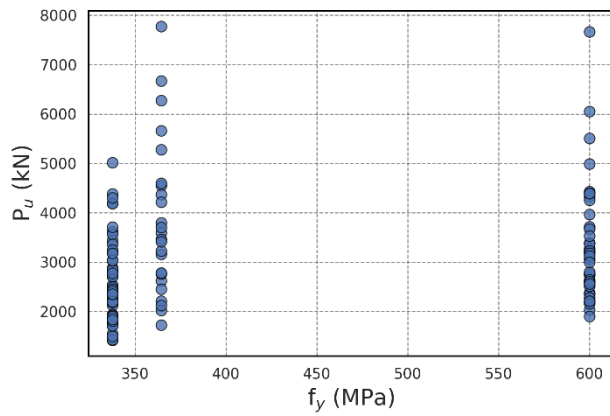
	Thông số đầu vào					Đầu ra
	A_s	f_y	A_c	f_c	t_{FRP}	P_u
GT trung bình	2.380,60	432,39	21.592,29	49,13	1,50	3.144,09
Độ lệch chuẩn	272,84	121,08	9.435,04	18,38	0,67	1.253,10
GT nhỏ nhất	1.809,56	337,60	10.684,62	28,95	0,71	1.416,30
Trung vị	2.375,04	364,50	21.113,51	43,45	1,42	2.802,45
GT lớn nhất	2.686,06	600,00	45.846,84	72,40	2,83	7.775,60



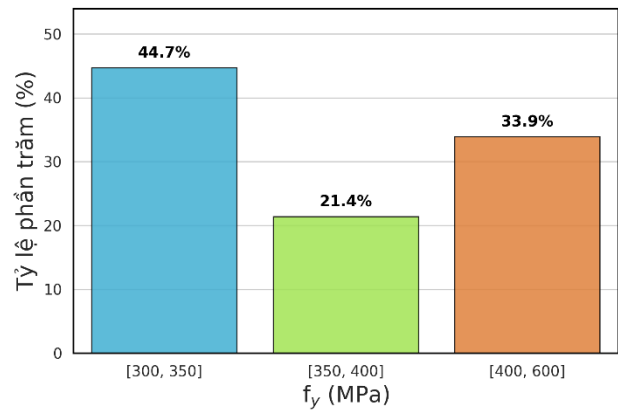
a)



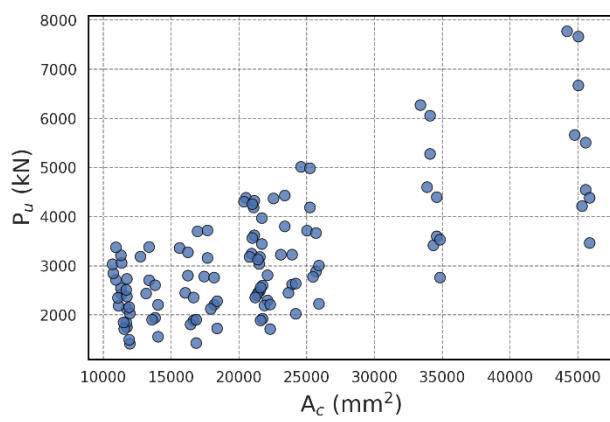
b)



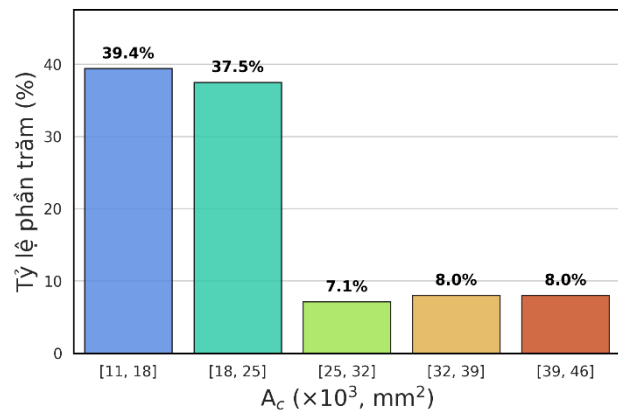
c)



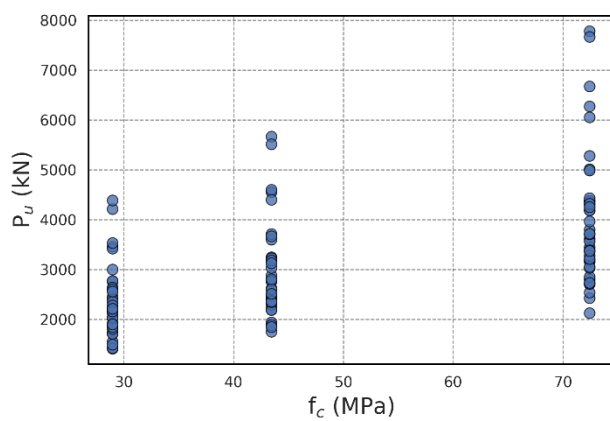
d)



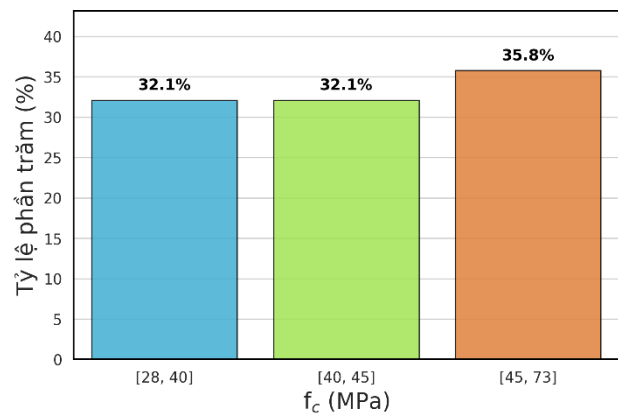
e)



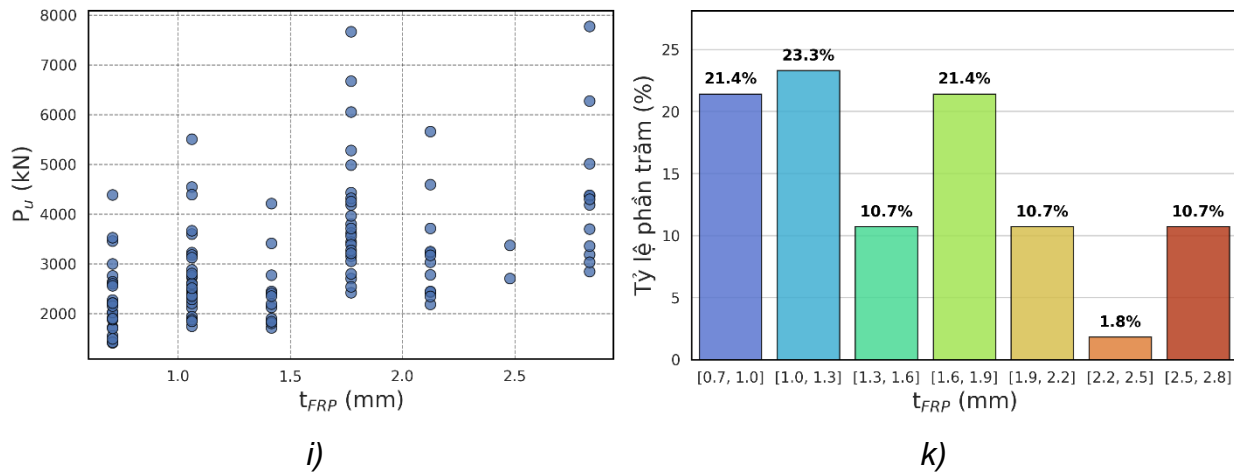
f)



e)



f)



Hình 1. Biểu đồ phân bố và tương quan của các thông số trong bộ dữ liệu nghiên cứu

Toàn bộ cơ sở dữ liệu được chia ngẫu nhiên theo tỷ lệ 7:3, trong đó tập dữ liệu đào tạo chiếm 70% và tập dữ liệu kiểm tra chiếm 30% còn lại. Quá trình tối ưu hóa các siêu tham số được thực hiện thông qua kỹ thuật kiểm tra chéo trên tập đào tạo. Để tránh hiện tượng rò rỉ dữ liệu, tập kiểm tra được tách biệt trong quá trình huấn luyện. Tập dữ liệu này chỉ được sử dụng một lần duy nhất để xác định khả năng dự báo thực tế của mô hình trên các dữ liệu chưa được biết đến.

2.2. Thuật toán rừng ngẫu nhiên

Rừng ngẫu nhiên (RF) là một thuật toán học máy mạnh mẽ thuộc nhóm học tập kết hợp, được thiết kế để giải quyết cả bài toán phân loại và hồi quy. Thuật toán này hoạt động dựa trên việc tạo ra nhiều bộ dữ liệu con từ tập dữ liệu gốc thông qua kỹ thuật lấy mẫu có hoàn lại. Đối với mỗi bộ dữ liệu con, một cây hồi quy độc lập được xây dựng, tạo thành một "khu rừng". Kết quả dự báo cuối cùng của mô hình RF đối với một mẫu mới là giá trị trung bình cộng của tất cả các dự báo từ các cây hồi quy thành phần. Cơ chế này giúp mô hình RF giảm thiểu phương sai và khắc phục hiệu quả hiện tượng quá khớp thường gặp ở các cây quyết định đơn lẻ.

2.3. Kỹ thuật giải thích mô hình

2.3.1. Phân tích SHAP (SHapley Additive exPlanations)

Dựa trên lý thuyết trò chơi hợp tác, SHAP được sử dụng để đánh giá mức độ đóng góp của từng thông số đầu vào đối với kết quả dự báo. SHAP tính toán giá trị đóng góp biên trung bình của một biến qua tất cả các tổ hợp biến có thể có. Phương pháp này cung cấp cái nhìn toàn về mức độ quan trọng của các thông số thiết kế đối với khả năng chịu lực dọc trục của cột.

2.3.2. Phân tích LIME (Local Interpretable Model-agnostic Explanations)

Bổ trợ cho SHAP, LIME được áp dụng để giải thích các dự báo cục bộ. Đối với một mẫu cấu kiện cụ thể, LIME sẽ xây dựng một mô hình đại diện tuyến tính đơn giản xung quanh lân cận của điểm dữ liệu đó. Điều này cho phép kỹ sư phân tích chi tiết tại sao mô hình học máy lại đưa ra một giá trị dự báo khả năng chịu lực dọc trục cụ thể cho một cột DSTC có cấu hình kích thước và vật liệu nhất định.

2.4. Đánh giá hiệu suất mô hình

Hiệu suất dự báo của các mô hình được đánh giá thông qua 4 chỉ số thống kê phổ biến: Hệ số xác định (R^2), sai số bình phương trung bình (RMSE), sai số tuyệt đối trung

bình (MAE) và sai số phần trăm tuyệt đối trung bình (MAPE). Các chỉ số này được tính toán như sau:

$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}$	(1)
$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$	(2)
$MAE = \frac{1}{N} y_i - \hat{y}_i $	(3)
$RMSE = \frac{1}{N} \sum_{i=1}^N \left \frac{y_i - \hat{y}_i}{y_i} \right $	(4)

Trong đó:

- y_i : Khả năng chịu lực dọc trục thực nghiệm của mẫu thứ i ;
- \hat{y}_i : Khả năng chịu lực dọc trục dự báo bởi mô hình;
- \bar{y} : Giá trị trung bình của khả năng chịu lực dọc trục thực nghiệm;
- N : Số lượng mẫu dữ liệu.

3. KẾT QUẢ VÀ THẢO LUẬN

3.1. Kết quả tối ưu hóa siêu tham số

Hiệu suất dự báo và khả năng tổng quát hóa của mô hình RF phụ thuộc đáng kể vào việc thiết lập các siêu tham số. Nếu các tham số này không được tinh chỉnh phù hợp, mô hình có thể gặp phải tình trạng quá khớp hoặc chưa khớp. Để xác định cấu trúc mô hình tối ưu nhất cho bài toán dự báo khả năng chịu lực dọc trục của cột DSTC hình elip, nghiên cứu này áp dụng phương pháp tìm kiếm lưới kết hợp với kiểm chứng chéo 10 lần. Quá trình này thực hiện tìm kiếm toàn diện trên một không gian tham số được xác định trước. Các siêu tham số quan trọng của thuật toán RF được đưa vào lưới tìm kiếm bao gồm: số lượng cây hồi quy ($n_estimators$), độ sâu tối đa của cây (max_depth), số lượng mẫu tối thiểu của nút ($min_samples_split$), số lượng mẫu tối thiểu của lá ($min_samples_leaf$), và số lượng đặc trưng ngẫu nhiên tối đa được xem xét tại mỗi nút phân tách ($max_features$). Quá trình kiểm chứng chéo giúp đảm bảo bộ siêu tham số được chọn có tính ổn định cao và không bị phụ thuộc vào cách phân chia tập dữ liệu. Tiêu chí đánh giá để lựa chọn bộ tham số tốt nhất là sai số dự báo thấp nhất trên tập huấn luyện đã được xác định. Bảng 2 tóm tắt chi tiết dải giá trị đầu vào cũng như kết quả bộ siêu tham số tối ưu cuối cùng được sử dụng để xây dựng mô hình dự báo chính thức.

Bảng 2. Không gian tìm kiếm và kết quả tối ưu siêu tham số cho mô hình RF

Siêu tham số	Dải giá trị tìm kiếm	Giá trị tối ưu được chọn
$n_estimators$	[100; 200; 300; 400; 500]	300
max_depth	[3; 4; 5; 6; 7; 8; 9; 10]	7
$min_samples_split$	[2; 4; 6; 8; 10]	2
$min_samples_leaf$	[1; 2; 3; 4; 5; 6]	1

Siêu tham số	Dải giá trị tìm kiếm	Giá trị tối ưu được chọn
max_features	['sqrt'; 0,2; 0,5; 0,8]	0,8

Để đánh giá độ ổn định và tính lặp lại của bộ siêu tham số tối ưu đã được lựa chọn, nghiên cứu thực hiện mô phỏng Monte Carlo với 100 lần chạy độc lập. Việc đánh giá hiệu suất mô hình chỉ dựa trên một lần phân chia ngẫu nhiên có thể dẫn đến những sai lệch nhất định do đặc tính phân bố của mẫu. Thông qua việc tái lấy mẫu và huấn luyện lại mô hình 100 lần với cùng một bộ siêu tham số tối ưu, phương pháp Monte Carlo cho phép định lượng mức độ biến động của các sai số dự báo một cách khách quan và toàn diện hơn. Kết quả thống kê các chỉ số đánh giá hiệu suất (bao gồm giá trị nhỏ nhất, lớn nhất, giá trị trung bình, độ lệch chuẩn (Std) và hệ số biến thiên (COV)) trên cả tập đào tạo và tập kiểm tra sau 100 lần mô phỏng được trình bày chi tiết trong Bảng 3.

Bảng 3. Thống kê các chỉ số đánh giá hiệu suất của mô hình RF sau 100 lần mô phỏng Monte Carlo

Chỉ số	Tập dữ liệu	GTNN	GTLN	TB	Std	COV (%)
RMSE (kN)	Đào tạo	136,57	172,44	154,38	8,02	5,2
	Kiểm tra	151,32	260,98	215,02	20,77	9,66
MAE (kN)	Đào tạo	97,94	124,29	111	4,91	4,42
	Kiểm tra	124,03	203,16	170,71	16,5	9,66
R ²	Đào tạo	0,9802	0,9888	0,9854	0,002	0,2
	Kiểm tra	0,9602	0,9736	0,9647	0,0035	0,36
MAPE	Đào tạo	0,0297	0,0386	0,0341	0,0016	4,71
	Kiểm tra	0,0422	0,0761	0,0593	0,0069	11,64

Phân tích số liệu từ Bảng 3 cho thấy cấu hình siêu tham số của mô hình RF thể hiện độ ổn định đồng đều qua các lần lấy mẫu. Hệ số xác định đạt giá trị trung bình 0,9854 trên tập đào tạo và 0,9647 trên tập kiểm tra. Đáng chú ý, hệ số biến thiên của chỉ số R² ở cả hai tập dữ liệu đều duy trì ở mức thấp, tương ứng là 0,20% và 0,36%. Sự phân tán nhỏ này minh chứng cho việc mô hình RF duy trì được mức độ phù hợp cao với dữ liệu thực nghiệm và ít nhạy cảm với sự thay đổi ngẫu nhiên của tập mẫu đầu vào. Đối với các đại lượng đo lường sai số, kết quả trên tập kiểm tra - vốn phản ánh năng lực dự báo độc lập của mô hình - ghi nhận sai số phần trăm trung bình tuyệt đối ở mức trung bình 5,93% với độ lệch chuẩn là 0,0069. Hệ số COV của các chỉ số RMSE và MAE trên tập kiểm tra đều là 9,66%, cho thấy biên độ dao động của sai số nằm trong giới hạn kiểm soát dưới 10%.

Sự ổn định cao của các chỉ số qua 100 lần mô phỏng Monte Carlo khẳng định bộ siêu tham số tối ưu thu được từ tìm kiếm lưới là đáng tin cậy, ít nhạy cảm với sự ngẫu nhiên của quá trình chia dữ liệu và huấn luyện. Do đó, bộ siêu tham số này được lựa chọn để xây dựng mô hình RF cuối cùng và tiến hành các phân tích giải thích tiếp theo bằng SHAP và LIME.

3.2. Hiệu suất của mô hình

Sau khi bộ siêu tham số tối ưu được xác định, mô hình RF được tiến hành huấn luyện lại trên tập dữ liệu đào tạo và được đánh giá trên tập kiểm tra độc lập. Kết quả định

lượng về độ chính xác dự báo thông qua bốn chỉ số R^2 , RMSE, MAE và MAPE được tổng hợp chi tiết trong Bảng 4.

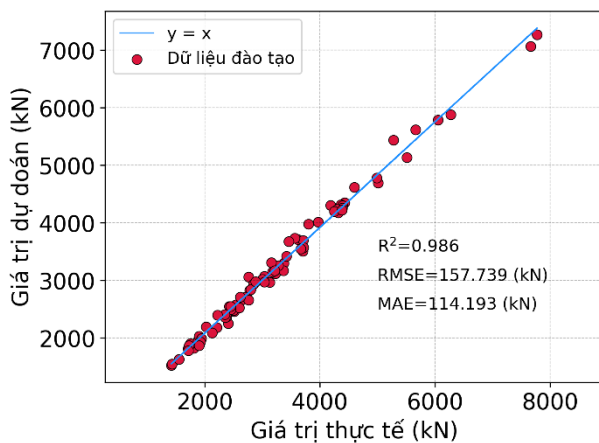
Bảng 4. Tổng hợp kết quả đánh giá hiệu suất mô hình RF

Tập dữ liệu	R^2	RMSE (kN)	MAE (kN)	MAPE
Đào tạo	0,986	157,739	114,193	0,034
Kiểm tra	0,972	162,691	135,976	0,048
Toàn bộ dữ liệu	0,984	159,259	120,806	0,039

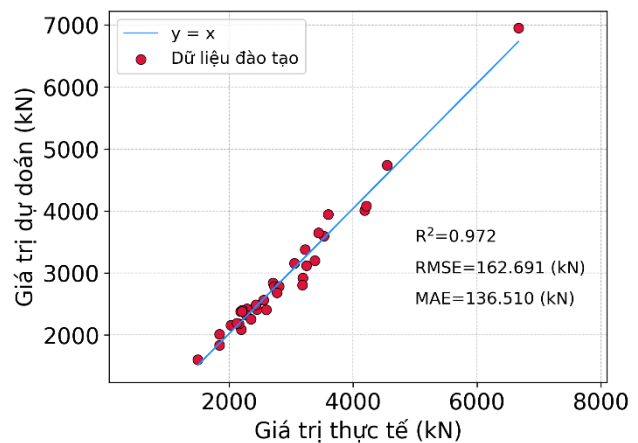
Các giá trị trong Bảng 4 cho thấy sự nhất quán giữa tập đào tạo và tập kiểm tra. Cụ thể, R^2 trên tập kiểm tra chỉ giảm nhẹ so với tập đào tạo, trong khi RMSE và MAE tăng không đáng kể. Điều này khẳng định mô hình có khả năng khái quát hóa tốt và không bị quá khớp, phù hợp với yêu cầu dự báo chính xác khả năng chịu lực dọc trục của cột DSTC.

Sự tương quan giữa khả năng chịu lực dọc trục thực tế và giá trị dự báo được minh họa qua biểu đồ hồi quy ở Hình 2. Các điểm dữ liệu trên cả tập đào tạo (Hình 2a) và tập kiểm tra (Hình 2b) đều phân bố tập trung và bám sát đường hồi quy lý tưởng $y = x$. Mức độ phân tán hẹp này xác nhận dự báo của mô hình có độ sai số thấp so với thực tế trên toàn bộ dải biến thiên của khả năng chịu lực dọc trục.

Bên cạnh đó, Hình 3 biểu diễn sự phân bố sai số dự báo thông qua biểu đồ tọa độ cực, trong đó khoảng cách từ mỗi điểm đến tâm thể hiện độ lớn của sai số tuyệt đối. Phần lớn các điểm dữ liệu hội tụ ở khu vực trung tâm (biên độ sai số thấp, chủ yếu dưới 240 kN), trong khi tần suất xuất hiện ở vùng biên ngoài (sai số lớn) là rất thấp. Việc các điểm sai số phân bố tương đối đồng đều theo các hướng quanh tâm chứng tỏ sai lệch của mô hình mang tính ngẫu nhiên, không có xu hướng sai lệch hệ thống và hoàn toàn nằm trong giới hạn cho phép của bài toán kỹ thuật kết cấu.

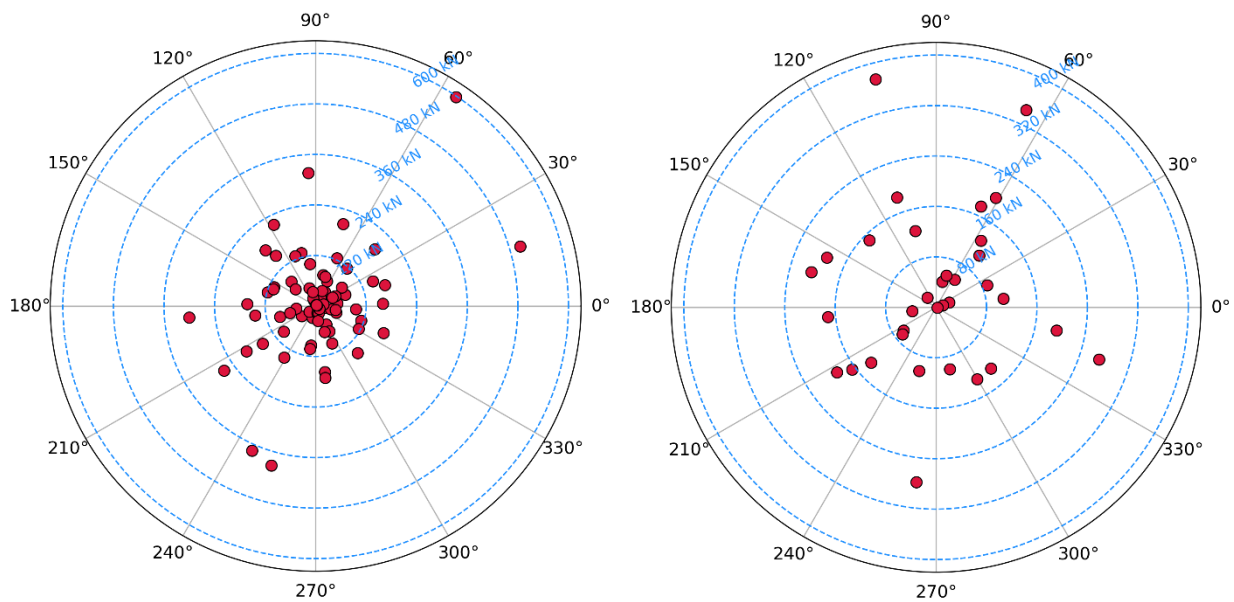


a) - Tập dữ liệu đào tạo



b) - Tập dữ liệu kiểm tra

Hình 2. Biểu đồ hồi quy so sánh giá trị thực tế và dự báo của mô hình RF trên tập đào tạo và kiểm tra



a) - Tập dữ liệu đào tạo

b) - Tập dữ liệu kiểm tra

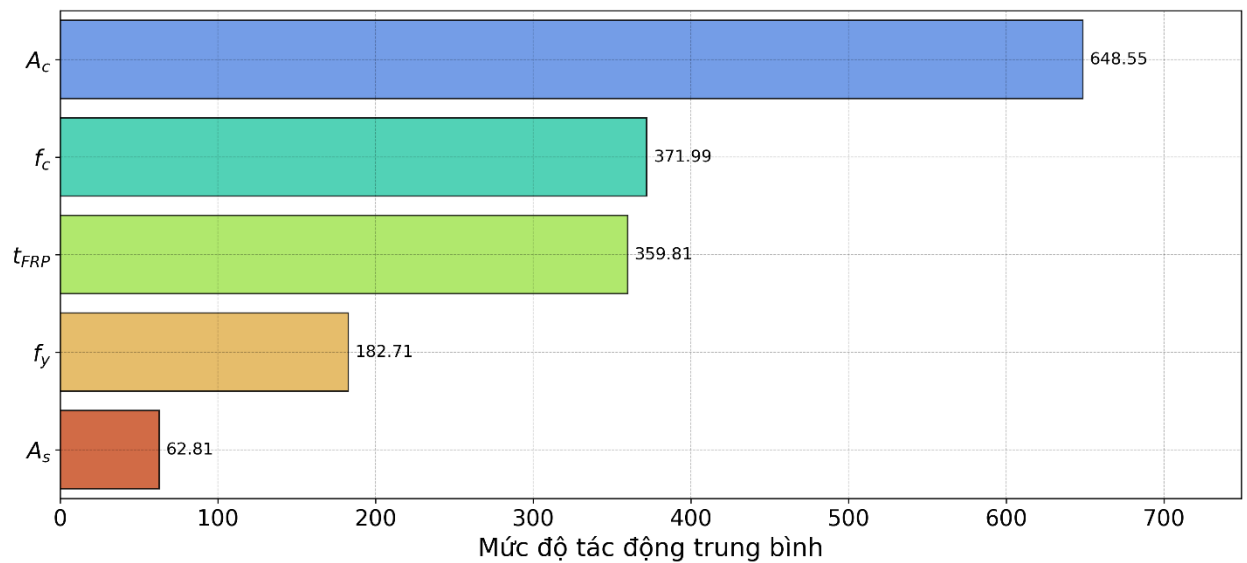
Hình 3. Biểu đồ phân bố sai số dự báo của mô hình RF trên tập đào tạo và kiểm tra

3.3. Phân tích mức độ ảnh hưởng của các thông số

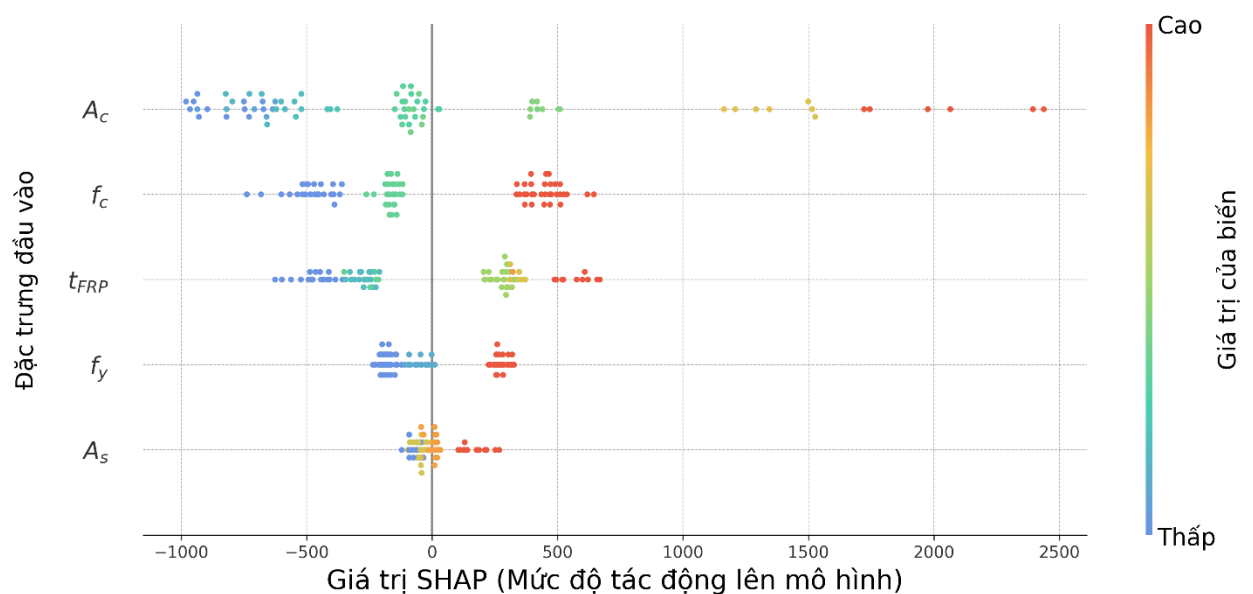
Để làm rõ mức độ đóng góp của từng biến đầu vào vào kết quả dự đoán khả năng chịu lực dọc trục P_u , phân tích SHAP được thực hiện trên mô hình RF đã huấn luyện với bộ siêu tham số tối ưu. SHAP cho phép phân bố giá trị dự báo cho từng biến, từ đó định lượng cả tầm quan trọng toàn cục lẫn chiều hướng tác động cục bộ.

Hình 4a trình bày mức độ quan trọng toàn cục của các biến, được tính bằng giá trị tuyệt đối trung bình của SHAP. Kết quả cho thấy diện tích lõi bê tông là yếu tố chi phối mạnh mẽ nhất với giá trị tác động trung bình lên tới 648,55. Tiếp theo đó, cường độ chịu nén của bê tông và chiều dày ống FRP đóng vai trò quan trọng thứ hai và thứ ba với mức độ tác động khá tương đồng, lần lượt đạt 371,99 và 359,81. Ngược lại, các thông số liên quan đến ống thép bên trong bao gồm cường độ chảy và diện tích tiết diện thép có mức độ ảnh hưởng thấp nhất (đạt 182,71 và 62,81). Thứ tự này phù hợp với cơ chế chịu lực của cột DSTC hình elip, trong đó lõi bê tông và lớp vỏ FRP là hai yếu tố quyết định chính khả năng chịu nén [10].

Hình 4b là biểu đồ tóm tắt phân phối giá trị SHAP cho từng biến. Mỗi điểm đại diện cho một mẫu dữ liệu; vị trí trên trục hoành thể hiện mức độ và chiều hướng đóng góp vào P_u (phía phải: tăng dự báo, phía trái: giảm dự báo), trong khi màu sắc biểu thị giá trị thực của biến (đỏ: cao, xanh dương: thấp). Có thể thấy A_c có ảnh hưởng tích cực rõ rệt, khi A_c tăng (điểm màu đỏ), giá trị SHAP chủ yếu dương và có độ lớn lớn. Hai biến f_c và t_{FRP} cũng chủ yếu mang lại đóng góp tích cực. Ngược lại, A_s và f_y có phạm vi ảnh hưởng hẹp hơn, với các giá trị SHAP tập trung gần giá trị 0, cho thấy hai biến này ít tác động đến P_u trong phạm vi dữ liệu nghiên cứu.



a) - Mức độ quan trọng của các biến đầu vào



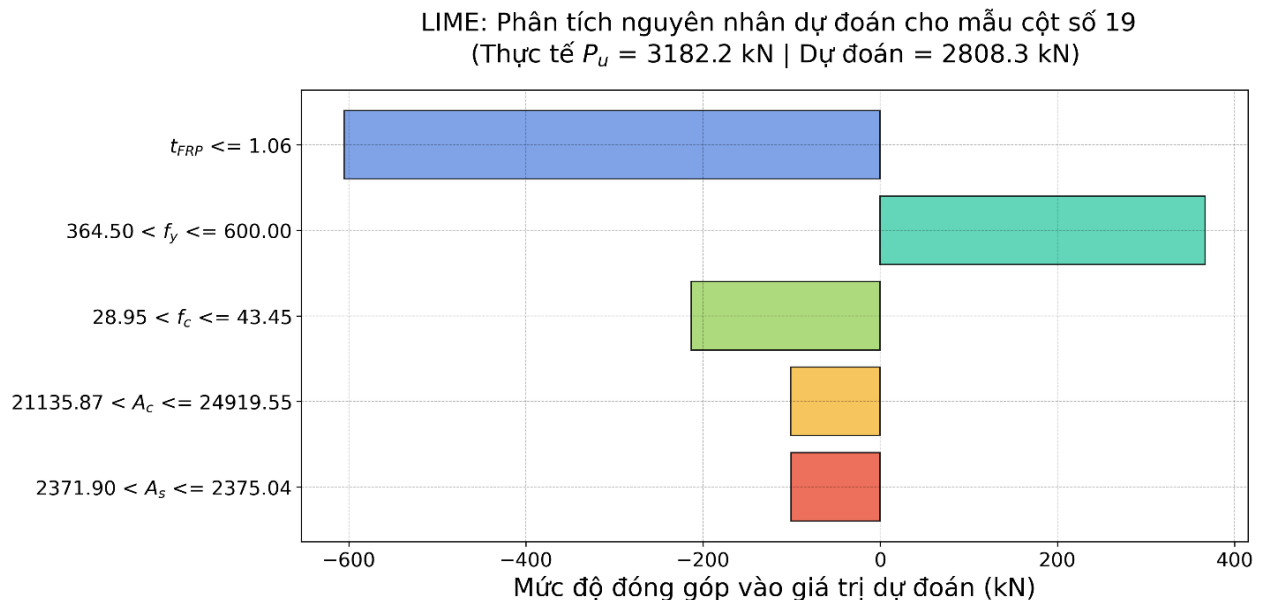
b) - Phân phối chi tiết ảnh hưởng của các biến (SHAP Values)

Hình 4. Biểu đồ SHAP mô tả chiều hướng tác động của các thông số đầu vào đến khả năng chịu lực của cột

Để bổ sung góc nhìn cục bộ, kỹ thuật LIME được áp dụng cho từng mẫu cụ thể. Hình 5 trình bày kết quả phân tích LIME cho mẫu số 19 thuộc tập kiểm tra là mẫu có sai số lớn nhất (giá trị thực tế $P_u = 3.182,2$ kN, giá trị dự báo của mô hình là 2.808,3 kN). Biểu đồ cho thấy chiều dày ống FRP mỏng ($t_{FRP} \leq 1,06$ mm) là yếu tố tiêu cực mạnh nhất, làm giảm gần 600 kN so với giá trị cơ sở. Cường độ bê tông trung bình ($28,95 < f_c \leq 43,45$ MPa) và tiết diện bê tông không lớn cũng góp phần làm giảm dự báo. Yếu tố duy nhất đóng góp tích cực là cường độ chảy của thép cao ($364,50 < f_y \leq 600,00$ MPa), giúp tăng thêm khoảng 370 kN.

Điều đáng chú ý là kết quả giải thích cục bộ từ LIME cho mẫu này khác biệt rõ rệt so với mức độ quan trọng trung bình toàn cục thu được từ SHAP. Trong khi SHAP xác định A_c là biến chi phối mạnh nhất trên toàn bộ tập dữ liệu, thì đối với mẫu cụ thể này (có

giá trị t_{FRP} thấp), chiều dày ống FRP lại trở thành yếu tố tiêu cực chiếm ưu thế. Sự khác biệt này thường xuất hiện ở các mẫu có giá trị đặc trưng cực đoan hoặc khi tồn tại tương tác phức tạp giữa các biến, giúp làm nổi bật những trường hợp mà mô hình cần được kiểm tra kỹ hơn trong thực tiễn [11].



Hình 5. Giải thích LIME cục bộ cho khả năng chịu lực dọc trục dự đoán của mẫu cột số 19

4. KẾT LUẬN VÀ KIẾN NGHỊ

Nghiên cứu này đã ứng dụng thuật toán Rừng ngẫu nhiên kết hợp với các kỹ thuật giải thích mô hình SHAP và LIME để dự báo và phân tích khả năng chịu lực dọc trục của cột ống thép bê tông composite hai lớp FRP hình elip. Với $R^2 = 0,984$ trên toàn bộ dữ liệu và độ ổn định được kiểm chứng qua 100 lần mô phỏng Monte Carlo (hệ số biến thiên của R^2 dưới 0,36%), mô hình cho thấy khả năng khái quát hóa tốt và ít nhạy cảm với sự ngẫu nhiên của tập dữ liệu. Phân tích SHAP xác định diện tích lõi bê tông, cường độ bê tông và độ dày ống FRP là ba thông số có ảnh hưởng chính. Phân tích LIME làm rõ sự khác biệt giữa giải thích toàn cục và cục bộ tại các mẫu có giá trị đặc trưng biên hoặc tương tác phức tạp.

Kết quả nghiên cứu cho thấy phương pháp kết hợp học máy với các kỹ thuật giải thích mô hình cung cấp một công cụ dự báo chính xác và minh bạch, hỗ trợ kỹ sư kết cấu trong việc lựa chọn thông số thiết kế. Dựa trên những kết quả thu được, nghiên cứu kiến nghị mở rộng bộ dữ liệu huấn luyện bằng cách bổ sung các kết quả thí nghiệm thực tế. Đồng thời, phương pháp này có thể được phát triển thêm để dự báo ứng xử của cột DSTC hình elip dưới các điều kiện tải trọng phức tạp hơn, chẳng hạn như nén lệch tâm, tải trọng chu kỳ hoặc tải trọng va đập.

TÀI LIỆU THAM KHẢO

[1] A. Hamoda, M. Ghalla, S.A. Yehia, M. Ahmed, A.A. Abadel, A. Baktheer, R.I. Shahin. Experimental and numerical investigations of the shear performance of reinforced concrete deep beams strengthened with hybrid SHCC-mesh. Case Studies in Construction Materials, 21, 2024, doi: 10.1016/j.cscm.2024.e03495.

- [2] F. Yuan, J. Pan, C.K.Y. Leung. Flexural Behaviors of ECC and Concrete/ECC Composite Beams Reinforced with Basalt Fiber-Reinforced Polymer. *J. Compos. Constr*, 17, 591-602, 2013, doi: 10.1061/(ASCE)CC.1943-5614.0000381.
- [3] B.-L. Lai, M.-Y. Zhang, Z.-P. Chen, J.R. Liew, Y.-Y. Zheng. Axial compressive behavior and design of semi-precast steel reinforced concrete composite columns with permanent ECC formwork, in: *Structures*, Elsevier, 2023, doi: 10.1016/j.istruc.2023.105130.
- [4] A. Hamoda, R.I. Shahin, M. Ahmed, A.A. Abadel, S.A. Yehia. Flexural behaviour of normal concrete circular beams strengthened with ECC and stainless steel tubes. *Magazine of Concrete Research*, 77, 171-188, 2025, doi: 10.1680/jmacr.24.00137.
- [5] J.-J. Zeng, J.-F. Lv, G. Lin, Y.-C. Guo, L.-J. Li. Compressive behavior of double-tube concrete columns with an outer square FRP tube and an inner circular high-strength steel tube. *Construction and Building Materials*, 184, 668-680, 2018, doi: 10.1016/j.conbuildmat.2018.07.034.
- [6] Z. Bing, F. Guisen, W. Wei, Z. Ningyuan, Z. Qianbiao, H. Xiamin. Behavior of elliptical GFRP-concrete-steel doubleskin tubular columns under axial compression. *Journal of Building Structures*, 40, 185-191, 2019.
- [7] H.-B. Ly. An interpretable machine learning-based measurement system for predicting the compressive strength of concrete at elevated temperatures. *Measurement*, 120273, 2025, doi: 10.1016/j.measurement.2025.120273.
- [8] TA. Nguyen. Versatile predictive modeling of CFDST columns using diverse FRP types: CATBOOST and advanced optimization techniques. *Structural Concrete*, 2026, doi: 10.1002/suco.70522.
- [9] F. Yu, H.F. Isleem, W.J. Almoghayer, R.I. Shahin, S.A. Yehia, M. Khishe, M.K. Elshaarawy. Predicting axial load capacity in elliptical fiber reinforced polymer concrete steel double skin columns using machine learning. *Scientific Reports*, 15, 2025, 12899, doi:10.1038/s41598-025-97258-y.
- [10] T. Ozbakkaloglu, B.L. Fanggi. Axial Compressive Behavior of FRP-Concrete-Steel Double-Skin Tubular Columns Made of Normal- and High-Strength Concrete. *J. Compos. Constr*, 18, 2014 04013027, doi: 10.1061/(ASCE)CC.1943-5614.0000401.
- [11] C. Cakiroglu, K. Islam, G. Bekdas, S. Kim, Z.W. Geem. Interpretable Machine Learning Algorithms to Predict the Axial Capacity of FRP-Reinforced Concrete Columns. *Materials*, 2022, 15, 2742, doi:10.3390/ma15082742.