

ĐÁNH GIÁ CÁC YẾU TỐ ẢNH HƯỞNG ĐẾN CƯỜNG ĐỘ NÉN VÀ MÔ ĐUN ĐÀN HỒI CỦA BÊ TÔNG CỐT LIỆU CAO SU BẰNG MÔ HÌNH HỌC MÁY

Trần Văn Quân^a, Nguyễn Hoàng Giang^{b,c,*}, Nguyễn Ngọc Tân^{b,c}

^aTrường Đại học Công nghệ Giao thông vận tải, Số 54, Triều Khúc, Thanh Xuân, Hà Nội, Việt Nam

^bKhoa Xây dựng Dân dụng và Công nghiệp, Trường Đại học Xây dựng Hà Nội,

55 đường Giải Phóng, quận Hai Bà Trưng, Hà Nội, Việt Nam

^cNghiên cứu và ứng dụng giải pháp đột phá cho tái chế phế thải (WASO), Trường Đại học Xây dựng Hà Nội,

55 đường Giải Phóng, quận Hai Bà Trưng, Hà Nội, Việt Nam

Nhận ngày 06/5/2023, Sửa xong 21/5/2023, Chấp nhận đăng 22/5/2023

Tóm tắt

Nghiên cứu này tập trung vào phát triển mô hình học máy để thiết kế thành phần bê tông sử dụng cốt liệu cao su. Mô hình tăng cường độ dốc (GB) dự đoán cường độ chịu nén và mô đun đàn hồi với độ chính xác cao trên tập dữ liệu kiểm chứng với hệ số xác định $R^2 = 0,9983$ và sai số căn quân phương RMSE = 0,8947 MPa khi dự đoán cường độ chịu nén và $R^2 = 0,9971$ và RMSE = 0,7473 GPa cho dự đoán mô đun đàn hồi. Các giá trị SHAP toàn cục và tuyệt đối được sử dụng để đánh giá độ ảnh hưởng của tám tham số đầu vào đến cường độ chịu nén và mô đun đàn hồi. Ảnh hưởng của các yếu tố đến cường độ chịu nén được sắp xếp theo thứ tự là: Vụn cao su cắt nhỏ (CR) > Vụn cao su (TR) > Xi măng (C) > Nước (W) > Muối silic (SF) > Cốt liệu thô (CA) > Cát (S) và với mô đun đàn hồi theo thứ tự là CR > TR > W > C > CA > S > SF. Hàm lượng nước, TR và CR làm giảm cường độ và mô đun đàn hồi của bê tông cao su khi tăng hàm lượng, trong khi các yếu tố còn lại đều có lợi cho việc cải thiện cường độ chịu nén và mô đun đàn hồi khi tăng hàm lượng.

Từ khóa: bê tông cao su; vụn cao su cắt nhỏ; vụn cao su, cường độ chịu nén; mô đun đàn hồi; mô hình học máy.

ASSESSING FACTORS AFFECTING COMPRESSIVE STRENGTH AND ELASTIC MODULUS OF CEMENTITIOUS CONCRETES MADE WITH RUBBER AGGREGATES USING MACHINE LEARNING MODELS

Abstract

The present study focuses on developing a machine learning model to design concrete compositions using rubber aggregates. The Gradient Boosting (GB) model predicts compressive strength and modulus of elasticity with high accuracy on the control dataset, having a coefficient of determination $R^2 = 0,9983$ and a root mean square error RMSE = 0,8947 MPa in predicting compressive strength and $R^2 = 0,9971$ and RMSE = 0,7473 GPa in predicting elastic modulus. Global and absolute SHAP values are used to evaluate the influence of the eight input variables on compressive strength and modulus of elasticity. The influence of factors on compressive strength is arranged in the order: Crumb rubber (CR) > Tire rubber (TR) > Cement (C) > Water (W) > Silica fume (SF) > Coarse aggregate (CA) > Sand (S) and with the elastic modulus in the order CR > TR > W > C > CA > S > SF. Water content, TR, and CR are to reduce the strength and elastic modulus of rubber concrete with increasing content, while the remaining factors are beneficial for improving compressive strength and elastic modulus when increasing content.

Keywords: rubber concrete; crumb rubber; tire rubber; compressive strength; modulus of elasticity; machine learning model.

[https://doi.org/10.31814/stce.huce2023-17\(2V\)-07](https://doi.org/10.31814/stce.huce2023-17(2V)-07) © 2023 Trường Đại học Xây dựng Hà Nội (ĐHXDHN)

*Tác giả đại diện. Địa chỉ e-mail: giangnh@huce.edu.vn (Giang, N. H.)

1. Mở đầu

Sự gia tăng chất thải cao su từ lốp xe và các sản phẩm làm bằng cao su tổng hợp đã trở thành một vấn đề cần được quan tâm, do loại chất thải này không thể tự phân hủy trong tự nhiên và gây ra mối đe dọa lớn cho môi trường. Việt Nam sản xuất hơn 400.000 tấn cao su phế thải mỗi năm. Khoảng 50% lốp xe đã qua sử dụng được chôn lấp, 40% lốp xe được đốt và chỉ 10% được tái chế [1]. Để góp phần giải quyết các vấn đề nghiêm trọng về môi trường do chất thải rắn gây ra, sử dụng các loại cốt liệu tái chế trong sản xuất vật liệu xây dựng là một giải pháp được khuyến khích [2, 3]. Việc sử dụng vụn cao su trong hỗn hợp nhựa đường đã chứng minh mang lại các hiệu quả kỹ thuật đối với kết cấu mặt đường nhựa [4]. Trong khi đó, việc sử dụng cao su tái chế trong hỗn hợp bê tông xi măng gặp phải nhiều thách thức, do sự không tương thích gây ra bởi thành phần hóa học và độ cứng của cao su [5]. Do đó, nhiều nghiên cứu đã được thực hiện trong những năm qua, nhằm cải thiện việc sử dụng cao su tái chế trong sản xuất bê tông xi măng [6–9]. Giải pháp này không những làm giảm thiểu lượng nguyên vật liệu sử dụng, mà còn giảm khối lượng lốp thải trực tiếp ra môi trường xung quanh. Do đó, bê tông sử dụng cốt liệu cao su tái chế (gọi tắt là bê tông cao su) là một loại vật liệu thân thiện với môi trường, được thực hiện bằng cách thêm cốt liệu cao su nghiền vào hỗn hợp bê tông, để đạt được những tính năng nổi trội, như tăng tính chịu nhiệt, cách điện và chống mài mòn, khả năng tỏa nhiệt thấp hơn và do đó tăng hiệu suất sử dụng trong cơ sở hạ tầng [10].

Trong số các tính chất cơ học của bê tông, thì cường độ chịu nén và mô đun đàn hồi là hai tính chất quan trọng thường xuyên được kiểm tra và sử dụng trong tính toán kết cấu. Phương pháp truyền thống để kiểm tra các tính chất cơ học này là thực hiện thí nghiệm nén trên các mẫu hình trụ hoặc hình lập phương được đúc trong phòng thí nghiệm hoặc tại hiện trường, tuân theo các tiêu chuẩn kỹ thuật hiện hành. Đối với bê tông cao su, quy trình thí nghiệm nén trên các mẫu thử được thực hiện tương tự như đối với bê tông thường. Cường độ chịu nén và mô đun đàn hồi của bê tông cao su bị quyết định bởi một số yếu tố, như: loại cao su sử dụng bao gồm vụn cao su và vụn cao su cắt nhỏ, hàm lượng xi măng, tỷ lệ cao su trong bê tông, hàm lượng nước, hàm lượng phụ gia sử dụng (ví dụ: muối silic), hàm lượng cốt liệu bao gồm cốt liệu thô và cát [11]. Theo nghiên cứu của Güneysi và cs. [11], vụn cao su cắt nhỏ là một vật liệu mịn với kích thước hạt gần giống với kích thước của cát và vụn cao su được sản xuất bằng cách băm nhỏ cơ học và chứa các kích thước hạt thô hơn. Sự phân cấp của vụn cao su được xác định dựa trên phương pháp ASTM C136. Tuy nhiên, phương pháp trên không thể xác định đường cong phân cấp cho vụn cao su như đối với cốt liệu bình thường, bởi vì kích thước hạt được kéo dài từ 10 đến 40 mm. Do vậy, sự hiểu biết đầy đủ các tính chất cơ học của bê tông cao su, bao gồm cường độ chịu nén và mô đun đàn hồi, vẫn còn hạn chế. Những nghiên cứu chuyên sâu là cần thiết nhằm phát hiện tất cả tiềm năng của cốt liệu cao su có thể ứng dụng trong xây dựng.

Mặc dù, cường độ chịu nén và mô đun đàn hồi của bê tông cao su thường được nghiên cứu bằng phương pháp thực nghiệm, nhưng phương pháp này có những hạn chế, chẳng hạn như chi phí cao, thời gian thí nghiệm kéo dài và các kết quả khác nhau bị ảnh hưởng bởi một số tham số, chẳng hạn như loại cao su được sử dụng, quy trình trộn và thời gian bảo dưỡng. Do đó, một phương pháp tin cậy và chính xác sử dụng các đặc tính của vật liệu cấu thành và tỷ lệ hỗn hợp để kiểm tra cường độ chịu nén và mô đun đàn hồi của bê tông cao su là cần thiết. Điều này có thể giúp giảm thời gian và chi phí liên quan đến thí nghiệm và cung cấp một cách hiệu quả để đánh giá cường độ chịu nén của bê tông cao su cho các ứng dụng thực tế. Gần đây, việc sử dụng các kỹ thuật học máy (ML) ngày càng quan tâm trong kỹ thuật dân dụng. Những cách tiếp cận này đã được áp dụng thành công trong nhiều ứng dụng, bao gồm cả kỹ thuật kết cấu, địa kỹ thuật, khoa học vật liệu [12–14]. Do khả năng của nó để đánh giá số lượng lớn của tổ hợp dữ liệu và tạo ra các dự đoán chính xác dẫn đến thiết kế và ra quyết định tốt hơn, ML đã cho thấy là một công cụ hữu ích cho các ứng dụng kỹ thuật dân dụng.

Trong việc ứng dụng mô hình học máy ML vào nghiên cứu bê tông sử dụng cốt liệu cao su, một số nghiên cứu đã được thực hiện trong thời gian gần đây. Sun và cs. [15] xây dựng mô hình Rừng cây ngẫu nhiên (Random Forest) kết hợp thuật toán tối ưu Ấng ten bọ cánh cứng (Beetle Antennae Search) dựa trên bộ dữ liệu có 138 mẫu và chín tham số đầu vào. Hadzima-Nyarko và cs. [16] sử dụng bộ dữ liệu có 457 mẫu và sáu tham số đầu vào để xây dựng mô hình Mạng thần kinh nhân tạo (Artificial Neural Network ANN) với ba lớp ẩn. Topçu and Sarıdemir [17] phát triển mô hình học máy ANN dựa trên bộ dữ liệu có 36 mẫu và bảy tham số đầu vào. Ly và cs. [18] xây dựng mô hình học máy ANN dựa trên 223 mẫu và mười hai tham số đầu vào. Các nghiên cứu nêu trên nghiên cứu vào việc xây dựng mô hình học máy cho dự đoán cường độ bê tông sử dụng cốt liệu cao su, ngoại trừ Gesoğlu và cs. [19] phát triển mô hình ANN và mô hình di truyền (Genetic programming) dựa trên tám tham số đầu vào và 70 mẫu để dự đoán cường độ chịu nén và mô đun đàn hồi của bê tông cao su. Thực tế cho thấy nhiều tham số cần tinh chỉnh dù các nghiên cứu nêu trên đã đưa ra được cấu tạo các lớp ẩn, tuy nhiên còn nhiều tham số của mô hình cần tinh chỉnh [20]. Do đó, việc tiếp cận sử dụng mô hình ANN cho các kỹ sư công trình trong thực tế triển khai sẽ tương đối khó khăn. Ngoài ra, nghiên cứu định lượng ảnh hưởng của hàm lượng vụn cao su và vụn cao su cắt nhỏ đến cường độ chịu nén và mô đun đàn hồi cũng chưa được thực hiện ở các nghiên cứu nêu trên.

Do vậy, mục tiêu chính của nghiên cứu là lựa chọn được một mô hình học máy có hiệu năng cao có sẵn trong các thư viện mở giúp các kỹ sư dễ dàng sử dụng trong việc thiết kế cấp phối bê tông sử dụng cốt liệu cao su. Ngoài ra, dựa trên độ tin cậy của mô hình học máy này, nghiên cứu định lượng ảnh hưởng của hàm lượng vụn cao su và vụn cao su cắt nhỏ, cũng như các yếu tố khác đến cường độ chịu nén và mô đun đàn hồi bằng kỹ thuật học máy lý thuyết trò chơi Shapley Additive Explanations (SHAP) được phát triển bởi Lundberg và cs. [21]. Trong các kỹ thuật học máy, thuật toán rừng cây ngẫu nhiên RF (Random Forest) [22] và thuật toán tăng cường độ dốc GB (Gradient Boosting) [23] là hai thuật toán rất phổ biến và sử dụng rộng rãi trong bài toán hồi quy.

Vì vậy, nghiên cứu này đề xuất sử dụng lựa chọn mô hình học máy có hiệu năng cao từ hai mô hình học máy sử dụng các giá trị siêu tham số mặc định có trong thư viện mã nguồn mở Sklearn viết bằng ngôn ngữ lập trình Python [20]. Việc sử dụng các kỹ thuật ML có thể giải quyết hiệu quả các hạn chế của các kỹ thuật thử nghiệm và cung cấp một công cụ dự đoán cường độ bê tông nhanh chóng và tiết kiệm chi phí. Một bộ dữ liệu được tổng hợp từ nghiên cứu trước đây của Güneşli và cs. [11], bao gồm 70 mẫu, với tám tham số đầu vào là hàm lượng xi măng (C), hàm lượng muối silic (SF), hàm lượng nước (W), hàm lượng cốt liệu thô (CA), hàm lượng cát (S), hàm lượng vụn cao su (TR), và hàm lượng vụn cao su cắt nhỏ (CR). Lựa chọn đánh giá hai mô hình RF và GB để lựa chọn một mô hình tốt nhất cho việc dự đoán cường độ chịu nén (F_c), mô đun đàn hồi (E_c) và đánh giá ảnh hưởng của các tham số đầu vào đến F_c và E_c của bê tông cao su bằng kỹ thuật SHAP. Để triển khai nghiên cứu, cơ sở dữ liệu sẽ được mô tả trong Phần 2, lý thuyết về các thuật toán và đánh giá hiệu suất mô hình học máy sẽ được giới thiệu ngắn gọn trong Phần 3, Phần 4 của nghiên cứu sẽ được dành để đánh giá và so sánh hiệu suất hai mô hình học máy RF và GB dựa trên kỹ thuật mô phỏng ngẫu nhiên Monte Carlo, mô hình có hiệu suất cao hơn sẽ được sử dụng cho dự đoán và đánh giá ảnh hưởng các yếu tố đầu vào đến F_c và E_c của bê tông cao su.

2. Mô tả cơ sở dữ liệu

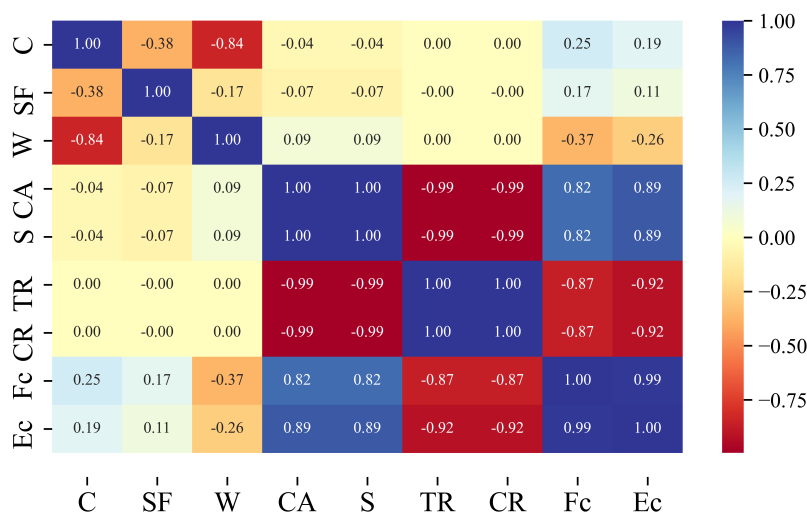
Cơ sở dữ liệu gồm 70 mẫu được thu thập từ nghiên cứu của Güneşli và cs. [11]. Tập dữ liệu này cho phép khảo sát tám tham số đầu vào là: hàm lượng xi măng (C), hàm lượng muối silic (SF), hàm lượng nước (W), hàm lượng cốt liệu thô (CA), hàm lượng cát (S), hàm lượng vụn cao su (TR), và hàm lượng vụn cao su cắt nhỏ (CR). Tham số đầu ra của việc xây dựng mô hình học máy bao gồm:

cường độ chịu nén F_c và mô đun đàn hồi E_c . Đối với mỗi tham số, miền giá trị được khảo sát thông qua giá trị trung bình (mean), giá trị lớn nhất (max), giá trị nhỏ nhất (min) và độ lệch chuẩn (Std), được thống kê trong Bảng 1. Hàm lượng vụn cao su và vụn cao su cắt nhỏ được tính theo phần trăm thể tích của cốt liệu thô và cốt liệu mịn.

Bảng 1. Thống kê miền giá trị của tham số đầu vào và đầu ra

Tham số	Ký hiệu	Đơn vị	Mean	Min	Max	Std
Xi măng	C	kg/m ³	360,20	280,00	450,00	53,65
Muội silic	SF	kg/m ³	40,00	0,00	90,00	29,15
Nước	W	kg/m ³	195,00	180,00	210,00	15,11
Cốt liệu thô	CA	kg/m ³	980,08	783,75	1076,00	86,92
Cát	S	kg/m ³	634,80	507,75	697,00	56,30
Vụn cao su	TR	% thể tích	0,08	0,00	0,25	0,08
Vụn cao su cắt nhỏ	CR	% thể tích	0,08	0,00	0,25	0,08
Cường độ chịu nén	F_c	MPa	44,13	7,10	85,70	22,26
Mô đun đàn hồi	E_c	GPa	27,38	6,10	47,50	12,07

Nghiên cứu của Güneyisi và cs. [11] sử dụng hàm lượng cốt liệu cao su thay thế đạt tới 50% thể tích tổng thể tích cốt liệu. Trong đó, thể tích cốt liệu cao su bao gồm thể tích vụn cao su và thể tích vụn cao su cắt nhỏ. Güneyisi và cs. [11] cũng đã nêu trong nghiên cứu với mọi hàm lượng cốt liệu cao su thay thế đều sử dụng 50% thể tích vụn cao su và 50% thể tích vụn cao su cắt nhỏ. Do vậy, khi thay thế 50% thể tích cốt liệu bê tông thông thường, 50% thể tích cốt liệu thô được thay thế bằng thể tích vụn cao su và 50% thể tích cát được thay thế bằng vụn cao su cắt nhỏ. Từ đó, tương ứng với hàm lượng cốt liệu cao su thay thế 50% thể tích cốt liệu sẽ tương ứng bao gồm là vụn cao su thay thế 25% thể tích cốt liệu và vụn cao su cắt nhỏ thay thế 25% (0,25) thể tích cốt liệu.



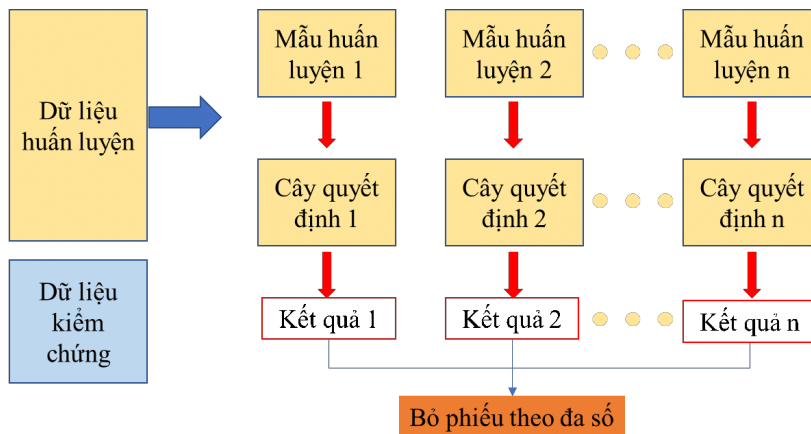
Hình 1. Tương quan tuyến tính giữa các tham số đầu vào và đầu ra được thể hiện bằng ma trận hệ số tương quan tuyến tính Pearson

Thông qua ma trận hệ số tương quan tuyến tính Pearson được biểu diễn trong Hình 1, có thể thấy bốn tham số cốt liệu bao gồm CA, S, TR và CR là các biến phụ thuộc về ý nghĩa số liệu. Các tham số đầu vào như xi măng C, nước W ảnh hưởng yếu hơn đến cường độ chịu nén F_c và mô đun đàn hồi E_c của bê tông so với các tham số cốt liệu. Tuy nhiên, tương quan trên mới chỉ dựa trên đánh giá tương quan tuyến tính nên chưa thực sự mô tả được sự ảnh hưởng của các yếu tố đến các tính chất cơ học (F_c và E_c) của bê tông sử dụng cốt liệu cao su. Nghiên cứu của Aas và cs. [24] cho thấy kỹ thuật SHAP cho phép đánh giá ảnh hưởng của các biến phụ thuộc. Do đó, để tăng tính ứng dụng thực tế trong thiết kế với các thành phần cốt liệu CA, S, TR và CR, cũng như đánh giá được ảnh hưởng của các biến này đến F_c và E_c bằng kỹ thuật SHAP, nghiên cứu đề xuất sử dụng tám tham số đầu vào được nêu trên.

3. Lý thuyết về phương pháp học máy

Trong mục này, lý thuyết của hai thuật toán Rừng cây ngẫu nhiên (Random Forest) và Tăng cường độ dốc (Gradient Boosting) sẽ được trình bày một cách ngắn gọn và súc tích.

3.1. Mô hình rừng cây ngẫu nhiên (RF)



Hình 2. Mô hình cơ chế hoạt động thuật toán Random Forest

Random forest còn gọi là “rừng cây ngẫu nhiên” là thuật toán có giám sát (supervised learning), có thể được sử dụng linh hoạt và dễ dàng cho cả phân loại và hồi quy được phát triển bởi Brieman [22]. Random forest tạo ra cây quyết định trên các mẫu dữ liệu được chọn ngẫu nhiên, được dự đoán từ mỗi cây và chọn giải pháp tốt nhất bằng cách bỏ phiếu. Thuật toán Random Forest gồm nhiều cây quyết định, mỗi cây đều có những yếu tố ngẫu nhiên. Do mỗi cây quyết định không dùng tất cả dữ liệu huấn luyện, cũng như các thuộc tính của dữ liệu để xây dựng cây nên mỗi cây có thể dự đoán không tốt, khi đó mỗi mô hình cây quyết định không bị overfitting mà có thể bị underfitting (mô hình có high bias). Tuy nhiên, kết quả cuối cùng của thuật toán Random Forest lại tổng hợp từ nhiều cây quyết định, thế nên thông tin từ các cây sẽ bổ sung cho nhau, dẫn đến mô hình có low bias và low variance, hay mô hình có kết quả dự đoán tốt. Thuật toán Random Forest hoạt động theo bốn bước:

Bước 1: Chọn các mẫu ngẫu nhiên từ tập dữ liệu đã cho: giả sử bộ dữ liệu có n dữ liệu (sample) và mỗi dữ liệu có m thuộc tính (feature). Lấy ngẫu nhiên n dữ liệu từ bộ n dữ liệu với kỹ thuật bootstrapping (random sampling with replacement). Một dữ liệu được lấy ra không bị mất đi trong

tập dữ liệu ban đầu, cứ thế tiếp tục lấy mẫu cho tới khi đủ n dữ liệu. Khi dùng kỹ thuật này thì tập n dữ liệu mới thu được có thể có những dữ liệu bị trùng nhau.

Bước 2: Xây dựng cây quyết định cho từng mẫu và nhận kết quả dự đoán từ mỗi quyết định của cây: sau khi lấy được n dữ liệu từ bước 1, chọn ngẫu nhiên ở k thuộc tính ($k < m$). Bộ dữ liệu mới thu được gồm n dữ liệu với mỗi dữ liệu có k thuộc tính. Dùng thuật toán Decision Tree để xây dựng cây quyết định với mỗi bộ dữ liệu mới. Do quá trình xây dựng mỗi cây quyết định đều có yếu tố ngẫu nhiên, nên kết quả là các cây quyết định trong thuật toán Random Forest có thể khác nhau.

Bước 3: Bỏ phiếu cho mỗi kết quả dự đoán: mỗi cây được xây dựng từ bước 2 dùng thuật toán Decision Tree trên tập dữ liệu khác nhau và dùng tập thuộc tính khác nhau sẽ cho một kết quả dự đoán (có thể giống nhau). Kết quả dự đoán của thuật toán Random Forest sẽ được tổng hợp từ các cây quyết định.

Bước 4: Chọn kết quả được dự đoán nhiều nhất là dự đoán cuối cùng (Hình 2).

3.2. Mô hình tăng cường độ dốc (GB)

Gradient Boosting là một trong số các thuật toán rất mạnh trong lĩnh vực máy học, được sử dụng rộng rãi cho các vấn đề hồi quy và phân loại [23]. Nó tạo ra một mô hình dự đoán dưới dạng mô hình kết hợp các mô hình dự đoán yếu, điển hình là các cây quyết định (Decision Tree). Nó xây dựng mô hình theo kiểu từng bước giống như các phương pháp boosting khác, và khái quát hóa chúng bằng cách cho phép tối ưu hóa một hàm mất mát (loss function). Mô hình Gradient Boosting có thể được xây dựng gồm 5 bước:

Bước 1: Xây dựng mô hình đầu tiên (model_1) với hàm mục tiêu mong muốn và tiến hành dự đoán cho bộ dữ liệu.

Bước 2: Tính toán lỗi (error) dự đoán của mô hình trước (model_1) và sử dụng lỗi này làm hàm mục tiêu dự đoán.

Bước 3: Xây dựng mô hình dự đoán về các lỗi này và tiến hành dự đoán.

Bước 4: Cập nhật lại kết quả dự đoán của mô hình đầu tiên (model_1) dựa trên kết quả dự đoán lỗi ở bước 3, ta được mô hình thứ hai (model_2).

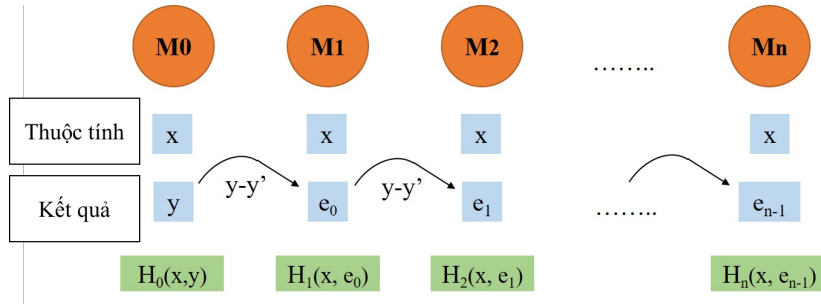
$$\begin{array}{|c|} \hline \text{Model_2} \\ \hline \text{(Orginal target)} \\ \hline \end{array} = \begin{array}{|c|} \hline \text{Model_1} \\ \hline \text{(Orginal target)} \\ \hline \end{array} + \begin{array}{|c|} \hline \text{Preticted} \\ \hline \text{Errors} \\ \hline \end{array}$$

Bước 5: Lặp lại các bước từ 2 đến 4: tính toán các lỗi dự đoán mới và đặt chúng làm hàm mục tiêu dự đoán. Xây dựng mô hình dự đoán về các lỗi mới này và tiến hành dự đoán. Kết quả dự đoán lỗi nhận được sẽ được cập nhật vào mô hình thứ hai (model_2) ta thu được mô hình ba (model_3). Quá trình này (từ bước 2 đến bước 4) sẽ tiếp tục lặp lại nhiều lần cho đến khi lỗi dự đoán tiến về 0 hoặc đạt đến điều kiện dừng (stopping criteria) là số lượng mô hình muốn xây dựng.

Để hiểu rõ hơn về các bước xây dựng mô hình Gradient Boosting, có thể lý giải theo ví dụ đơn giản ở Hình 3.

Đầu tiên ta xây dựng mô hình thứ nhất có thuộc tính (feature) x và mục tiêu dự đoán (target) y , gọi là mô hình $H_0(x, y)$ là hàm số của x và y . Mô hình tiếp theo sẽ được xây dựng dựa trên lỗi dự đoán của mô hình trước đó. Cứ như vậy quá trình này được lặp và các mô hình mới được xây dựng một cách tuần tự dựa trên lỗi của tất cả các mô hình trước nó. Có thể nói rằng mỗi mô hình trên là những hàm riêng biệt có biến độc lập là thuộc tính x (feature) và mục tiêu dự đoán là lỗi của mô hình trước nó.

Như vậy, mô hình đầu tiên $H_0(x, y)$, khi dự đoán cho kết quả với các lỗi dự đoán (error). Kết quả dự đoán này ta gọi là hàm $F_0(x)$. Tiếp theo, để xây dựng mô hình thứ hai ta dựa trên lỗi của mô hình



Hình 3. Các bước xây dựng n mô hình Gradient Boosting

trước và tiến hành dự đoán lỗi, kết quả dự đoán lỗi mới được thêm vào kết quả $F_0(x)$, lúc này ta có hàm mới là $F_1(x)$. Tương tự, ta sẽ xây dựng các mô hình tiếp theo cho đến khi đạt được n mô hình. Ở mỗi bước, chúng ta đang cố gắng tạo ra mô hình các lỗi, điều này giúp làm giảm thiểu lỗi ở các mô hình xây dựng sau. Lý tưởng nhất là ở mô hình cuối cùng có $e_n = 0$ (Hình 4).

$$\begin{aligned}
 F_0(x) &= H_0(x, y) + e_0 \\
 F_1(x) &= F_0(x) + H_1(x, e_0) + e_1 \\
 F_2(x) &= F_1(x) + H_2(x, e_1) + e_2 \\
 &\vdots \\
 F_n(x) &= F_{n-1}(x) + H_n(x, e_{n-1}) + e_n
 \end{aligned}$$

Hình 4. Hàm kết quả dự đoán ở mỗi mô hình tạo ra

3.3. Đánh giá hiệu suất của các mô hình học máy

Đánh giá mô hình học máy được thực hiện bằng các phép đo thống kê như: hệ số xác định (R^2) và sai số căn quân phương (RMSE). Nhìn chung, các tiêu chí này là các phương pháp phổ biến để định lượng hiệu suất của các mô hình học máy cho các bài toán hồi quy. Cụ thể hơn, sự khác biệt bình phương trung bình giữa các giá trị thực và giá trị ước tính xác định bởi RMSE, R^2 đánh giá mối tương quan giữa giá trị thực tế và giá trị ước tính. Về mặt định lượng, RMSE thấp hơn cho thấy hiệu suất tốt hơn của các mô hình. Ngược lại, R^2 cao hơn cho thấy hiệu suất tốt hơn của mô hình. RMSE và R^2 được thể hiện như sau:

$$R^2 = 1 - \frac{\sum_{i=1}^N (a_i - \hat{a}_i)^2}{\sum_{i=1}^N (a_i - \bar{a})^2} \quad (1)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (a_i - \hat{a}_i)^2} \quad (2)$$

trong đó: a_i là đầu ra thực tế, \hat{a}_i là đầu ra dự đoán, N là số lượng mẫu sử dụng.

Về cơ bản, RMSE càng tiến tới 0 thì mô hình có năng lực càng cao. Giá trị của R^2 nằm trong khoảng 0 đến 1, và giá trị của R^2 càng tiến tới 1 thì mô hình có độ chính xác càng cao.

4. Kết quả và thảo luận

4.1. Đánh giá hiệu suất của các mô hình học máy

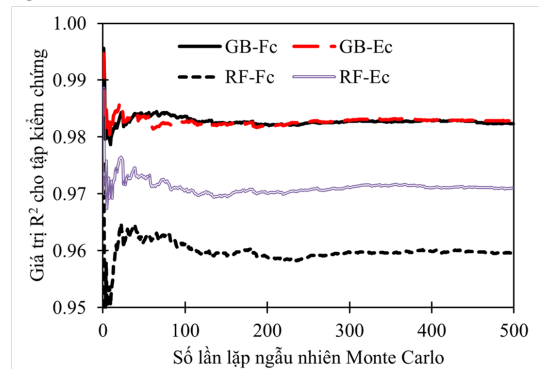
Bộ dữ liệu được chia 70%/30% tương ứng là 49 mẫu/21 mẫu cho tập dữ liệu huấn luyện và kiểm chứng. Tỷ lệ phân chia dữ liệu 70/30 cho bài toán học máy hồi quy là phù hợp và đã được chứng minh trong nghiên cứu của Nguyễn và cs. [25]. Ngoài ra, do có tính ngẫu nhiên trong việc phân chia tập dữ liệu do đó, kỹ thuật mô phỏng ngẫu nhiên Monte Carlo được sử dụng để đánh giá hiệu suất mô hình học máy RF và GB một cách toàn diện và chặt chẽ.

Trong việc lựa chọn và xác thực việc lựa chọn mô hình học máy, Hastie và cs. [26] và Zhang [27] đã đưa ra so sánh hai phương pháp xác thực là K-Fold Cross Validation (xác thực chéo) và mô phỏng lặp Monte Carlo (Monte Carlo cross validation). Ưu nhược điểm của mỗi phương pháp được thảo luận trong 2 nghiên cứu này. Tuy nhiên, khi số lượng mẫu của cơ sở dữ liệu mô phỏng Monte Carlo sẽ không yêu cầu thời gian tính toán quá lớn, cũng như mô hình K-Fold cross validation thường được sử dụng trong các quá trình xác thực khi tối ưu các siêu tham số của mô hình học máy. Do vậy, trong nghiên cứu này nhóm tác giả xin được sử dụng mô phỏng Monte Carlo cho việc xác thực mô hình.

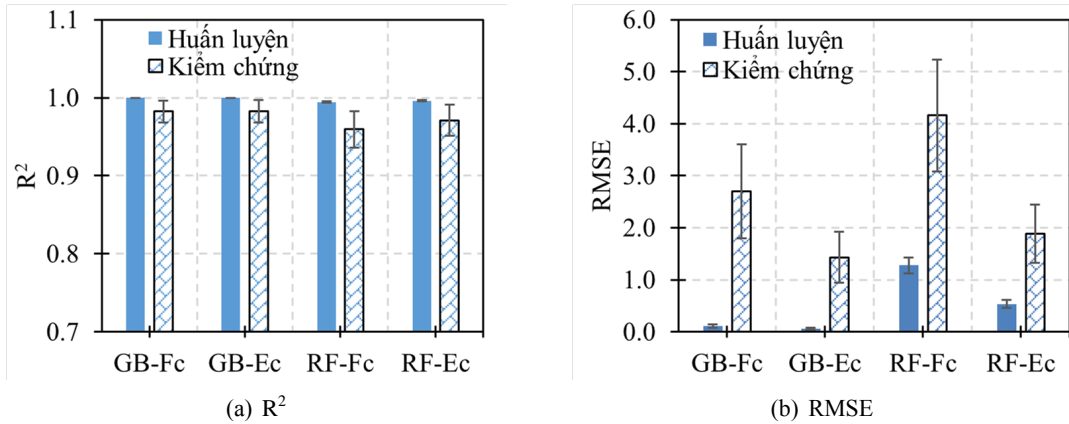
Hình 5 cho thấy giá trị R^2 cho tập dữ liệu kiểm chứng trở nên ổn định sau khoảng 200 lần chạy lặp cho cả hai mô hình học máy RF và GB trong việc dự đoán F_c và E_c . Do đó, việc lựa chọn số lần chạy lặp ngẫu nhiên là 500 đảm bảo tính xác thực chặt chẽ của các giá trị hiệu suất của mô hình học máy được đưa ra.

Biểu đồ so sánh hiệu suất của hai mô hình RF và GB sau 500 lần mô phỏng ngẫu nhiên Monte Carlo (a) R^2 và (b) RMSE được thể hiện trong Hình 6. Kết quả so sánh cho thấy mô hình GB cho hiệu suất cao hơn mô hình RF trong việc dự đoán cường độ chịu nén F_c và mô đun đàn hồi E_c . Hiệu suất cao hơn của mô hình GB được thể hiện qua hai giá trị trung bình của 500 lần mô phỏng Monte Carlo bao gồm giá trị R^2 và RMSE. Chi tiết các giá trị trung bình, giá trị lớn nhất, giá trị nhỏ nhất và độ lệch chuẩn của 500 giá trị R^2 và RMSE được tổng hợp trong Bảng 2. Trong đó, giá trị trung bình R^2 và RMSE của mô hình GB lần lượt là 0,9824 và 2,7029 MPa cho tập dữ liệu kiểm chứng trong việc dự đoán cường độ chịu nén F_c của bê tông cao su, trong khi mô hình RF có $R^2 = 0,9597$ và $RMSE = 4,1600$ MPa. Giá trị trung bình R^2 và RMSE của mô hình GB lần lượt là 0,9828 và 1,4327 GPa cho tập dữ liệu kiểm chứng trong việc dự đoán cường độ chịu nén F_c của bê tông cao su, trong khi mô hình RF có $R^2 = 0,9710$ và $RMSE = 1,8876$ GPa.

Do vậy, có thể thấy mô hình GB có hiệu suất tốt hơn hẳn mô hình RF trong việc xác định các tính chất cơ học, bao gồm cường độ chịu nén và mô đun đàn hồi của bê tông cao su. Mô hình GB sẽ được sử dụng trong việc dự đoán F_c , E_c và đánh giá các tham số đầu vào ảnh hưởng đến giá trị của F_c và E_c bằng các giá trị SHAP trong các phần tiếp theo.



Hình 5. Sự hội tụ của giá trị R^2 cho tập dữ liệu kiểm chứng sau 500 lần lặp ngẫu nhiên Monte Carlo



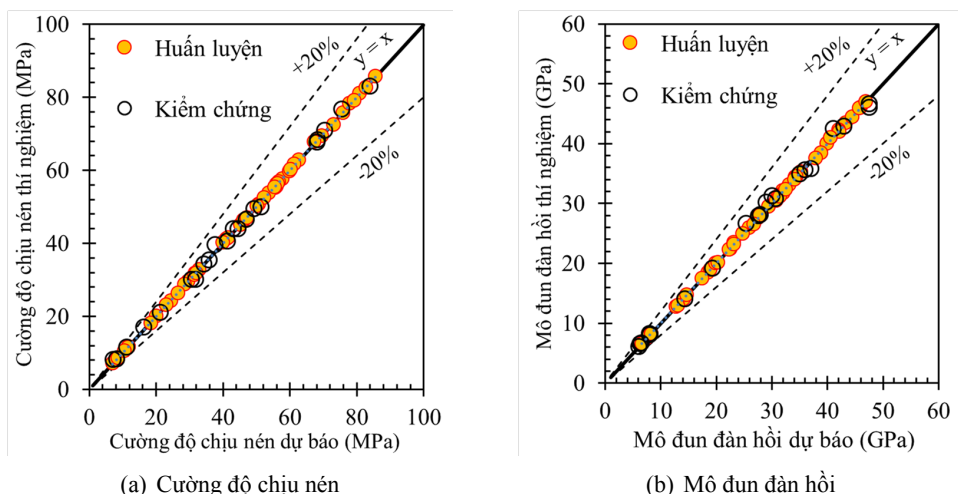
Hình 6. Biểu đồ so sánh hiệu suất của hai mô hình RF và GB sau 500 lần mô phỏng ngẫu nhiên Monte Carlo

Bảng 2. Tổng hợp thông kê giá trị hiệu suất của hai mô hình RF và GB

	Huấn luyện				Kiểm chứng			
	R^2				R^2			
	GB-F _c	GB-E _c	RF-F _c	RF-E _c	GB-F _c	GB-E _c	RF-F _c	RF-E _c
StD	0,0000	0,0000	0,0011	0,0009	0,0138	0,0142	0,0233	0,0199
Min	0,9997	0,9998	0,9895	0,9915	0,8780	0,8554	0,8264	0,8669
Average	0,9999	0,9999	0,9944	0,9959	0,9824	0,9828	0,9597	0,9710
Median	0,9999	0,9999	0,9944	0,9959	0,9857	0,9870	0,9658	0,9752
Max	1,0000	1,0000	0,9971	0,9981	0,9984	0,9971	0,9933	0,9963
RMSE								
StD	0,0369	0,0236	0,1514	0,0732	0,9029	0,4934	1,0785	0,5597
Min	0,1120	0,0625	1,2787	0,5375	0,8789	0,6127	1,7992	0,7604
Average	0,2028	0,1130	1,6381	0,7568	2,7029	1,4327	4,1600	1,8876
Median	0,2003	0,1118	1,6390	0,7568	2,5859	1,3478	3,9937	1,8123
Max	0,3390	0,1854	2,1032	1,0110	7,4896	4,1269	8,7489	4,0300

4.2. Dự đoán cường độ chịu nén của mô hình học máy

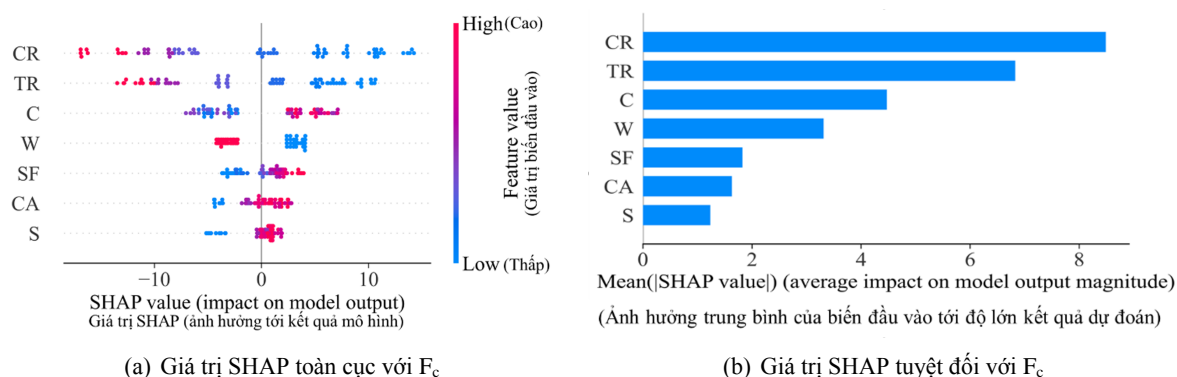
Hình 7(a) cho thấy tương quan giữa cường độ chịu nén thí nghiệm và dự đoán bằng mô hình học máy GB huấn luyện và kiểm chứng. Hình 7(b) cho thấy tương quan giữa mô đun đàn hồi thí nghiệm và dự đoán bằng mô hình học máy GB cho cả hai tập dữ liệu huấn luyện và kiểm chứng. Tương quan cho thấy việc dự đoán F_c và E_c của mô hình GB là rất tốt, khi các điểm dữ liệu tương quan nằm gần như trên đường thẳng $y = x$. Độ chính xác của mô hình GB trong việc dự đoán F_c và E_c được lượng hóa bằng các chỉ số $RMSE = 0,8947$ MPa và $R^2 = 0,9983$, trong đó giá trị R^2 của mô hình GB cao hơn kết quả của mô hình nơ ron nhân tạo được xây dựng bởi Gesoğlu và cs. [19] với $R^2 = 0,9956$ cho tập kiểm chứng cho dự đoán cường độ chịu nén F_c . Trong khi đó, với mô đun đàn hồi, chỉ số $R^2 = 0,9971$ và $RMSE = 0,7473$ GPa đối với mô hình GB cho tập dữ liệu kiểm chứng.



Hình 7. Tương quan so sánh (a) Cường độ chịu nén thí nghiệm và dự đoán, (b) Mô đun đàn hồi thí nghiệm và dự đoán

4.3. Định lượng ảnh hưởng của các yếu tố đầu vào đến giá trị cường độ chịu nén và mô đun đàn hồi của bê tông

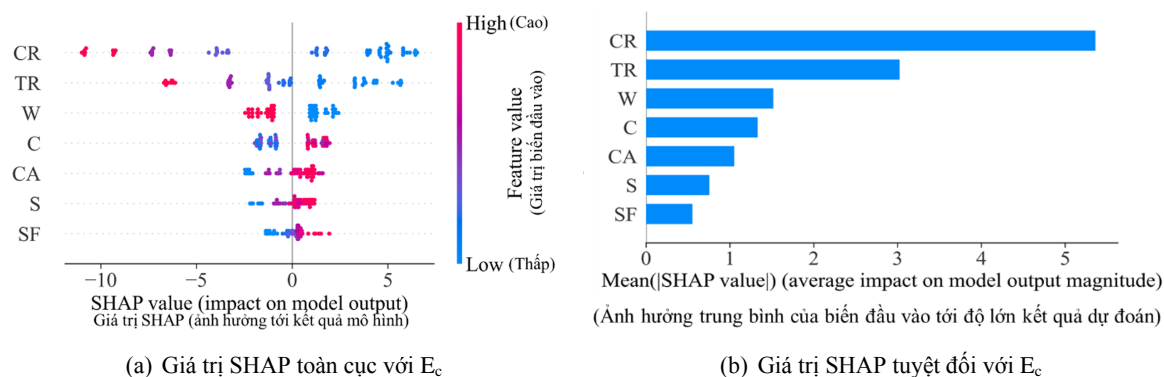
Dựa trên khả năng dự đoán các tính chất cơ học bao gồm cường độ chịu nén F_c và mô đun đàn hồi E_c , kỹ thuật SHAP được phát triển bởi Lundberg và cs. [21] sẽ được kết hợp với mô hình học máy GB để đánh giá ảnh hưởng của tám tham số đầu vào đến cường độ chịu nén F_c , cũng như mô đun đàn hồi E_c của bê tông cao su. Định lượng ảnh hưởng của từng yếu tố đầu vào đến cường độ chịu nén và mô đun đàn hồi của bê tông cao su được thể hiện lần lượt trong Hình 8 và Hình 9.



Hình 8. Định lượng ảnh hưởng của từng tham số đầu vào đến cường độ chịu nén của bê tông xi măng sử dụng cốt liệu cao su

Hình 8(a) và Hình 9(a) đều thống nhất cho thấy hàm lượng vụn cao su cắt nhỏ CR thay thế cốt liệu mịn có ảnh hưởng lớn nhất tới giá trị F_c và E_c . Tiếp theo sau là vụn cao su TR thay thế hàm lượng cốt liệu thô. Hàm lượng vụn cao su cắt nhỏ CR và vụn cao su TR sử dụng càng nhiều cường độ chịu nén và mô đun đàn hồi càng giảm. Hình 8(b) và Hình 9(b) thể hiện giá trị SHAP tuyệt đối của từng tham số đến giá trị dự đoán F_c và E_c . Có thể nhận thấy rằng, hàm lượng vụn cao su cắt nhỏ thể hiện

vai trò rất lớn trong đóng góp mô đun đàn hồi so với cường độ chịu nén khi mà SHAP tuyệt đối của CR và TR với E_c chênh lệch là 2, còn CR và TR với F_c là 1.



Hình 9. Định lượng ảnh hưởng của từng tham số đầu vào đến mô đun đàn hồi của bê tông xi măng sử dụng cốt liệu cao su

Muội silic có ảnh hưởng đến cường độ chịu nén nhiều hơn ảnh hưởng đến mô đun đàn hồi của bê tông cao su. Thông qua giá trị SHAP toàn cục và SHAP tuyệt đối, ảnh hưởng của các tham số đến cường độ chịu nén F_c được sắp xếp theo thứ tự là $CR > TR > C > W > SF > CA > S$ và với mô đun đàn hồi theo thứ tự là $CR > TR > W > C > CA > S > SF$. Trong các tham số trên, hàm lượng nước, TR và CR làm giảm cường độ chịu nén và mô đun đàn hồi của bê tông cao su khi tăng hàm lượng, các yếu tố còn lại đều có lợi cho việc cải thiện hai tính chất này khi tăng hàm lượng.

5. Kết luận và kiến nghị

Trong bài báo này, việc ứng dụng hai mô hình học máy thông dụng là rừng cây ngẫu nhiên RF và tăng cường độ dốc GB đã được xây dựng thành công trong việc đánh giá cường độ chịu nén và mô đun đàn hồi của bê tông xi măng sử dụng cốt liệu cao su tái chế dựa trên tám tham số đầu vào, bao gồm: hàm lượng xi măng (C), hàm lượng muội silic (SF), hàm lượng nước (W), hàm lượng cốt liệu thô (CA), hàm lượng cát (S), hàm lượng vụn cao su (TR), và hàm lượng vụn cao su cắt nhỏ (CR). Hiệu suất của hai mô hình GB và RF được đánh giá bằng các giá trị R^2 và RMSE cùng mô phỏng ngẫu nhiên Monte Carlo. Sau 500 lần mô phỏng ngẫu nhiên Monte Carlo, giá trị trung bình R^2 và RMSE của mô hình GB cao hơn trong việc dự đoán cường độ chịu nén và mô đun đàn hồi, $R^2 = 0,9824$ và $RMSE = 2,7029$ MPa cho tập dữ liệu kiểm chứng khi dự đoán F_c , $R^2 = 0,9828$ và $RMSE = 1,4327$ GPa khi dự đoán mô đun đàn hồi.

Muội silic có ảnh hưởng đến cường độ chịu nén hơn là ảnh hưởng đến mô đun đàn hồi của bê tông cao su. Thông qua giá trị SHAP toàn cục và SHAP tuyệt đối, ảnh hưởng của các yếu tố đến cường độ chịu nén F_c được sắp xếp theo thứ tự là $CR > TR > C > W > SF > CA > S$ và với mô đun đàn hồi theo thứ tự là $CR > TR > W > C > CA > S > SF$. Hàm lượng nước, TR và CR làm giảm cường độ và mô đun đàn hồi của bê tông cao su khi tăng hàm lượng, các yếu tố còn lại đều có lợi cho việc cải thiện cường độ và mô đun đàn hồi khi tăng hàm lượng.

Mô hình GB có thể sử dụng trong việc thiết kế thành phần bê tông cao su với hiệu quả cao trong miền giá trị của các tham số đầu vào được nêu trong bài. Để nâng cao tính ứng dụng và phổ quát rộng của mô hình học máy trong việc thiết kế cấp phối bê tông cao su, lượng mẫu dữ liệu lớn hơn cần phải được thu thập và so sánh với nhiều thuật toán học máy mới.

Tài liệu tham khảo

- [1] Tạp chí Môi trường. *Việt Nam có nhà máy tái chế rác thải từ lốp cao su đầu tiên*. Truy cập 14/5/2023.
- [2] Quang, N. T., Cường, T. V., Tân, N. N., Tân, N. H., Giang, N. H. (2021). *Nghiên cứu thực nghiệm ảnh hưởng của tỷ lệ sử dụng cốt liệu lớn tái chế đến sự phát triển cường độ nén và mô đun đàn hồi của bê tông theo thời gian*. *Tạp chí Khoa học Công nghệ Xây dựng (KHCN XD) - ĐHXDHN*, 15(1V):48–59.
- [3] Quang, N. T., Cuong, T. V., Tan, N. N., Tan, N. H., Kawamoto, K., Giang, N. H. (2022). *Experimental studies on behaviors of reinforced concrete column structures made of recycled aggregates under concentric loads*. *Journal of Science and Technology in Civil Engineering (STCE) - HUCE*.
- [4] Shu, X., Huang, B. (2014). *Recycling of waste tire rubber in asphalt and portland cement concrete: An overview*. *Construction and Building Materials*, 67:217–224.
- [5] Sofi, A. (2018). *Effect of waste tyre rubber on mechanical and durability properties of concrete – A review*. *Ain Shams Engineering Journal*, 9(4):2691–2700.
- [6] Youssf, O., Mills, J. E., Hassanli, R. (2016). *Assessment of the mechanical performance of crumb rubber concrete*. *Construction and Building Materials*, 125:175–183.
- [7] Ramdani, S., Guettala, A., Benmalek, M. L., Aguiar, J. B. (2019). *Physical and mechanical performance of concrete made with waste rubber aggregate, glass powder and silica sand powder*. *Journal of Building Engineering*, 21:302–311.
- [8] Siddika, A., Mamun, M. A. A., Alyousef, R., Amran, Y. H. M., Aslani, F., Alabduljabbar, H. (2019). *Properties and utilizations of waste tire rubber in concrete: A review*. *Construction and Building Materials*, 224:711–731.
- [9] Bu, C., Zhu, D., Liu, L., Lu, X., Sun, Y., Yu, L., OuYang, Y., Cao, X., Wang, F. (2022). *Research progress on rubber concrete properties: a review*. *Journal of Rubber Research*, 25(2):105–125.
- [10] Gerges, N. N., Issa, C. A., Fawaz, S. A. (2018). *Rubber concrete: Mechanical and dynamical properties*. *Case Studies in Construction Materials*, 9:e00184.
- [11] Güneyisi, E., Gesoğlu, M., Özturan, T. (2004). *Properties of rubberized concretes containing silica fume*. *Cement and Concrete Research*, 34(12):2309–2317.
- [12] Tran, V. Q., Do, H. Q. (2021). *Prediction of California Bearing Ratio (CBR) of Stabilized Expansive Soils with Agricultural and Industrial Waste Using Light Gradient Boosting Machine*. *Journal of Science and Transport Technology*, 1–9.
- [13] Tran, V. Q., Dang, V. Q., Ho, L. S. (2022). *Evaluating compressive strength of concrete made with recycled concrete aggregates using machine learning approach*. *Construction and Building Materials*, 323:126578.
- [14] Tran, V. Q. (2022). *Hybrid gradient boosting with meta-heuristic algorithms prediction of unconfined compressive strength of stabilized soil based on initial soil properties, mix design and effective compaction*. *Journal of Cleaner Production*, 355:131683.
- [15] Sun, Y., Li, G., Zhang, J., Qian, D. (2019). *Prediction of the Strength of Rubberized Concrete by an Evolved Random Forest Model*. *Advances in Civil Engineering*, 2019:1–7.
- [16] Hadzima-Nyarko, M., Nyarko, E. K., Ademović, N., Miličević, I., Šipoš, T. K. (2019). *Modelling the Influence of Waste Rubber on Compressive Strength of Concrete by Artificial Neural Networks*. *Materials*, 12(4):561.
- [17] Topçu, İ. B., Sarıdemir, M. (2008). *Prediction of rubberized concrete properties using artificial neural network and fuzzy logic*. *Construction and Building Materials*, 22(4):532–540.
- [18] Ly, H.-B., Nguyen, T.-A., Mai, H.-V. T., Tran, V. Q. (2021). *Development of deep neural network model to predict the compressive strength of rubber concrete*. *Construction and Building Materials*, 301:124081.
- [19] Gesoğlu, M., Güneyisi, E., Özturan, T., Özbay, E. (2009). *Modeling the mechanical properties of rubberized concretes by neural network and genetic programming*. *Materials and Structures*, 43(1-2):31–45.
- [20] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. et al. (2011). *Scikit-learn: Machine learning in Python*. *The Journal of Machine Learning Research*, 12:2825–2830.
- [21] Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J.,

- Bansal, N., Lee, S.-I. (2020). [From local explanations to global understanding with explainable AI for trees](#). *Nature Machine Intelligence*, 2(1):56–67.
- [22] Breiman, L. (2001). *Machine Learning*, 45(1):5–32.
- [23] Friedman, J. H. (2002). [Stochastic gradient boosting](#). *Computational Statistics & Data Analysis*, 38(4): 367–378.
- [24] Aas, K., Jullum, M., Løland, A. (2021). [Explaining individual predictions when features are dependent: More accurate approximations to Shapley values](#). *Artificial Intelligence*, 298:103502.
- [25] Nguyen, Q. H., Ly, H.-B., Ho, L. S., Al-Ansari, N., Le, H. V., Tran, V. Q., Prakash, I., Pham, B. T. (2021). [Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil](#). *Mathematical Problems in Engineering*, 2021:1–15.
- [26] Hastie, T., Tibshirani, R., Friedman, J. (2009). *The Elements of Statistical Learning*. Springer New York.
- [27] Zhang, P. (1993). [Model Selection Via Multifold Cross Validation](#). *The Annals of Statistics*, 21(1): 299–313.