



Article info

Type of article:

Original research paper

DOI:

<https://doi.org/10.58845/jstt.utt.2025.vn.5.2.56-68>

***Corresponding author:**

Email address:

vanmth@utt.edu.vn

Received: 06/03/2025

Received in Revised Form:
29/04/2025

Accepted: 24/05/2025

Predicting the Marshall flow of asphalts concrete using Gradient Boosting model

Hai-Van Thi Mai*, Van Huynh Pham

University of Transport Technology, Hanoi 100000, Vietnam

Abstract: Dense-graded asphalt concrete is one of the primary materials used to construct road pavement structures. The design and quality assessment of dense-graded asphalt concrete mixtures require consideration of multiple criteria, among which Marshall flow (MF) is a main indicator. This study compiled a dataset consisting of 90 dense-graded asphalt concrete samples with nine input variables and one output variable, MF. Based on this dataset, three machine learning models were developed and evaluated for predicting MF, including Gradient Boosting, Light Gradient Boosting, and Decision Tree. The predictive performance of the models was compared over 20 simulation runs, leading to the identification of the Gradient Boosting model as the optimal approach, with a high predictive accuracy ($R^2_{\text{test}} = 0.91$). The study presents a guide for using the machine learning model to predict the MF of dense-graded asphalt concrete with high accuracy, providing significant time and cost savings. Additionally, SHAP diagram was used to predict the influence of each input variable on the MF of dense-graded asphalt concrete.

Keywords: Machine learning; Marshall flow; Dense-graded Asphalt concrete; Gradient Boosting.



Thông tin bài viết
Dạng bài viết:
Bài báo nghiên cứu

DOI:
<https://doi.org/10.58845/jstt.utt.2025.vn.5.2.56-68>

Tác giả liên hệ:
Địa chỉ Email:
vanmth@utt.edu.vn

Ngày nộp bài: 06/03/2025
Ngày nộp bài sửa: 29/04/2025
Ngày chấp nhận: 24/05/2025

Dự báo độ dẻo Marshall của hỗn hợp bê tông nhựa chặt bằng mô hình tăng cường độ dốc

Mai Thị Hải Vân*, Phạm Văn Huỳnh
Trường Đại học Công nghệ Giao thông vận tải, số 54 Triều Khúc, Thanh Xuân, Hà Nội, Việt Nam

Tóm tắt: Bê tông nhựa chặt (BTNC) là một trong số các vật liệu chủ yếu dùng xây dựng kết cấu áo đường ô tô. Để thiết kế thành phần và đánh giá chất lượng hỗn hợp BTNC cần dựa vào nhiều chỉ tiêu. Trong đó, độ dẻo Marshall (MF) là một chỉ tiêu quan trọng. Nghiên cứu đã thu thập bộ dữ liệu gồm 90 mẫu BTNC với 9 biến đầu vào và một biến đầu ra là MF. Trên cơ sở đó đã tiến hành xây dựng, phát triển 3 mô hình (MH) máy học: Tăng cường độ dốc - Gradient Boosting (GB), Tăng cường độ dốc nhẹ - Light Gradient Boosting (LGB) và Cây quyết định- Decision Tree (DT) để dự báo MF. Hiệu suất dự báo (DB) của các MH sau 20 lần mô phỏng được so sánh, từ đó đề xuất được một MH tối ưu trong dự báo MF của BTNC (mô hình GB), với có khả năng dự báo MF của BTNC rất cao ($R^2_{kiểm\ tra} = 0,91$). Nghiên cứu đã xây dựng được một hướng dẫn sử dụng MH để DB độ dẻo Marshall của BTNC một cách chính xác, tiết kiệm thời gian và chi phí. Ngoài ra, biểu đồ SHAP cũng được xây dựng để DB mức độ ảnh hưởng của các biến đầu vào đối với MF của BTNC.

Từ khóa: Máy học; Độ dẻo Marshall; Bê tông nhựa chặt; Tăng cường độ dốc.

1. Đặt vấn đề

Trong xây dựng đường ô tô, bê tông nhựa (BTN) là vật liệu chủ yếu được sử dụng cho kết cấu áo đường. Mặt đường BTN có nhiều ưu điểm như chịu lực tốt, mặt đường bằng phẳng, công nghệ thi công nhanh và giá thành và chi phí duy tu bảo dưỡng hợp lý. Tuy nhiên, BTN là hỗn hợp vật liệu đàn nhớt dẻo, khi nhiệt độ cao BTN có tính dẻo, sức chống cắt thấp, biến dạng không phục hồi tăng dễ gây lượn sóng và hằn lún vệt bánh xe mặt đường. Trái lại, khi nhiệt độ thấp, BTN trở nên giòn, chịu mỗi kém gây nứt và bong tróc mặt đường. Để hạn chế các hư hỏng trên trong điều kiện khai thác cụ thể, cần thiết kế hỗn

hợp BTN với các thành phần và các chỉ tiêu cơ lý phù hợp.

Ở Việt Nam, thí nghiệm Marshall được dùng phổ biến để xác định hàm lượng chất kết dính tối ưu của hỗn hợp BTNC [1]. Trong đó, độ dẻo Marshall (MF) là một chỉ tiêu cần thiết để thiết kế BTNC. Tuy nhiên, để xác định MF bằng phương pháp thí nghiệm cần nhiều thời gian và chi phí cho quá trình đúc mẫu và làm thí nghiệm. Hơn nữa, khi kết quả thí nghiệm không như mong muốn hoặc cần thay đổi thiết kế thành phần BTNC đều phải tiến hành lại thí nghiệm, gây lãng phí. Do đó, cần tiếp cận tới một phương pháp mới để DB các chỉ tiêu cơ lý của vật liệu, nhằm giảm

bớt các hạn chế của phương pháp thí nghiệm truyền thống.

Hiện nay, công nghệ trí tuệ nhân tạo phát triển mạnh mẽ và việc áp dụng các mô hình máy học vào giải quyết các bài toán kỹ thuật đã được nghiên cứu [2], [3], [4], [5], đặc biệt trong việc DB các tính chất cơ lý của vật liệu [6] [7] [8] [9]. Phương pháp máy học tỏ ra phù hợp trong việc này do có thể xác định nhanh kết quả, chính xác, tiết kiệm thời gian và tiền bạc mà không cần tiến hành làm thí nghiệm. Đặc biệt, phương pháp này có thể giải quyết được các bài toán vật liệu phức tạp, bài toán có nhiều biến mà tương quan giữa các biến không được biểu thị rõ ràng bởi các công thức toán học. Đối với việc dự báo MF của BTNC đã có nghiên cứu của Baldo và cộng sự [10] sử dụng mô hình Artificial Neural Network (ANN); Gul và các cộng sự sử dụng các mô hình

ANN, Adaptive Neuro-Fuzzy Inference System (ANFIS), Multi Expression Programming (MEP) [11]...Tuy nhiên, chưa có nghiên về mô hình GB và các nghiên cứu cũng chưa đưa ra được hướng dẫn cụ thể cách sử dụng mô hình để dự báo MF của BTNC.

Trong bài báo này, ba MH máy học GB, LGB và DT đã được xây dựng và đào tạo để dự báo MF của BTNC. Để đề xuất được một MH tối ưu nhất, khả năng DB của ba MH được so sánh sau 20 lần mô phỏng. Các lần mô phỏng khác nhau sẽ sử dụng các bộ dữ liệu khác nhau được tạo ra từ sự xáo trộn ngẫu nhiên. Bài báo cũng xây dựng được một hướng dẫn sử dụng MH (gọi là biểu đồ GUI), nhằm cụ thể các thao tác để DB được giá trị MF khi có sự thay đổi của các biến đầu vào.

2. Dữ liệu

Bảng 1. Thông tin của các biến

Tên biến	Ký hiệu	Đơn vị	Số lượng	Trung bình	Độ lệch chuẩn	Nhỏ nhất	25%	50%	75%	Lớn nhất
Độ kim lún	V1	0,1 mm	90	51,33	9,01	45,00	45,00	45,00	64,00	64,00
Nhiệt độ hóa mềm	V2	0 °C	90	67,73	15,74	45,60	45,60	78,80	78,80	78,80
Lọt sàng 2,36 mm	V3	%	90	33,13	2,77	27,85	32,03	33,22	33,22	39,26
Lọt sàng 4,75 mm	V4	%	90	47,86	5,56	39,93	44,41	46,27	46,27	64,93
Lọt sàng 9,5 mm	V5	%	90	70,13	3,18	61,58	68,64	69,66	69,66	79,87
Hàm lượng nhựa	V6	%	90	4,75	0,50	3,80	4,42	4,81	5,00	5,80
Độ rỗng dư	V7	%	90	5,49	2,29	0,40	3,80	5,50	7,43	9,70
Độ rỗng cốt liệu	V8	%	90	16,00	2,25	12,10	14,33	15,45	17,98	20,40
Độ rỗng lấp đầy nhựa	V9	%	90	66,57	11,96	48,00	57,40	65,35	73,78	97,40
Độ dẻo Marshall	MF	mm	90	3,47	0,84	1,90	3,00	3,40	3,90	5,40

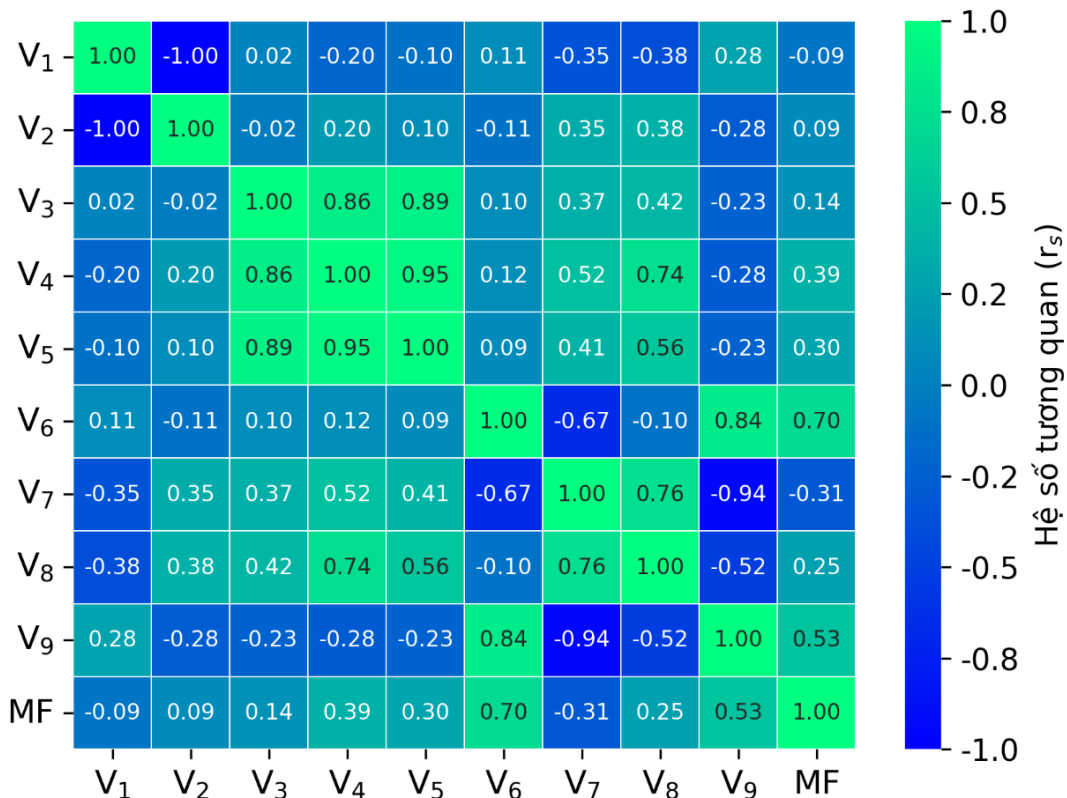
Để xây dựng các MH máy học, bài báo tận dụng các dữ liệu có sẵn từ các thí nghiệm về xác định MF của BTNC đã được công bố trên tạp chí quốc tế [10]. Bộ dữ liệu gồm 90 mẫu thí nghiệm BTNC, thu thập từ 3 dự án quốc tế khác nhau với hàm lượng nhựa và cấp phối cốt liệu khác nhau. Có 2 loại nhựa đường được sử dụng là nhựa đường thông thường và nhựa đường biến tính SBS (Styrene-Butadiene-Styrene). Các mẫu thí nghiệm được thu thập từ cả hai nguồn: trong phòng thí nghiệm và trong nhà máy sản xuất bê

tông nhựa để đảm bảo tính đa dạng của bộ dữ liệu. Bộ dữ liệu gồm 9 biến đầu vào là các yếu tố ảnh hưởng đến MF của BTN (ký hiệu là V₁÷V₉), và một biến đầu ra chính là Độ dẻo Marshall (ký hiệu là MF). Thông tin của các biến sử dụng trong nghiên cứu được thể hiện trong Bảng 1.

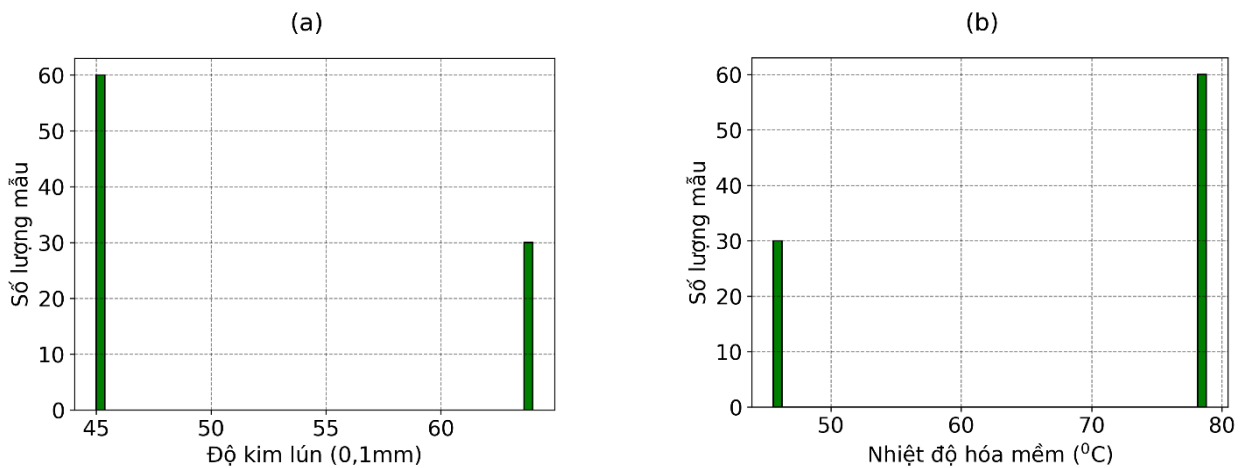
Quan hệ các biến với nhau được thể hiện trên Hình 1 nhờ hệ số tương quan r_s. Dựa vào giá trị của r_s có các mức độ quan hệ là: Tương quan không đáng kể; Tương quan yếu; Tương quan vừa phải; Tương quan mạnh; Tương quan rất

mạnh [12]. Kết quả trên Hình 1 cho thấy đa số quan hệ giữa các biến với nhau là tương quan yếu, cũng có tương quan giữa một vài biến như V_1 (Độ kim lún) với V_2 (Nhiệt độ hóa mềm) và V_3, V_4, V_5 (Hàm lượng lọt sàng của cốt liệu) hoặc tương quan giữa độ rỗng dư (V_7) và độ rỗng lấp đầy nhựa (V_9) có tương quan mạnh. Mặc dù vậy, nghiên cứu sẽ giữ lại tất cả các biến này để nghiên cứu vì mỗi biến đều mang một ý nghĩa riêng biệt, đều là các chỉ tiêu cần thiết khi thiết kế hỗn hợp BTN. Hơn nữa, mô hình mà tác giả sử

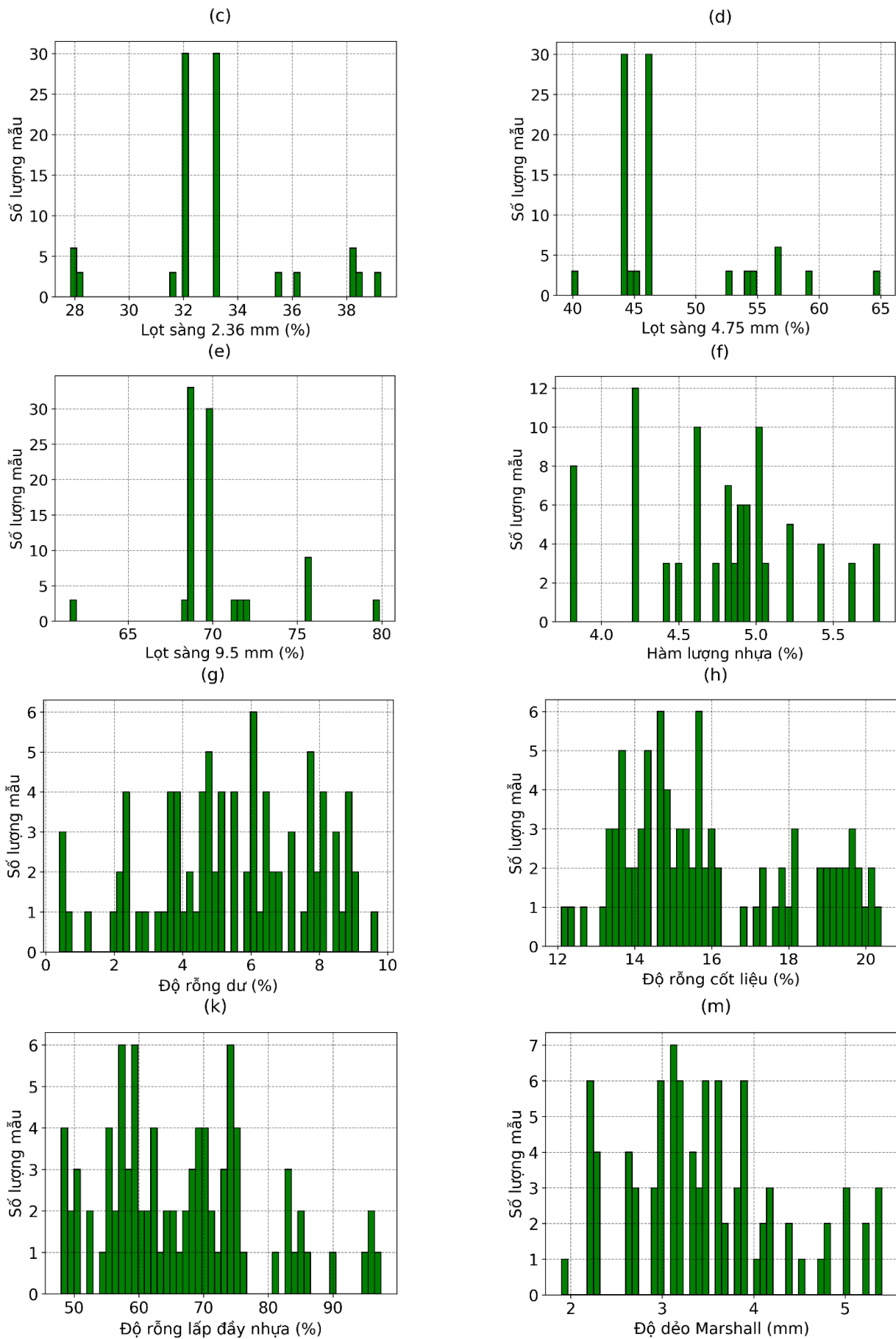
dụng trong nghiên cứu là mô hình GB. Đây là mô hình tập hợp nhiều cây quyết định nhỏ (weak learners), mà cây quyết định không bị ảnh hưởng nhiều bởi mối quan hệ tuyến tính giữa các biến đầu vào. Nếu hai biến gần như trùng nhau, GB sẽ ưu tiên chọn một biến hiệu quả hơn để chia trước, còn biến kia ít được dùng hơn (hoặc dùng ở cây khác) hay nói một cách khác mô hình GB có thể tự học cách bỏ qua biến thừa mà không nhất thiết phải loại bỏ biến. Vì vậy, nghiên cứu sẽ giữ lại cả 9 biến đầu vào.



Hình 1. Quan hệ giữa các biến



Hình 2. Phân bố dữ liệu của các biến



Hình 2. (tiếp)

Để mô tả chi tiết bộ dữ liệu, tần suất phân bố của các biến được trình bày trên Hình 2. Quan sát cho thấy, độ kim lún (V_1) chỉ tập trung ở hai giá trị 45, 64 (0,1mm) và Nhiệt độ hóa mềm (V_2) là 45 và 78°C. Cốt liệu lọt sàng 2,36mm (V_3) thay đổi trong khoảng từ 28 đến 40% nhưng chủ yếu tập trung ở hai giá trị 32 và 33%. Tương tự lượng cốt liệu lọt sàng 4,75mm (V_4) dao động trong khoảng từ 40÷65%; hàm lượng cốt liệu lọt sàng 9,5 mm (V_5) dao động trong khoảng từ 61÷80%. hàm lượng nhựa (V_6) tập trung trong khoảng từ 3,8÷5,8%, Độ rỗng dư (V_7) chủ yếu dao động trong khoảng từ 0,4÷9,7% trong khi độ rỗng cốt liệu (V_8) nhỏ nhất là 12%, cao nhất khoảng 20%. Độ rỗng lấp đầy nhựa (V_9) có giá trị trong khoảng từ 48÷97% nhưng đã số mẫu tập trung trong khoảng từ 50÷75%.

3. Cơ sở lý thuyết

3.1. Các MH máy học

Bài báo sử dụng 3 MH máy học là Tăng cường độ dốc (GB); Tăng cường độ dốc nhẹ (LGB) và Cây quyết định (DT).

Mô hình GB

GB là một mô hình máy học giải quyết cả bài toán hồi quy và phân loại [13]. Thuật toán sử dụng sự giảm dần độ dốc để giảm bớt sai số và đưa ra dự đoán chính xác hơn. Mục đích là xây dựng một tập hợp những người học cơ sở theo trình tự để dự đoán các sai số còn lại của mô hình trước đó. Để cung cấp dự đoán chính xác hơn, việc học được thực hiện liên tục, phù hợp với các mô hình mới. Thuật toán GB tạo ra các mô hình có độ chính xác cao, vượt trội hơn các thuật toán khác.

Mô hình LGB

LGB là mô hình tăng cường độ dốc, dùng thuật toán học cây. Một số ưu điểm của thuật toán này là tốc độ đào tạo nhanh hơn, hiệu quả cao hơn, sử dụng bộ nhớ tối ưu, độ chính xác cao, có thể xử lý dữ liệu lớn và hỗ trợ học song song. Cụ thể là, thay vì quét tuần tự qua từng mức độ sâu của cây như các thuật toán truyền thống, LGB áp dụng chiến lược tăng trưởng theo chiều sâu, lựa chọn nhánh có khả năng giảm lỗi lớn nhất để tiếp

tục phân chia, nhờ đó cải thiện hiệu quả học tập. Hơn nữa, LGB còn hỗ trợ xử lý dữ liệu lớn, có thể tự động hóa tối ưu nhiều tham số. MH này có thể áp dụng trong các bài toán xếp hạng, phân loại, hồi quy [14].

Mô hình DT

DT là một mô hình máy học phổ biến được đề xuất bởi Quinlan [15]. Mô hình DT sử dụng quy tắc để xác định các vùng có các biến đầu ra đồng nhất với các biến đầu vào mà trên đó xác định một hằng số phù hợp với từng vùng. Ưu điểm của mô hình này là cung cấp thông tin trực quan, dễ hình dung, mạnh mẽ để khai thác các hiệu ứng phi tuyến và tương tác giữa các biến phụ thuộc và độc lập, không yêu cầu giả định toán học giữa đầu ra và các biến đầu vào, có khả năng xử lý các giá trị bị thiếu và các giá trị ngoại lai. Một số nhược điểm của mô hình này là khó mô hình hóa các hàm, cấu trúc cây nhạy cảm với dữ liệu mẫu vì chỉ cần thay đổi nhỏ trong dữ liệu huấn luyện cũng có thể cho kết quả rất khác biệt, có phương sai cao và độ lệch thấp.

3.2. Đánh giá khả năng dự báo của MH

Để kiểm tra khả năng DB của các MH máy học, bài báo sử dụng 4 hệ số gồm: hệ số xác định (R^2), sai số tuyệt đối trung bình (MAE) và sai số bình phương trung bình (RMSE) và sai số phần trăm tuyệt đối trung bình (MAPE). Giá trị của R^2 nằm trong khoảng từ 0 đến 1. Khi R^2 càng gần tới 1, sai số MAE, RMSE, MAPE càng gần tới 0 thì MH có khả năng dự báo càng chính xác và ngược lại. Công thức xác định các hệ số:

$$R^2 = 1 - \frac{\sum_{i=1}^k (v_i - \bar{v}_i)^2}{\sum_{i=1}^k (v_i - \bar{v})^2} \quad (1)$$

$$RMSE = \sqrt{\frac{1}{k} \sum_{i=1}^k (v_i - \bar{v}_i)^2} \quad (2)$$

$$MAE = \frac{1}{k} \sum_{i=1}^k (v_i - \bar{v}_i) \quad (3)$$

$$MAPE = \frac{1}{k} \sum_{i=1}^k \frac{(v_i - \bar{v}_i)}{k \cdot v_i} \quad (4)$$

với v_i là Độ dẻo Marshall xác định từ thí nghiệm

(giá trị thực tế), \bar{v}_i là Độ dẻo Marshall được dự báo bởi MH học máy, k là số dữ liệu, \bar{v} giá trị trung bình của v_i .

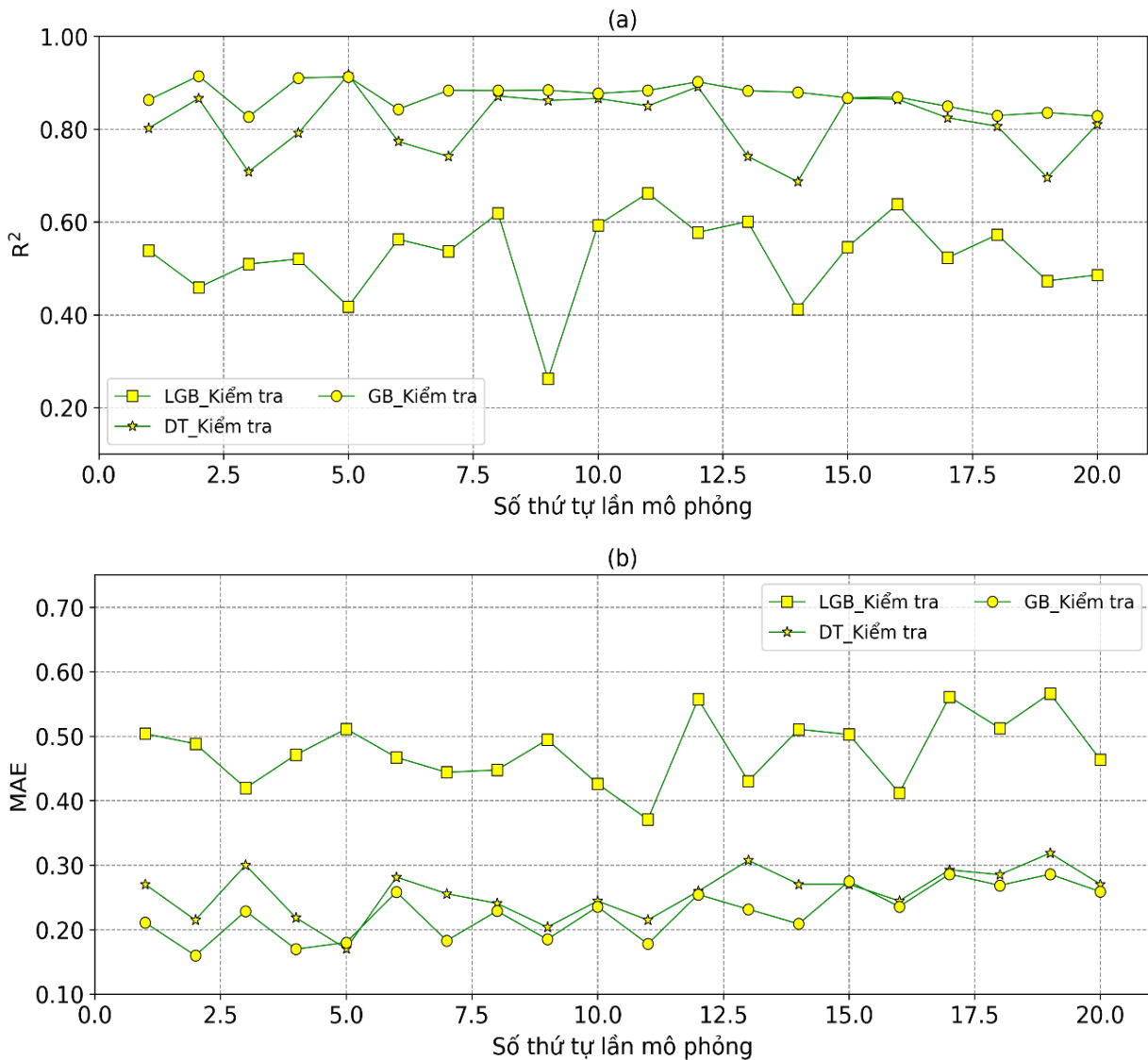
4. Kết quả và thảo luận

4.1. Phân tích đề xuất MH máy học tối ưu

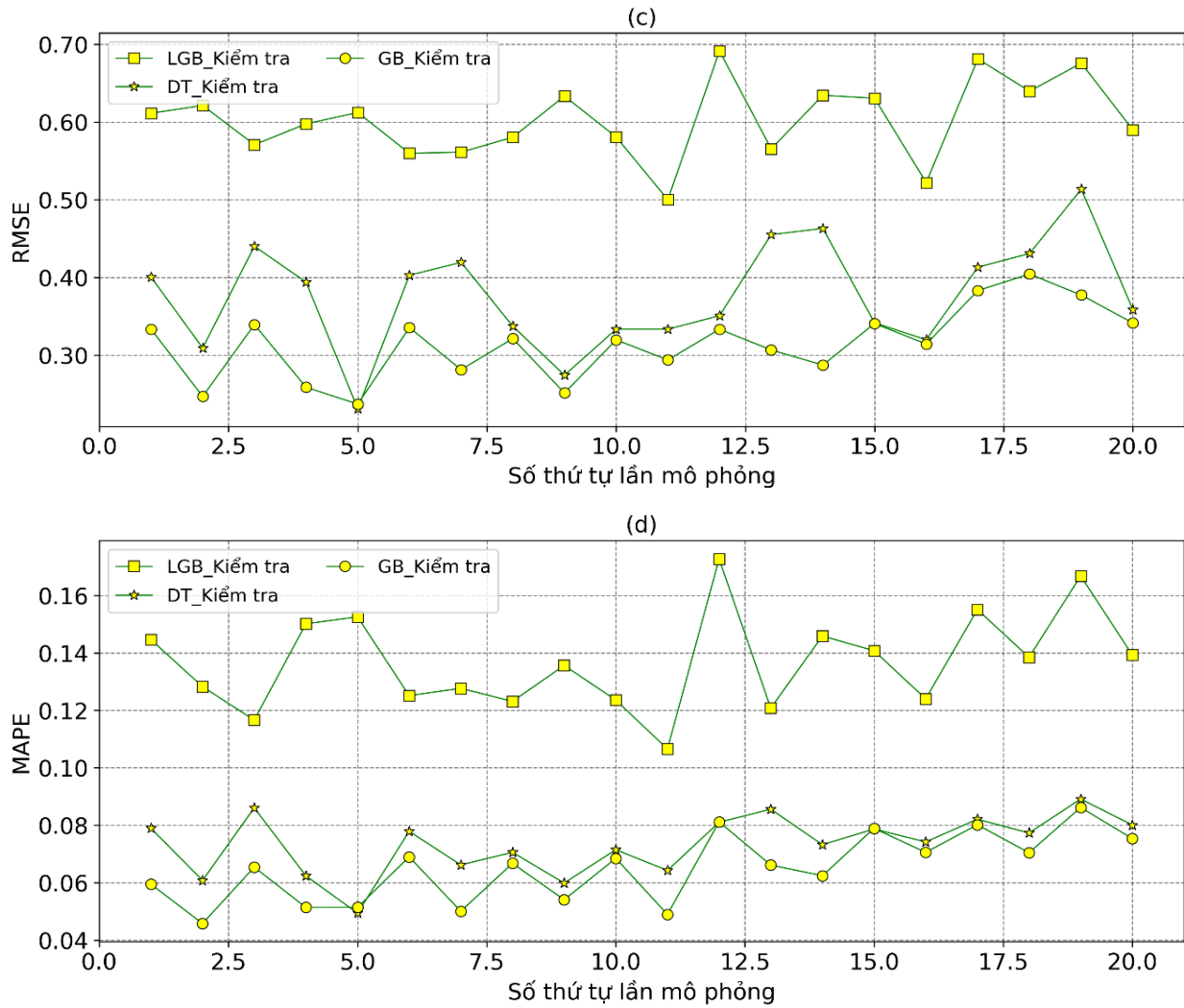
Nghiên cứu đã xây dựng, đào tạo 3 mô hình GB, LGB, DT và so sánh kết quả dự báo của 3 MH sau 20 lần mô phỏng để lựa chọn MH tối ưu nhất cho DB độ dẻo Marshall của BTNC. Bộ dữ liệu được thành 2 phần: phần dữ liệu đào tạo (70%) và phần dữ liệu kiểm tra (30%) một cách ngẫu nhiên. Phần dữ liệu đào tạo được sử dụng để huấn luyện và xác thực các MH máy học còn phần dữ liệu kiểm tra được dùng để kiểm tra MH. Giai đoạn đào tạo (GDĐT) được thực hiện trước,

giai đoạn kiểm tra (GDKT) thực hiện sau. Trong GDĐT và giai đoạn xác thực phần dữ liệu kiểm tra sẽ được giấu kín, MH không được biết. Do đó, trong 3 giai đoạn thì kết quả dự báo của mô hình ở GDKT có tính khách quan và quan trọng nhất. Đây là lý do, nghiên cứu sẽ căn cứ vào kết quả dự báo ở GDKT để đề xuất MH tối ưu.

Kết quả dự báo của 3 MH sau 20 lần mô phỏng được thể mô tả ở Hình 3. Cơ sở dữ liệu được xáo trộn ngẫu nhiên để tạo ra 20 bộ với các dữ liệu đào tạo và kiểm tra khác nhau (vẫn đảm bảo tỷ lệ 70/30). Mỗi lần mô phỏng các mô hình sẽ dùng 1 bộ dữ liệu được xáo trộn khác nhau này giúp làm tăng tính khách quan trong kết quả dự báo của các MH.



Hình 3. Kết quả dự báo của ba mô hình sau 20 lần mô phỏng ở giai đoạn kiểm tra



Hình 3. (tiếp)

Bảng 2. Kết quả dự báo chi tiết của ba MH ở GDKT

Mô hình	Giá trị	RMSE	MAE	R ²	MAPE
GB	TB	0,3153	0,2261	0,8712	0,0651
	Std	0,0447	0,0391	0,0276	0,0116
	Min	0,2369	0,1599	0,8266	0,0458
	Max	0,4043	0,2860	0,9146	0,0862
LGB	TB	0,6030	0,4780	0,5254	0,1369
	Std	0,0491	0,0513	0,0895	0,0166
	Min	0,5000	0,3707	0,2621	0,1065
	Max	0,6915	0,5662	0,6622	0,1727
DT	TB	0,3760	0,2567	0,8118	0,0734
	Std	0,0675	0,0371	0,0665	0,0100
	Min	0,2309	0,1704	0,6868	0,0494
	Max	0,5135	0,3185	0,9172	0,0891

(Std: Độ lệch chuẩn, TB: Trung bình)

Khả năng dự báo của các mô hình được thể hiện ở 4 chỉ tiêu R² (Hình 3a), MAE (Hình 3b), RMSE (Hình 3c) và MAPE (Hình 3d). Như đã

trình bày, một mô hình có R² càng lớn và MAE, RMSE, MAPE càng thấp thì khả năng dự báo càng đúng. Quan sát Hình 3, kết quả đánh giá

theo cả 4 tiêu chí đều cho thấy GB là mô hình có khả năng dự báo chính xác và ổn định nhất, tiếp theo đến mô hình DT và cuối cùng là LGB. Cụ thể giá trị $R^2_{\text{trung bình}}$ của mô hình GB, DT, LGB ở giai đoạn kiểm tra là 0,87; 0,81 và 0,52 và độ lệch chuẩn lần lượt là 0,02; 0,06; 0,08 (Bảng 2). Bên cạnh đó, so sánh sau 20 lần mô phỏng cho thấy GB ở lần mô phỏng thứ 2 là MH có khả năng dự báo tốt nhất thể hiện ở giá trị R^2 cao nhất ($R^2_{\text{max}} = 0,91$), các sai số MAE, RMSE, MAPE tương đối nhỏ so với các lần mô phỏng còn lại ($MAE_{\text{min}} = 0,15$; $RMSE_{\text{min}} = 0,23$; $MAPE_{\text{min}} = 0,04$). Với kết quả này mô hình GB ở lần mô phỏng thứ 2 cũng cho thấy có khả năng dự báo MF với độ chính xác cao. Vì vậy, đề xuất lựa chọn mô hình GB ở lần mô phỏng thứ 2 là mô hình tối ưu nhất trong 3 mô hình nghiên cứu để dự báo MF của BTNC.

4.2. Kết quả tiêu biểu khi dự báo độ dẻo Marshall bằng mô hình GB

Tiếp theo, kết quả tiêu biểu khi DB độ dẻo Marshall của BTNC bằng mô hình GB (là kết quả ứng với lần mô phỏng thứ 2) được trình bày thông qua biểu đồ hồi quy (Hình 4) và hình vẽ so sánh sự khác nhau giữa giá trị MF dự báo với giá trị MF thực nghiệm (Hình 5).

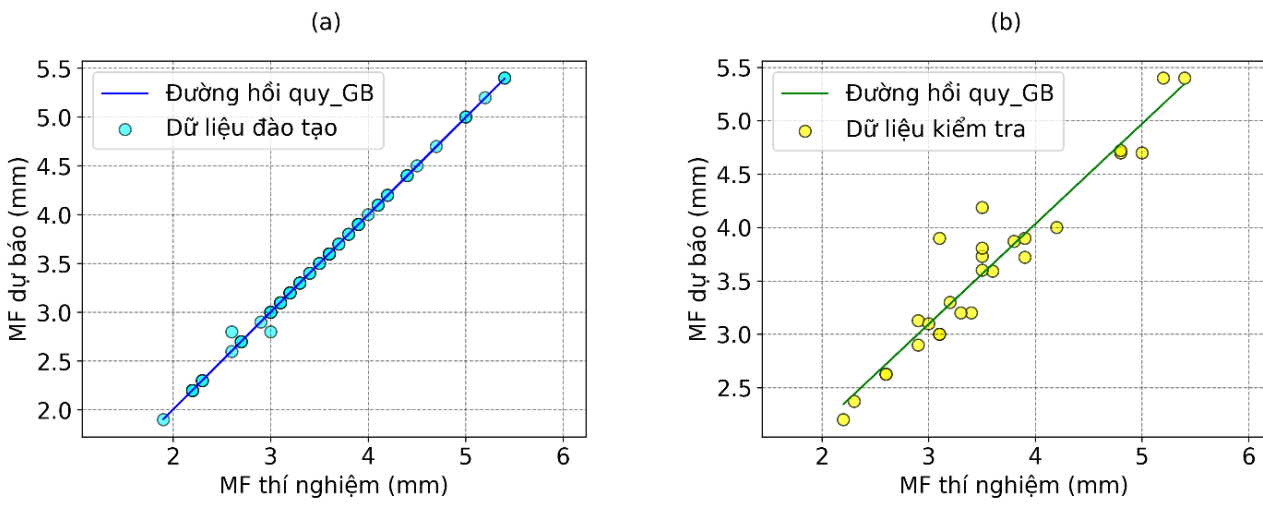
Biểu đồ hồi quy (Hình 4) có trục hoành thể hiện MF thực tế do thí nghiệm, trục tung thể hiện MF do mô hình GB dự báo. Khi hai giá trị này trùng nhau, các điểm sẽ nằm trên đường hồi quy (lý tưởng), các điểm càng nằm sát đường hồi quy tức là MH dự báo càng chính xác và ngược lại.

Hình 4a cho thấy GB là MH có kết quả ở GĐĐT rất chính xác khi đa số các điểm đều ở gần đường hồi quy ($R^2_{\text{đào tạo}} = 0,99$). Hình 4b thể hiện khả năng dự báo của mô hình GB ở GĐKT cũng rất cao, phần lớn các điểm nằm gần với đường hồi quy ($R^2_{\text{kiểm tra}} = 0,91$). Điều này khẳng định khả năng DB của mô hình GB đề xuất rất chính xác.

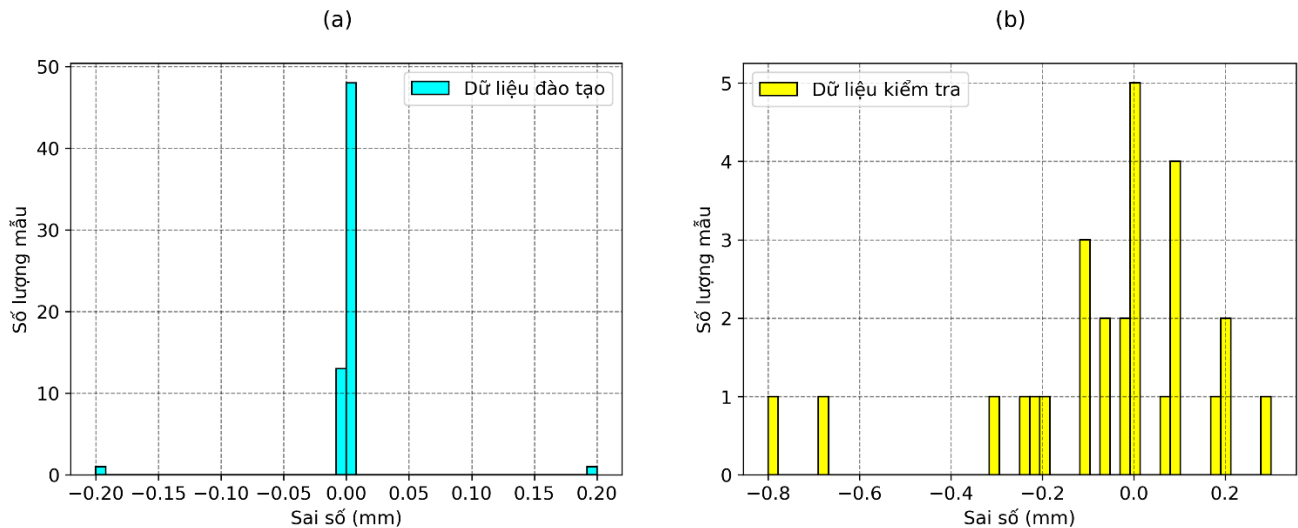
Ngoài ra, khả năng DB của mô hình GB còn được mô tả bằng hình vẽ phân bố sai số giữa giá trị MF do mô hình dự báo với giá trị MF thực nghiệm ở GĐĐT (Hình 5a) và ở GĐKT (Hình 5b). Ở GĐĐT, đa số các sai số rất nhỏ (xấp xỉ bằng 0), chỉ có 1 vài mẫu có sai số là -0,2 và 0,2 mm. Ở GĐKT, sai số nằm trong khoảng $[-0,8; 0,25]$ mm, nhưng tập trung trong khoảng từ $0,3; 0,2$ là chủ yếu. Sai số càng nhỏ tức là giá trị MF do mô hình dự báo càng gần đúng với giá trị MF thực nghiệm. Như vậy mô hình GB tối ưu có khả năng dự báo với độ chính xác cao.

Kết quả tiêu biểu khi sử dụng mô hình GB để dự báo MF được trình bày trong Bảng 3.

So sánh với một số kết quả nghiên cứu sử dụng mô hình máy học để dự báo MF đã có được thể hiện trong Bảng 4. Kết quả so sánh cho thấy so với nghiên cứu của Baldo và cộng sự. [10], mô hình GB của nghiên cứu này cho kết quả dự báo tốt hơn nhưng so với kết quả nghiên cứu của Gul và cộng sự. [11] thì không tốt hơn. Tuy nhiên, trong kết quả nghiên cứu của Gul và cộng sự. [11], các mô hình đều cho kết quả $R^2_{\text{kiểm tra}}$ cao hơn $R^2_{\text{đào tạo}}$. Điều này có vẻ bất thường.



Hình 4. Biểu đồ hồi quy ở giai đoạn: a) GĐĐT; b) GĐKT



Hình 5. Biểu đồ sai số giữa giá trị MF dự báo với giá trị MF thực nghiệm trong a) GĐĐT; b) GĐKT

Bảng 3. Kết quả dùng mô hình GB để DB độ dẻo Marshall

Giai đoạn	RMSE (mm)	MAE (mm)	R ²	MAPE (mm)
Đào tạo	0,03563	0,00637	0,99814	0,00228
Kiểm tra	0,24678	0,15987	0,91462	0,04578

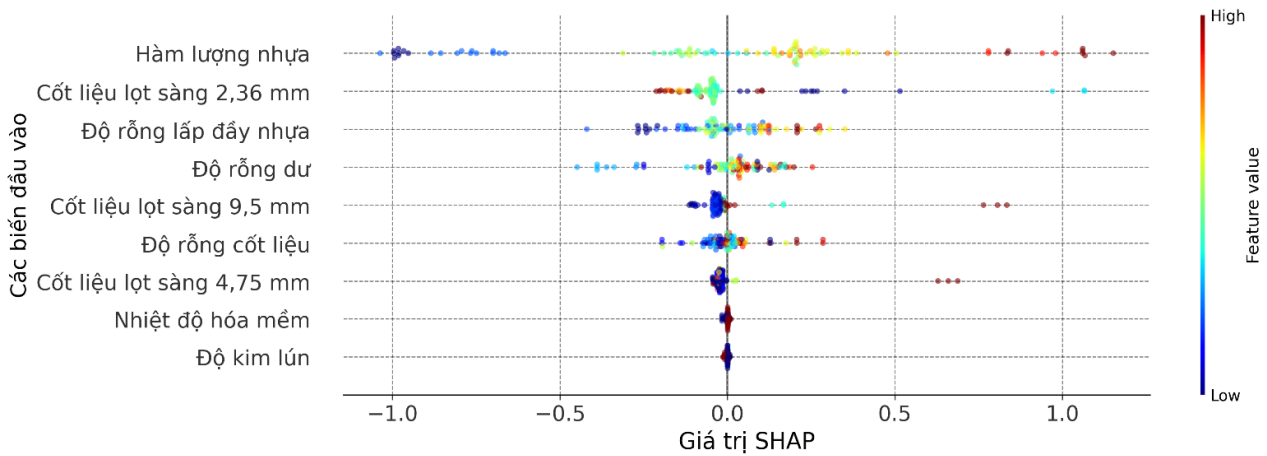
Bảng 4. So sánh kết quả với một số nghiên cứu đã có

Nghiên cứu	Mô hình sử dụng	Hiệu suất dự báo
Baldo và cộng sự. [10]	Artificial Neural Network (ANN)	R _{đào tạo} = 0,9383 (R ² _{đào tạo} = 0,88) R _{kiểm tra} = 0,9367 (R ² _{kiểm tra} = 0,88)
	+ Artificial Neural Network (ANN)	+ANN R _{đào tạo} = 0,961 (R ² _{đào tạo} = 0,92) R _{kiểm tra} = 0,973 (R ² _{kiểm tra} = 0,94)
Gul và cộng sự. [11]	+Adaptive Neuro-Fuzzy Inference System (ANFIS)	+ ANFIS: R _{đào tạo} = 0,972 (R ² _{đào tạo} = 0,94) R _{kiểm tra} = 0,980 (R ² _{kiểm tra} = 0,96)
	+ Multi Expression Programming (MEP)	+ MEP: R _{đào tạo} = 0,972 (R ² _{đào tạo} = 0,94) R _{kiểm tra} = 0,980 (R ² _{kiểm tra} = 0,96)
Nghiên cứu này	GB	R ² _{đào tạo} = 0,998 R ² _{kiểm tra} = 0,914

4.3. Phân tích mức độ ảnh hưởng của các yếu tố đến độ dẻo Marshall của Bê tông nhựa chặt (BTNC)

Trong phần này, nhờ vào MH máy học GB, tầm quan trọng và ảnh hưởng của các biến đầu vào đến MF được DB thông qua biểu đồ giá trị SHAP (Hình 6). Biểu đồ giá trị SHAP cho thấy trong 9 biến đầu vào, Hàm lượng nhựa có ảnh hưởng mạnh nhất đến MF. Trong đó, ảnh hưởng của Hàm lượng nhựa đến MF là ảnh hưởng tích

cực, có nghĩa là trong phạm vi nghiên cứu, khi Hàm lượng nhựa tăng lên thì MF tăng lên và ngược lại. Điều này đúng với với kết quả của nghiên cứu [16]. Ngoài ra, các biến đầu vào như Độ rỗng lấp đầy nhựa, Độ rỗng cốt liệu cũng có ảnh hưởng tích cực đến MF. Ngược lại, lượng cốt liệu lọt sàng 2,36 mm lại có ảnh hưởng tiêu cực, Độ rỗng dư có ảnh hưởng phức tạp đến MF. Biểu đồ SHAP còn cho thấy các yếu tố có ảnh hưởng ít nhất đến MF là Độ kim lún và Nhiệt độ hóa mềm.



Hình 6. Biểu đồ SHAP

4.4. Hướng dẫn sử dụng mô hình GB để DB độ dẻo Marshall của BTN

Độ kim lún (0,1mm)	47.52	↕
45	<input type="range" value="47.52"/>	64
Điểm hóa mềm (\$^0\$C)	52.73	↕
45.6	<input type="range" value="52.73"/>	78.8
Lượng cốt liệu lọt sàng 2.36 mm (%)	29.07	↕
27.85	<input type="range" value="29.07"/>	39.26
Lượng cốt liệu lọt sàng 4.75 mm (%)	46.23	↕
39.93	<input type="range" value="46.23"/>	64.93
Lượng cốt liệu lọt sàng 9.5 mm (%)	63.78	↕
61.58	<input type="range" value="63.78"/>	79.87
Hàm lượng nhựa (%)	4.23	↕
3.8	<input type="range" value="4.23"/>	5.8
Độ rỗng dư (%)	1.4	↕
0.4	<input type="range" value="1.4"/>	9.7
Độ rỗng cốt liệu (%)	13.41	↕
12.1	<input type="range" value="13.41"/>	20.4
Độ rỗng lấp đầy nhựa (%)	58.4	↕
48	<input type="range" value="58.4"/>	97.4

Độ dẻo Marshall (mm)

Hình 7. Hướng dẫn sử dụng mô hình GB để dự báo MF của BTNC

Nội dung phần này sẽ cụ thể hóa việc dự báo MF của BTN bằng mô hình GB thông qua biểu đồ GUI (Graphical User Interface) và 2 file đính kèm MFBTN.pkl và Van33b.ipynb

Hướng dẫn các bước sử dụng mô hình để dự báo MF như sau:

- Đăng nhập tài khoản Google cá nhân
- Mở thư mục Google Drive cá nhân
- Copy hai tệp MFBTN.pkl và Van33b.ipynb vào thư mục Google Drive, trong My Drive
- Nhấp chuột phải vào tệp Van33b.ipynb và chọn “Mở bằng Google Colab”. Nếu thành công, cửa sổ code sẽ mở ra.
- Click vào 2 khối code để thực hiện quá trình mô phỏng.
- Kết quả chạy 2 khối code này sẽ xuất hiện giao diện như Hình 7.
- Dựa vào các thanh có tên từng biến để thay đổi giá trị của 9 biến đầu vào (Hình 7).
- Bấm “Submit” để nhận kết quả độ dẻo Marshall (Hình 7).

[https://drive.google.com/file/d/1t8qOPdpNcHICUll5uovN_0PQwG_bV2XQ/view?usp=drive link](https://drive.google.com/file/d/1t8qOPdpNcHICUll5uovN_0PQwG_bV2XQ/view?usp=drive_link)

5. Kết luận

Bài báo đã đề xuất được một MH máy học GB để dự báo MF của BTNC, bằng cách xây dựng và so sánh khả năng DB của 3 MH (GB, LGB, DT) sau 20 lần mô phỏng dựa trên 4 chỉ tiêu R^2 , MAE, RMSE, MAPE. Mô hình GB có khả năng dự báo MF của BTNC chính xác và ổn định nhất, kết quả tối ưu thể hiện ở lần mô phỏng thứ 2 với giá trị $R^2_{kiểm tra} = 0,91$; $RMSE_{kiểm tra} = 0,24$ (mm), $MAE_{kiểm tra} = 0,15$ (mm), $MAPE_{kiểm tra} = 0,04$ (mm). Bên cạnh đó, mô hình GB tiêu biểu cũng dự báo được mức độ ảnh hưởng của các yếu tố đến MF của BTNC thông qua biểu đồ giá trị SHAP. Cuối cùng, nghiên cứu cũng đề xuất được một giao diện sử dụng mô hình GB để dự báo MF với các thao tác rất dễ dàng, giúp nhanh chóng dự báo chính xác được MF của BTNC.

Tuy nhiên, cơ sở dữ liệu trong nghiên cứu còn hạn chế về số lượng, để tiếp tục hướng nghiên cứu tiềm năng này, trong các nghiên cứu tiếp theo có thể phát triển cơ sở dữ liệu nhiều hơn nữa dựa trên các công bố quốc tế có độ tin cậy

cao, hoặc thu thập ở các công trình nghiên cứu thực nghiệm trong nước. Ngoài ra, việc tiếp tục nghiên cứu sử dụng các công cụ trí tuệ nhân tạo khác cũng như sử dụng kỹ thuật tinh chỉnh siêu tham số của MH để cải thiện hiệu suất DB độ dẻo Marshall của hỗn hợp BTN cũng rất hữu ích, có thể phát triển ở các nghiên cứu tiếp theo.

Tài liệu tham khảo

- [1] Bộ Khoa học và Công nghệ. (2011). Tiêu chuẩn quốc gia, TCVN 8860-1:2011, Bê tông nhựa - Phương pháp thử - Phần 1: Xác định độ ổn định, độ dẻo Marshall.
- [2] M. Xue, C. Zhu. (2009). A study and application on machine learning of artificial intelligence. *2009 International Joint Conference on Artificial Intelligence*, IEEE, 2009, pp. 272-274.
- [3] D. Mhlanga. (2021). Financial inclusion in emerging economies: The application of machine learning and artificial intelligence in credit risk assessment. *International Journal of Financial Studies*, 9(3), 39.
- [4] S. Guikema. (2020). Artificial intelligence for natural hazards risk analysis: Potential, challenges, and research needs. *Risk Analysis*, 40(6), 1117-1123.
- [5] A. Alimadadi, S. Aryal, I. Manandhar, P.B. Munroe, B. Joe, X. Cheng. (2020). Artificial intelligence and machine learning to fight COVID-19. *Physiol Genomics*, 52(4), 200-202.
- [6] E. Ozgan. (2011). Artificial neural network based modelling of the Marshall Stability of asphalt concrete. *Expert Systems with Applications*, 38(5), 6025-6030.
- [7] B.N. Phung, T.-H. Le, H.-V.T. Mai, T.-A. Nguyen, H.-B. Ly. (2023). Advancing basalt fiber asphalt concrete design: A novel approach using gradient boosting and metaheuristic algorithms. *Case Studies in Construction Materials*, 19, e02528.
- [8] I. Prakash, T.-N. Phan, H.-V. T. Mai. (2023). Estimating the Compressive Strength of Self-compacting Concrete with fiber using an Extreme Gradient Boosting model. *Journal of*

- Science and Transport Technology*, 3(1), 12-25.
- [9] H.-V.T. Mai, S.H. Trinh, H.-B. Ly. (2023). Enhancing Compressive strength prediction of Roller Compacted concrete using Machine learning techniques. *Measurement*, 218, 113196.
- [10] N. Baldo, E. Manthos, M. Pasetto. (2018). Analysis of the Mechanical Behaviour of Asphalt Concretes Using Artificial Neural Networks. *Advances in Civil Engineering*, 2018, 1650945.
- [11] M.A. Gul et al. (2022). Prediction of Marshall Stability and Marshall Flow of Asphalt Pavements Using Supervised Machine Learning Algorithms. *Symmetry*, 14(11), 2324.
- [12] M.M. Mukaka. (2012). Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal*, 24(3), 69-71.
- [13] A. Natekin, A. Knoll. (2013). Gradient Boosting Machines, A Tutorial. *Frontiers in Neurorobotics*, 7, 21.
- [14] G. Ke et al. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Advances in Neural Information Processing Systems 30 (NIPS 2017)*.
- [15] J.R. Quinlan. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81-106.
- [16] A. Setiawan, L.B. Suparma, A.T. Mulyono. (2017). Modelling Effect of Aggregate Gradation and Bitumen Content on Marshall Properties of Asphalt Concrete. *International Journal on Advanced Science, Engineering and Information Technology*, 7(2), 359-365.