

XÂY DỰNG

TẠP CHÍ ĐIỆN TỬ CỦA BỘ XÂY DỰNG
JOURNAL OF CONSTRUCTION

TẠP CHÍ XÂY DỰNG - eISSN 3030-4482

Phân tích đặc tính lưu biến của nhựa đường biến tính sợi polypropylene sử dụng mô hình học máy và SHAP

Analysis of rheological properties of polypropylene fiber-modified asphalt using machine learning models and SHAP

► TS Hoàng Thị Hương Giang

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: gianghth@utt.edu.vn

THÔNG TIN BÀI BÁO

Chuyên mục: Khoa học công nghệ

Ngày nhận bài: 07/5/2026

Ngày sửa bài: 20/5/2026

Ngày chấp nhận đăng: 05/6/2026

Ngày xuất bản Online: 21/6/2026

Tác giả liên hệ:

Email: gianghth@utt.edu.vn

TÓM TẮT

Đặc tính lưu biến là yếu tố quan trọng phản ánh khả năng làm việc của nhựa đường dưới tác dụng tải trọng và nhiệt độ, trong đó mô-đun cắt động (G^*) là chỉ tiêu đại diện cho độ cứng và khả năng kháng hằn lún của vật liệu. Nghiên cứu này phân tích đặc tính lưu biến của nhựa đường biến tính sợi polypropylene (PP) thông qua các mô hình học máy và phương pháp giải thích SHAP. Bộ dữ liệu gồm 132 mẫu được tổng hợp từ nhiều nguồn, bao gồm các đặc trưng vật liệu và điều kiện chế tạo. Kết quả cho thấy mô hình CatBoost (CAT) đạt hiệu suất vượt trội với độ chính xác cao ($R^2 = 0,995$ và $RMSE = 1,000$ trên tập kiểm tra). Phân tích SHAP chỉ ra rằng mô-đun cắt động của nhựa ban đầu là yếu tố chi phối chính, trong khi tỷ lệ sợi và các thông số công nghệ đóng vai trò tương tác. Nghiên cứu góp phần làm rõ cơ chế ảnh hưởng của sợi PP và cung cấp cơ sở cho tối ưu hóa thiết kế nhựa đường biến tính.

Từ khóa: Polypropylene (PP); học máy; mô-đun cắt động (G^*); nhựa đường biến tính.

ABSTRACT

Rheological properties are critical indicators reflecting the performance of asphalt under loading and temperature conditions, in which the dynamic shear modulus (G^*) serves as a representative parameter for material stiffness and rutting resistance. This study

analyzes the rheological properties of polypropylene (PP) fiber-modified asphalt using machine learning models and the SHAP interpretability method. The dataset consists of 132 samples compiled from multiple sources, including material characteristics and processing conditions. The results show that the CatBoost (CAT) model achieves superior performance with high predictive accuracy ($R^2 = 0.995$ and $RMSE = 1.000$ on the testing set). SHAP analysis reveals that the dynamic shear modulus of the base asphalt is the most influential factor, while fiber content and processing parameters play interactive roles. This study contributes to a better understanding of the influence mechanism of PP fibers and provides a scientific basis for optimizing the design of modified asphalt.

Keywords: Polypropylene (PP); machine learning; dynamic shear modulus (G^*); modified asphalt.

1. ĐẶT VẤN ĐỀ

Trong những năm gần đây, nhu cầu nâng cao chất lượng và tuổi thọ của kết cấu mặt đường nhựa ngày càng trở nên cấp thiết, đặc biệt trong điều kiện tải trọng giao thông gia tăng và biến đổi khí hậu phức tạp. Một trong những dạng hư hỏng phổ biến và nghiêm trọng nhất của mặt đường nhựa là hiện tượng hằn lún vệt bánh xe, ảnh hưởng trực tiếp đến khả năng khai thác và an toàn giao thông [1, 2]. Do đó, việc cải thiện đặc tính cơ học, đặc biệt là đặc tính lưu biến của nhựa đường đóng vai trò quan trọng trong việc nâng cao hiệu năng làm việc của vật liệu.

Trong số các giải pháp cải tiến vật liệu, việc sử dụng sợi polymer, đặc biệt là sợi polypropylene (PP) đã thu hút nhiều sự quan tâm nhờ khả năng tăng cường độ cứng, cải thiện khả năng kháng biến dạng và nâng cao độ bền của nhựa đường [3, 4]. Tuy nhiên, hiệu quả của việc biến tính phụ thuộc mạnh mẽ vào nhiều yếu tố như hàm lượng sợi, đặc tính cơ học của sợi và điều kiện chế tạo, dẫn đến mối quan hệ phức tạp và phi tuyến giữa các biến đầu vào và đặc tính đầu ra. Đặc tính lưu biến của nhựa đường thường được đánh giá thông qua các chỉ tiêu như mô-đun cắt động (G^*), phản ánh đồng thời tính đàn hồi và nhớt của vật liệu dưới tác dụng tải trọng [5]. Tuy nhiên, việc xác định và dự báo các đặc tính này bằng phương pháp thực nghiệm truyền thống thường tốn kém thời gian, chi phí và khó bao quát hết các điều kiện làm việc khác nhau.

Trong bối cảnh đó, các phương pháp học máy đã nổi lên như một công cụ hiệu quả trong việc mô hình hóa các mối quan hệ phi tuyến và khai thác thông tin từ dữ liệu [6]. Một số chỉ tiêu cơ lý của nhựa đường biến tính polymer và nhựa đường cao su cũng được giải quyết bằng các thuật toán ML với độ chính xác rất cao với R^2 có thể lên tới 0,976 [7]. Bên cạnh đó, các phương pháp giải thích mô hình như SHAP cho phép làm rõ vai trò và mức độ ảnh hưởng của từng biến đầu vào, góp phần nâng cao tính minh bạch và khả năng ứng dụng của mô hình.

Xuất phát từ những vấn đề trên, nghiên cứu này tập trung phân tích đặc tính lưu biến của nhựa đường biến tính sợi PP thông qua các mô hình học máy kết hợp với phương pháp SHAP. Mục tiêu của nghiên cứu là (i) xây dựng và đánh giá hiệu suất của các mô hình học máy, (ii) xác định các yếu tố ảnh hưởng chính đến đặc tính lưu biến và (iii) làm rõ cơ chế tương tác giữa các biến đầu vào. Kết quả nghiên cứu sẽ cung cấp cơ sở khoa học cho việc tối ưu hóa thiết kế và ứng dụng nhựa đường biến tính trong thực tiễn.

2. CƠ SỞ DỮ LIỆU

Trong nghiên cứu này, bộ dữ liệu được xây dựng thông qua việc tổng hợp từ nhiều nguồn tài liệu khoa học đã được công bố, nhằm đảm bảo tính đa dạng, độ tin cậy và khả năng đại diện cho các điều kiện thí nghiệm khác nhau. Cụ thể, dữ liệu được thu thập từ 6 nghiên cứu trước đây với tổng số 132 mẫu thí nghiệm (Bảng 1). Việc tổng hợp dữ liệu từ nhiều nguồn giúp mở rộng không gian biến đầu vào, đồng thời nâng cao tính tổng quát của mô hình học máy được phát triển.

Bộ dữ liệu bao gồm 7 biến đầu vào và 1 biến đầu ra. Các biến đầu vào phản ánh đặc trưng cơ học của sợi PP cũng như điều kiện chế tạo nhựa đường biến tính, bao gồm: Độ bền kéo của sợi (X1), tỷ lệ sợi (X2), độ dài sợi (X3), nhiệt độ trộn (X4), thời gian trộn (X5), vận tốc trộn (X6) và mô-đun cắt động của nhựa đường ban đầu G^* (X7). Biến đầu ra (Y) là mô-đun cắt động của nhựa đường sau khi biến tính bằng sợi PP, đại diện cho khả năng kháng biến dạng vĩnh viễn của vật liệu.

Để đánh giá đặc điểm phân bố và mức độ biến thiên của dữ liệu, các chỉ tiêu thống kê mô tả đã được tính toán và trình bày trong Bảng 2. Kết quả chỉ ra các biến đầu vào có phạm vi giá trị rộng, phản ánh sự đa dạng của điều kiện thí nghiệm. Ví dụ, tỷ lệ sợi (X2) dao động từ 0% đến 5%, trong khi độ dài sợi (X3) thay đổi từ 0 đến 10 mm. Tương tự, các thông số công nghệ như nhiệt độ trộn (X4) và vận tốc trộn (X6) cũng có độ phân tán lớn, chứng minh dữ liệu bao phủ nhiều điều kiện chế tạo khác nhau. Đáng chú ý, biến đầu ra G^* của nhựa biến tính (Y) có giá trị trung bình 10,74 kPa nhưng độ lệch chuẩn lớn (30,72 kPa), chứng tỏ tồn tại sự biến động đáng kể trong hiệu quả cải thiện tính chất cơ học của nhựa đường khi sử dụng sợi PP.

Ngoài ra, ma trận tương quan (Hình 1) được sử dụng để đánh giá mối quan hệ tuyến tính giữa các biến, qua đó cung cấp sơ bộ về mức độ ảnh hưởng của từng biến đầu vào tới biến mục tiêu. Biểu đồ phân phối (Hình 2) chỉ ra phần lớn các biến có phân bố không đồng đều, thậm chí lệch phải đối với một số biến như G^* , điều này đặt ra yêu cầu đối với các mô hình học máy phải có khả năng xử lý dữ liệu phi tuyến và không phân phối chuẩn.

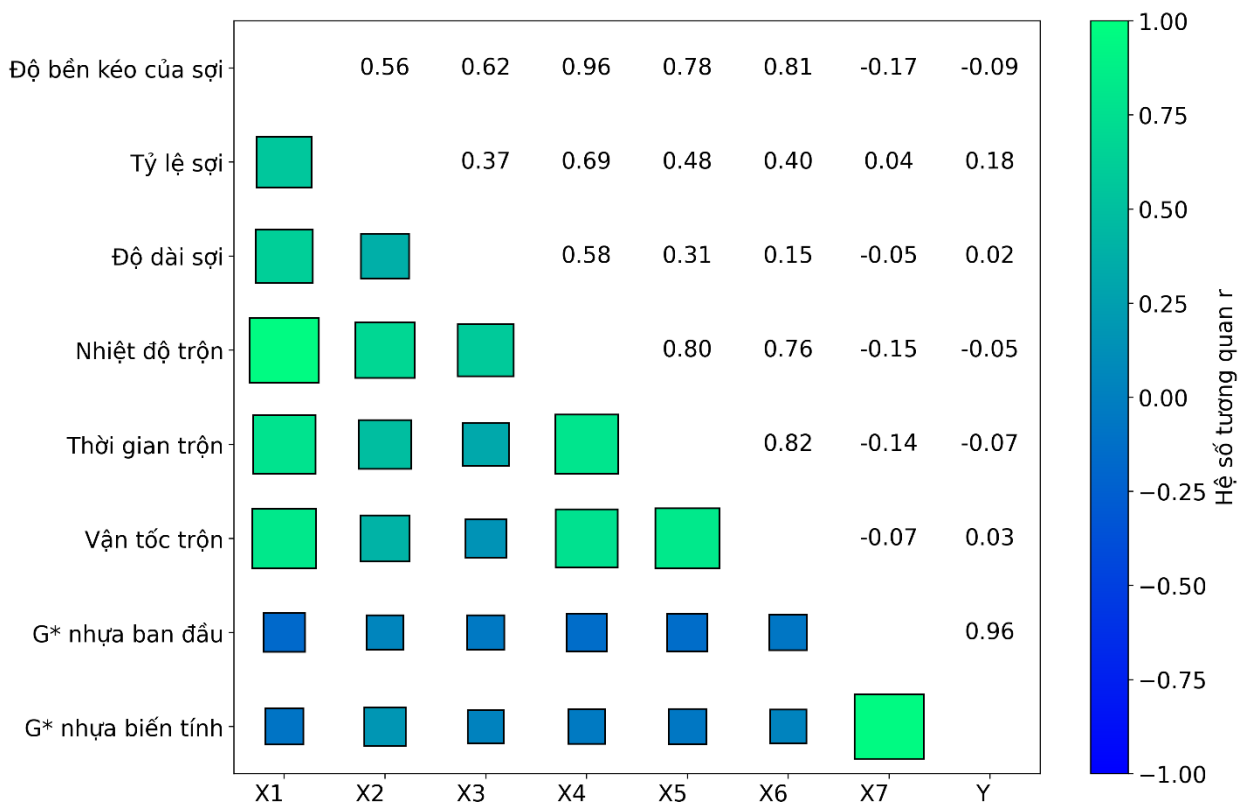
Cuối cùng, bộ dữ liệu được chia làm 2 phần, 70% dữ liệu dùng để huấn luyện mô hình, 30% dữ liệu còn lại dùng để đánh giá mô hình. Nhìn chung, bộ dữ liệu được xây dựng có quy mô vừa phải nhưng đảm bảo tính phổ quát và tính đại diện, phù hợp cho việc phát triển và đánh giá các mô hình học máy trong dự báo khả năng kháng hần lún của nhựa đường biến tính sợi PP.

Bảng 1. Nguồn và quy mô bộ dữ liệu sử dụng trong nghiên cứu

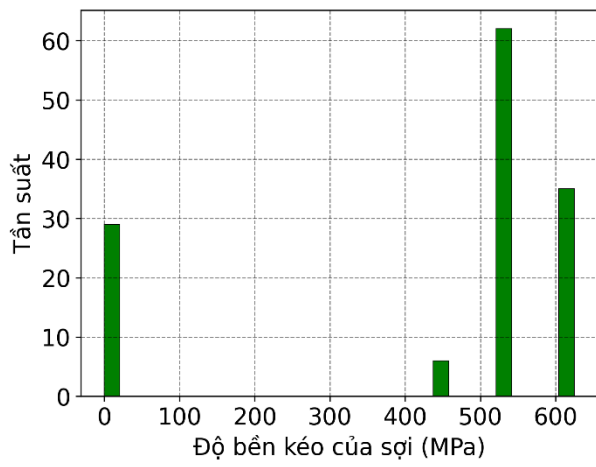
STT	Nguồn thu thập	Số lượng mẫu
1	Yeh và các cộng sự, 2005 [8]	20
2	Guo và các cộng sự, 2022 [9]	24
3	Zhang và các cộng sự, 2024 [10]	4
4	Li và các cộng sự, 2023 [3]	12
5	Kathari và các cộng sự, 2016 [11]	42
6	Xiaoming và các cộng sự, 2019 [4]	30
	Tổng	132

Bảng 2. Thống kê mô tả các biến đầu vào và đầu ra của bộ dữ liệu

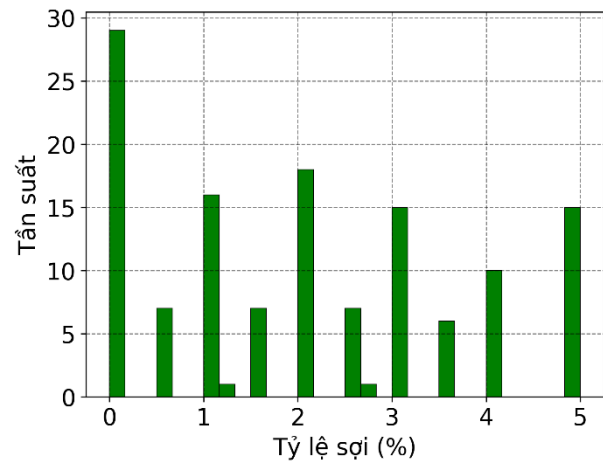
Tên biến	Đơn vị	Trung bình	Độ lệch chuẩn	Nhỏ nhất	25%	50%	75%	Lớn nhất
Độ bền kéo của sợi (X1)	%	438,64	237,71	0,00	450,00	537,50	625,00	625,00
Tỷ lệ sợi (X2)	%	2,03	1,65	0,00	0,50	2,00	3,00	5,00
Độ dài sợi (X3)	mm	4,05	3,51	0,00	0,10	3,50	6,00	10,00
Nhiệt độ trộn (X4)	°C	137,20	60,14	25,00	160,00	165,00	170,00	180,00
Thời gian trộn (X5)	Phút	44,05	30,35	0,00	30,00	40,00	60,00	90,00
Vận tốc trộn (X6)	Vòng /phút	1.547,16	1.053,79	0,00	500,00	2.200,00	2.500,00	2.575,00
G* nhựa ban đầu (X7)	kPa	6,70	23,11	0,10	0,43	1,03	2,99	168,53
G* nhựa biến tính (Y)	kPa	10,74	30,72	0,10	0,86	2,28	6,19	245,13



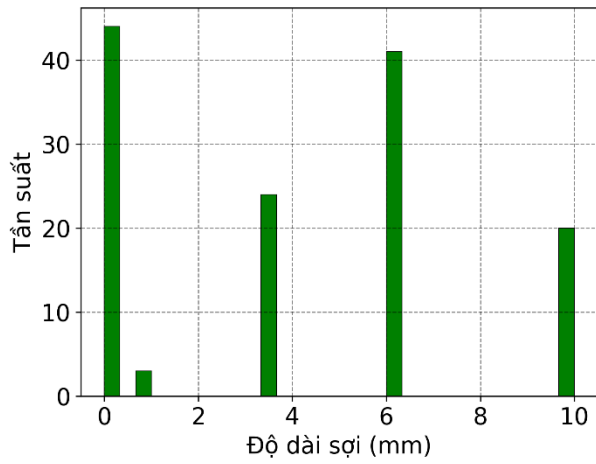
Hình 1. Ma trận tương quan giữa các biến trong bộ dữ liệu



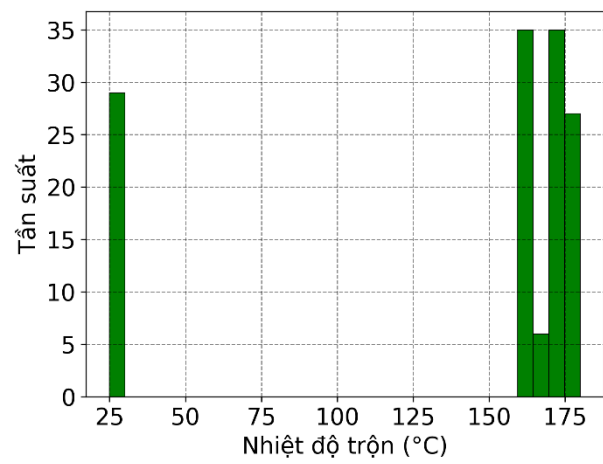
a)



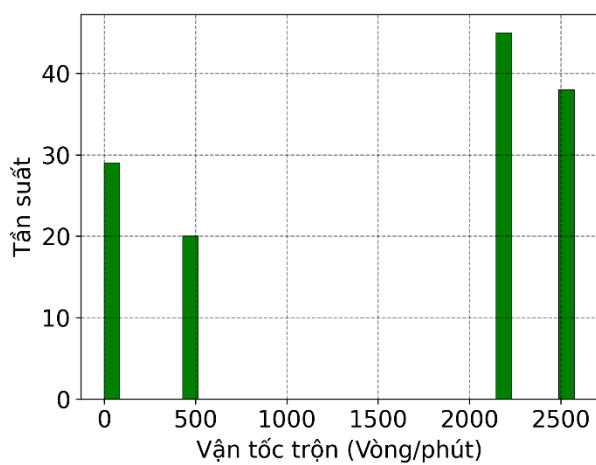
b)



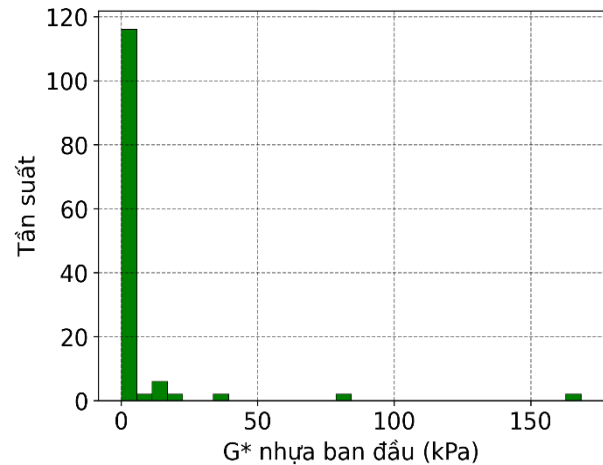
c)



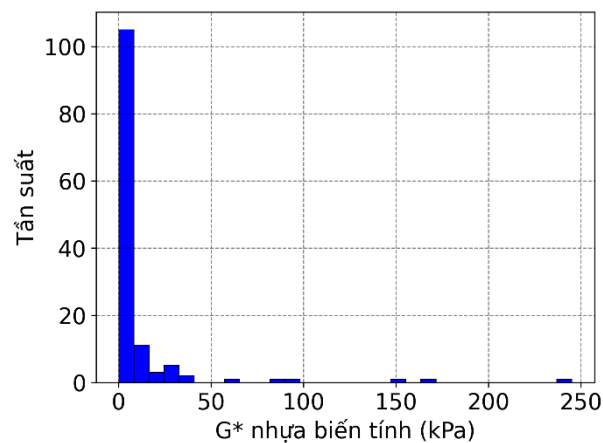
d)



e)



f)



g)

Hình 2. Phân bố dữ liệu của các biến đầu vào và đầu ra

3. PHƯƠNG PHÁP HỌC MÁY

Trong nghiên cứu này, các phương pháp học máy được áp dụng nhằm xây dựng mô hình dự báo mối quan hệ phi tuyến giữa các đặc trưng vật liệu, điều kiện chế tạo và khả năng kháng hàn lún của nhựa đường biến tính sợi PP. Do bản chất phức tạp, đa biến và không tuyến tính của dữ liệu, các thuật toán học máy tiên tiến dựa trên cây quyết định và phương pháp học kết hợp được lựa chọn để khai thác hiệu quả thông tin bộ dữ liệu.

Quy trình xây dựng mô hình bao gồm các bước chính: Huấn luyện mô hình với nhiều thuật toán khác nhau, đánh giá hiệu suất thông qua kỹ thuật xác thực chéo và lựa chọn mô hình tối ưu dựa trên các tiêu chí thống kê. Đồng thời, phương pháp giải thích mô hình SHAP được sử dụng nhằm phân tích mức độ ảnh hưởng của các biến đầu vào, qua đó làm rõ cơ chế tác động của sợi PP đến tính chất cơ học của nhựa đường.

3.1. Thuật toán CatBoost

CatBoost (CAT) là một thuật toán tăng cường tiên tiến dựa trên cây quyết định, được thiết kế nhằm cải thiện độ chính xác dự báo và hạn chế hiện tượng quá khớp. Thuật toán này sử dụng cơ chế tăng cường để giảm sai lệch trong quá trình học, đồng thời áp dụng phương pháp mã hóa đặc trưng hiệu quả, giúp khai thác tốt các mối quan hệ phi tuyến và tương tác phức tạp giữa các biến đầu vào.

3.2. Thuật toán XGBoost

XGBoost (XGB) là một trong những thuật toán tăng cường độ dốc mạnh mẽ và phổ biến nhất hiện nay. Mô hình hoạt động bằng cách xây dựng tuần tự các cây quyết định, trong đó mỗi cây mới được huấn luyện để giảm sai số của các cây trước đó. XGBoost kiểm soát tốt độ phức tạp của mô hình, đồng thời tối ưu hóa hiệu suất tính toán. Nhờ khả năng học các quan hệ phi tuyến mạnh mẽ, XGBoost thường cho kết quả dự báo chính xác trong các bài toán hồi quy phức tạp.

3.3. Thuật toán cây quyết định

Cây quyết định (DT) là một mô hình học có giám sát cơ bản, trong đó dữ liệu được phân chia thành các nhánh dựa trên các ngưỡng tối ưu của biến đầu vào. Quá trình phân tách được thực hiện nhằm tối thiểu hóa sai số dự báo tại mỗi nút. Ưu điểm của phương pháp này là dễ hiểu, dễ triển khai và có khả năng mô hình hóa quan hệ phi tuyến. Tuy nhiên, cây quyết định đơn lẻ thường dễ bị quá khớp nếu không được kiểm soát độ sâu hoặc số lượng nút.

3.4. Thuật toán tăng cường độ dốc

Thuật toán tăng cường độ dốc (GB) xây dựng mô hình dựa trên việc kết hợp tuần tự nhiều cây quyết định yếu. Mỗi cây mới được huấn luyện để xấp xỉ phần sai số của mô hình trước đó, từ đó cải thiện dần độ chính xác tổng thể. Phương pháp này có khả năng mô hình hóa tốt các quan hệ phi tuyến và tương tác giữa các biến, tuy nhiên cần lựa chọn tham số cẩn thận để tránh hiện tượng quá khớp.

3.5. Thuật toán rừng ngẫu nhiên

Rừng ngẫu nhiên (RF) là một phương pháp học kết hợp dựa trên việc xây dựng nhiều cây quyết định độc lập từ các tập con dữ liệu được chọn ngẫu nhiên. Mỗi cây được huấn luyện trên một tập dữ liệu khởi tạo và chỉ sử dụng một tập con các biến đầu vào tại mỗi lần phân tách. Kết quả dự báo cuối cùng được tổng hợp từ các cây thành phần, giúp giảm phương sai và tăng độ ổn định của mô hình.

3.6. Xác thực chéo

Xác thực chéo là kỹ thuật được sử dụng để đánh giá khả năng tổng quát hóa của mô hình học máy. Dữ liệu được chia thành nhiều tập con, trong đó mô hình được huấn luyện trên một số tập và kiểm tra trên tập còn lại. Quá trình này được lặp lại nhiều lần với các cách chia khác nhau, giúp giảm sự phụ thuộc vào một cách phân chia dữ liệu cụ thể và cung cấp đánh giá đáng tin cậy hơn về hiệu suất mô hình. Nghiên cứu này sử dụng xác thực chéo 5 lần.

3.7. Các tiêu chí đánh giá hiệu suất dự báo mô hình

Hiệu suất của các mô hình được đánh giá thông qua các chỉ tiêu thống kê phổ biến. Hệ số xác định (R^2) phản ánh mức độ phù hợp giữa giá trị dự báo và thực nghiệm. Sai số bình phương trung bình (RMSE) đo lường độ lệch trung bình theo bình phương, nhạy với các sai số lớn. Sai số tuyệt đối trung bình (MAE) thể hiện sai lệch trung bình theo giá trị tuyệt đối, trong khi sai số phần trăm tuyệt đối trung bình (MAPE) cho biết mức sai số tương đối theo phần trăm. Việc kết hợp nhiều chỉ tiêu giúp đánh giá toàn diện độ chính xác và độ tin cậy của mô hình.

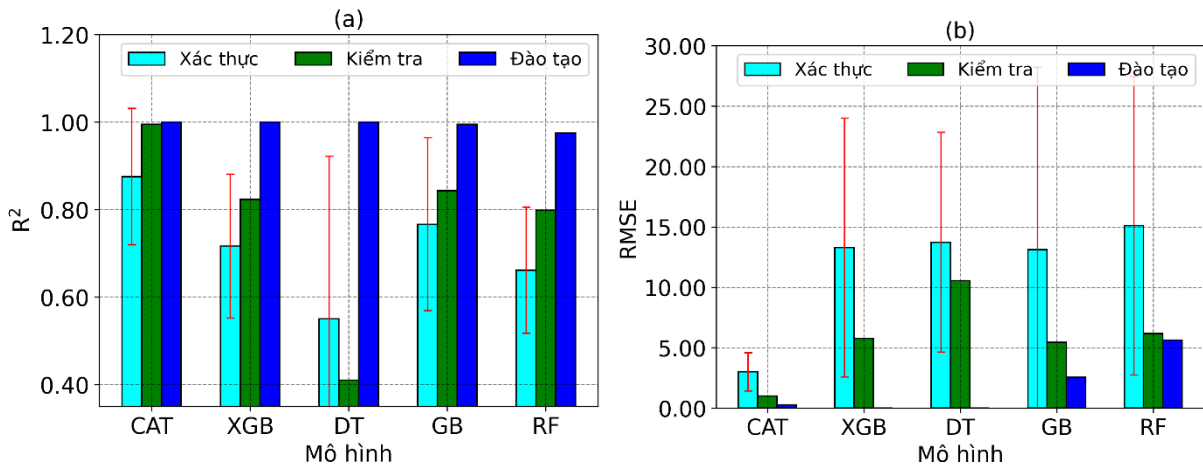
4. KẾT QUẢ VÀ THẢO LUẬN

4.1. Xây dựng mô hình

Các mô hình học máy được xây dựng với tham số mặc định nhằm đảm bảo tính khách quan trong quá trình so sánh hiệu suất. Dữ liệu được chia thành các tập huấn luyện, kiểm tra và xác thực, kết hợp đánh giá thông qua các chỉ tiêu R^2 và RMSE. Kết quả được thể hiện trực quan trong Hình 3 và chi tiết trong Bảng 3.

Quan sát Hình 3 cho thấy mô hình CAT luôn đạt R^2 cao nhất và RMSE thấp nhất trên cả ba tập dữ liệu, trong khi các mô hình khác có xu hướng suy giảm hiệu suất rõ rệt khi chuyển sang tập kiểm tra và xác thực. Cụ thể, CAT đạt $R^2 = 0,999$ và RMSE = 0,264 (tập đào tạo); $R^2 = 0,995$ và RMSE = 1,000 (tập kiểm tra); $R^2 = 0,875$ và RMSE = 3,000 (tập xác thực). Ngược lại, các mô hình như XGB, GB, DT và RF có RMSE trên tập xác thực dao động từ 13,113 đến 15,091 chỉ ra sai số lớn và kém ổn định. Đặc biệt, Hình 3b minh họa rõ sự chênh lệch về sai số khi RMSE của CAT thấp hơn khoảng 4 - 5 lần so với các mô hình còn lại trên tập xác thực. Điều này khẳng định khả năng dự báo chính xác và tính tổng quát hóa tốt của mô hình.

Trên cơ sở các kết quả định lượng và trực quan, mô hình CAT được lựa chọn là mô hình tối ưu để sử dụng trong các phân tích chuyên sâu tiếp theo.



Hình 3. So sánh hiệu suất dự báo của các mô hình học máy theo (a) R^2 và (b) RMSE

Bảng 3. So sánh hiệu suất dự báo của các mô hình học máy trên các tập dữ liệu

Tiêu chí	R^2	RMSE	R^2	RMSE	R^2	RMSE
Mô hình	Tập đào tạo		Tập kiểm tra		Tập xác thực	
CAT	0,999	0,264	0,995	1,000	0,875	3,000
XGB	0,999	0,003	0,823	5,772	0,716	13,287
DT	0,999	0,000	0,410	10,547	0,550	13,732
GB	0,994	2,547	0,843	5,436	0,766	13,113
RF	0,974	0,562	0,798	6,164	0,661	15,091

4.2. Kết quả dự báo mô hình tiêu biểu

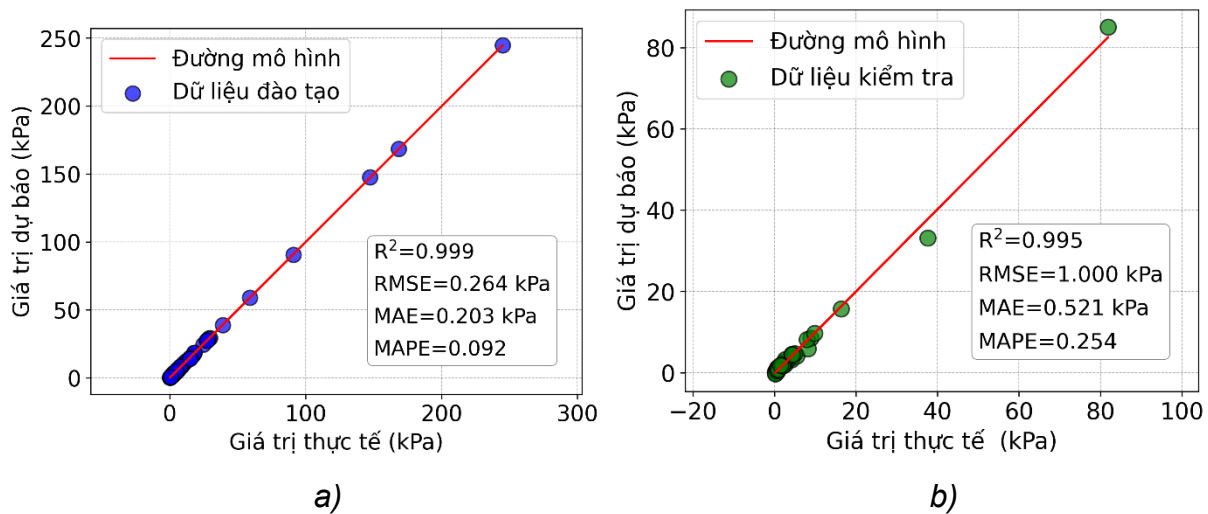
Kết quả dự báo của mô hình CAT được thể hiện trong Hình 4 và Hình 5, chứng minh mức độ phù hợp cao giữa giá trị dự báo và giá trị thực tế. Cụ thể, trong Hình 4, các điểm dữ liệu trên cả tập huấn luyện và kiểm tra phân bố tập trung xung quanh đường hồi quy, chứng tỏ mô hình có khả năng dự báo chính xác và ít sai lệch hệ thống.

Định lượng hơn, các chỉ số trong Bảng 4 cho thấy mô hình đạt hiệu suất rất cao với $R^2 = 0,999$, $RMSE = 0,264$, $MAE = 0,203$ và $MAPE = 0,092$ trên tập đào tạo; đồng thời vẫn duy trì độ chính xác tốt trên tập kiểm tra với $R^2 = 0,995$, $RMSE = 1,000$, $MAE = 0,521$ và $MAPE = 0,254$. Sự chênh lệch nhỏ giữa hai tập dữ liệu khẳng định mô hình không bị quá khớp và có khả năng tổng quát hóa tốt.

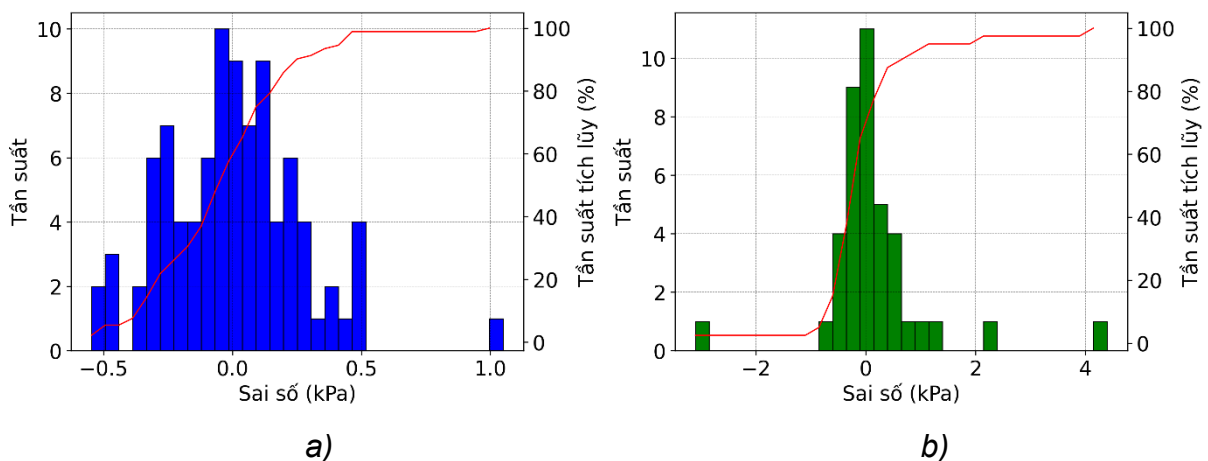
Ngoài ra, Hình 5 chỉ ra sai số dự báo phân bố tương đối đồng đều quanh giá trị 0 và không xuất hiện xu hướng lệch rõ rệt với 3 mẫu có sai số nằm ngoài khoảng $[-0,5; 0,5]$ (kPa) trên tập dữ liệu đào tạo và 3 mẫu dữ liệu có sai số nằm ngoài khoảng $[-2; 2]$ (kPa) trên tập kiểm tra. Điều này khẳng định tính ổn định của mô hình và khả năng dự báo chính xác, đáng tin cậy đối với chỉ tiêu kháng lún của nhựa đường biến tính.

Bảng 4. Chỉ số đánh giá hiệu suất mô hình CAT

Tiêu chí đánh giá	R^2	RMSE	MAE	MAPE
Dữ liệu đào tạo	0,999	0,264	0,203	0,092
Dữ liệu kiểm tra	0,994	1,000	0,521	0,254



Hình 4. So sánh giá trị thực nghiệm và giá trị dự báo của mô hình CAT: a) - Tập đào tạo; b) - Tập kiểm tra



Hình 5. Phân bố sai số dự báo của mô hình CAT: a) - Tập đào tạo; b) - Tập kiểm tra

4.3. Phân tích ảnh hưởng của các biến đầu vào tới dự báo kết quả đầu ra

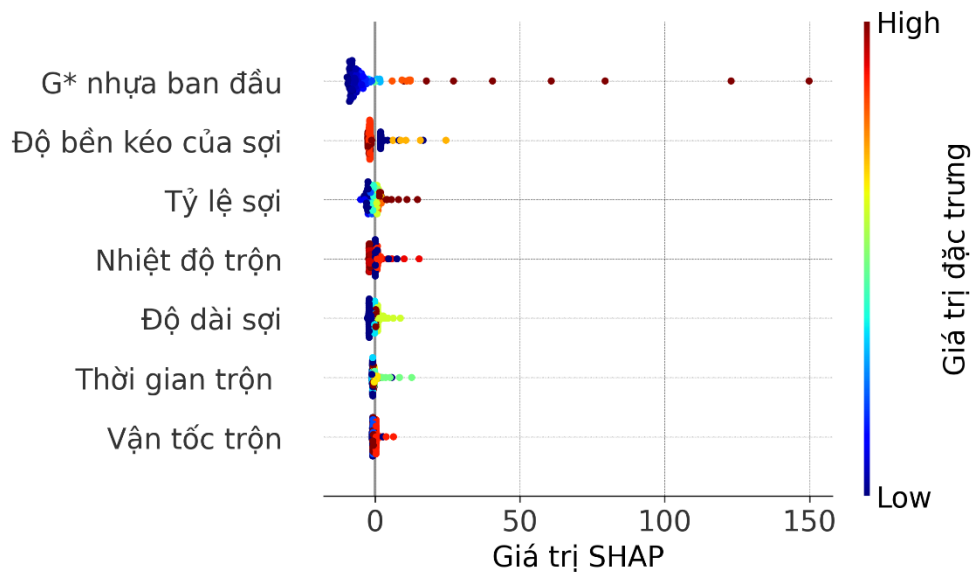
Ảnh hưởng của các biến đầu vào được phân tích thông qua các biểu đồ SHAP. Hình 6 thể hiện biểu đồ SHAP beeswarm, cung cấp cái nhìn tổng thể về mức độ quan trọng và chiều hướng ảnh hưởng của các biến đầu vào đến kết quả dự báo.

Kết quả khẳng định biến G^* nhựa ban đầu (X7) có ảnh hưởng vượt trội so với các biến còn lại, với giá trị SHAP trải rộng và xuất hiện nhiều điểm có giá trị rất lớn (có thể vượt 150). Điều này khẳng định đây là yếu tố chi phối chính đến khả năng dự báo hệ số hàn lún. Đồng thời, các điểm màu đỏ (giá trị cao của X7) chủ yếu nằm phía dương, cho thấy khi G^* ban đầu tăng thì giá trị dự báo cũng tăng mạnh. Điều này hoàn toàn phù hợp với nghiên cứu [12] khi sử dụng GO là phụ gia trong nhựa đường.

Biến độ bền kéo của sợi (X1) và tỷ lệ sợi (X2) có mức ảnh hưởng đứng sau, tuy nhiên biên độ SHAP nhỏ hơn đáng kể so với X7. Các điểm dữ liệu phân bố hai phía quanh giá trị 0 chỉ ra tác động của các biến này có thể thay đổi tùy theo miền giá trị, không hoàn toàn đơn điệu.

Các biến nhiệt độ trộn (X4) và độ dài sợi (X3) có ảnh hưởng trung bình, với phân bố SHAP hẹp hơn, khẳng định vai trò điều chỉnh nhưng không chi phối chính. Trong khi

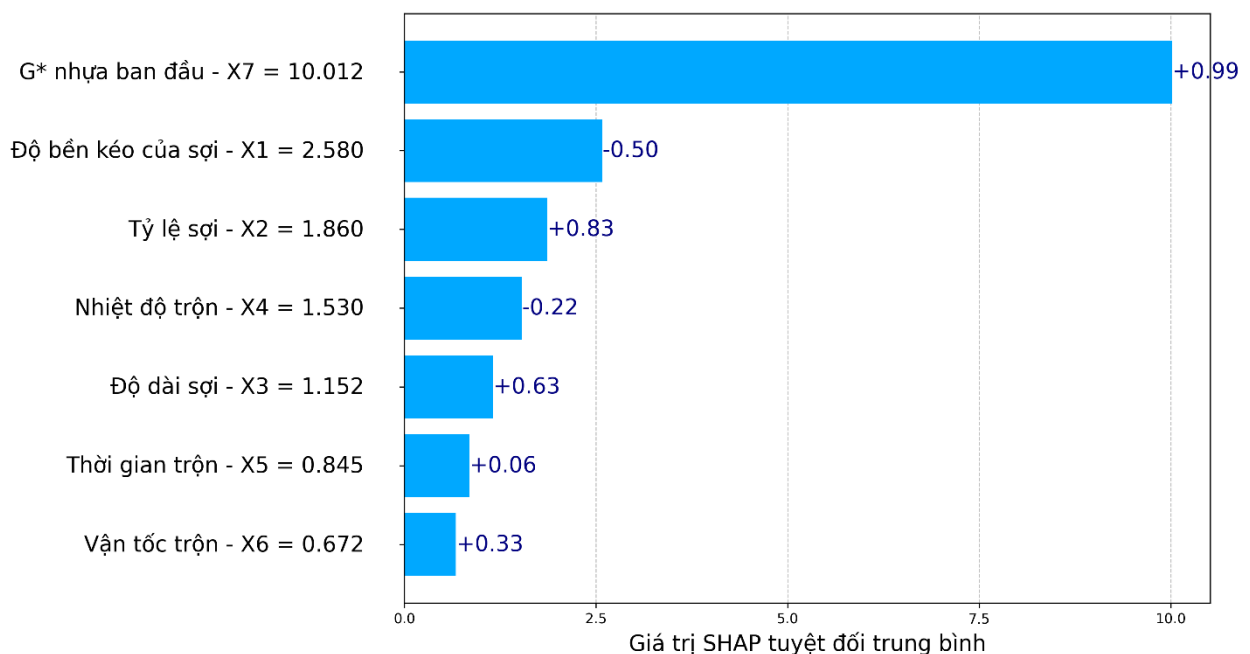
đó, thời gian trộn (X5) và vận tốc trộn (X6) có giá trị SHAP tập trung rất gần 0, chứng tỏ ảnh hưởng của chúng đến kết quả dự báo là không đáng kể.



Hình 6. Biểu đồ SHAP beeswarm thể hiện mức độ và chiều hướng ảnh hưởng của các biến đầu vào

Kết quả từ biểu đồ SHAP (Hình 6) được củng cố và lượng hóa rõ ràng hơn thông qua biểu đồ SHAP dạng thanh. Hình 7 thể hiện giá trị SHAP tuyệt đối trung bình, qua đó định lượng mức độ quan trọng của từng biến. Kết quả chỉ ra X7 (G* nhựa ban đầu) có giá trị lớn nhất (10,012), vượt trội so với các biến còn lại và khẳng định vai trò chi phối chính trong mô hình.

Tiếp theo là X1 (2,580) và X2 (1,258), cho thấy hai biến này có ảnh hưởng đáng kể nhưng thấp hơn nhiều so với X7. Các biến X4 (1,530) và X3 (1,152) có mức ảnh hưởng trung bình, đóng vai trò hỗ trợ trong dự báo. Trong khi đó, X5 (0,845) và X6 (0,672) có giá trị SHAP rất nhỏ, gần như không đáng kể so với X7, chứng minh ảnh hưởng hạn chế đến kết quả mô hình. Sự chênh lệch lớn giữa X7 và các biến còn lại nhấn mạnh tính quyết định của biến này trong bài toán.



Hình 7. Biểu đồ SHAP dạng thanh thể hiện mức độ quan trọng của các biến đầu vào

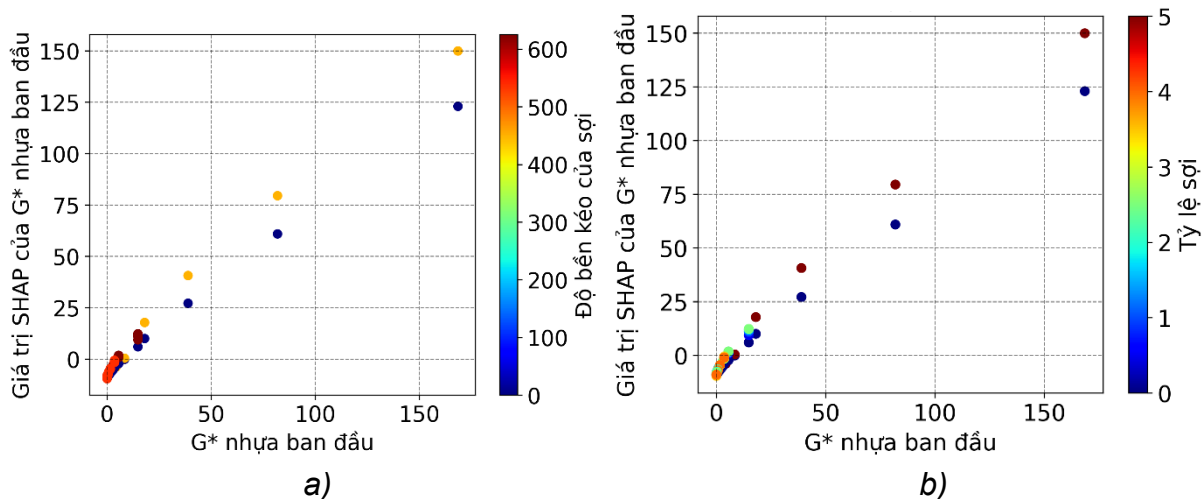
Cuối cùng, Hình 8 chỉ ra ảnh hưởng của các cặp biến quan trọng đến đầu ra. Hình 8a thể hiện ảnh hưởng đồng thời của G^* nhựa ban đầu (X7) và độ bền kéo của sợi (X1) đến giá trị SHAP. Có thể thấy giá trị SHAP tăng mạnh theo X7, khẳng định vai trò chi phối của biến này. Đồng thời, tại cùng một mức X7, các điểm có X1 cao (màu đỏ) thường cho giá trị SHAP lớn hơn so với X1 thấp, cho thấy X1 có tác động làm gia tăng hiệu ứng của X7, dù mức độ ảnh hưởng vẫn nhỏ hơn đáng kể.

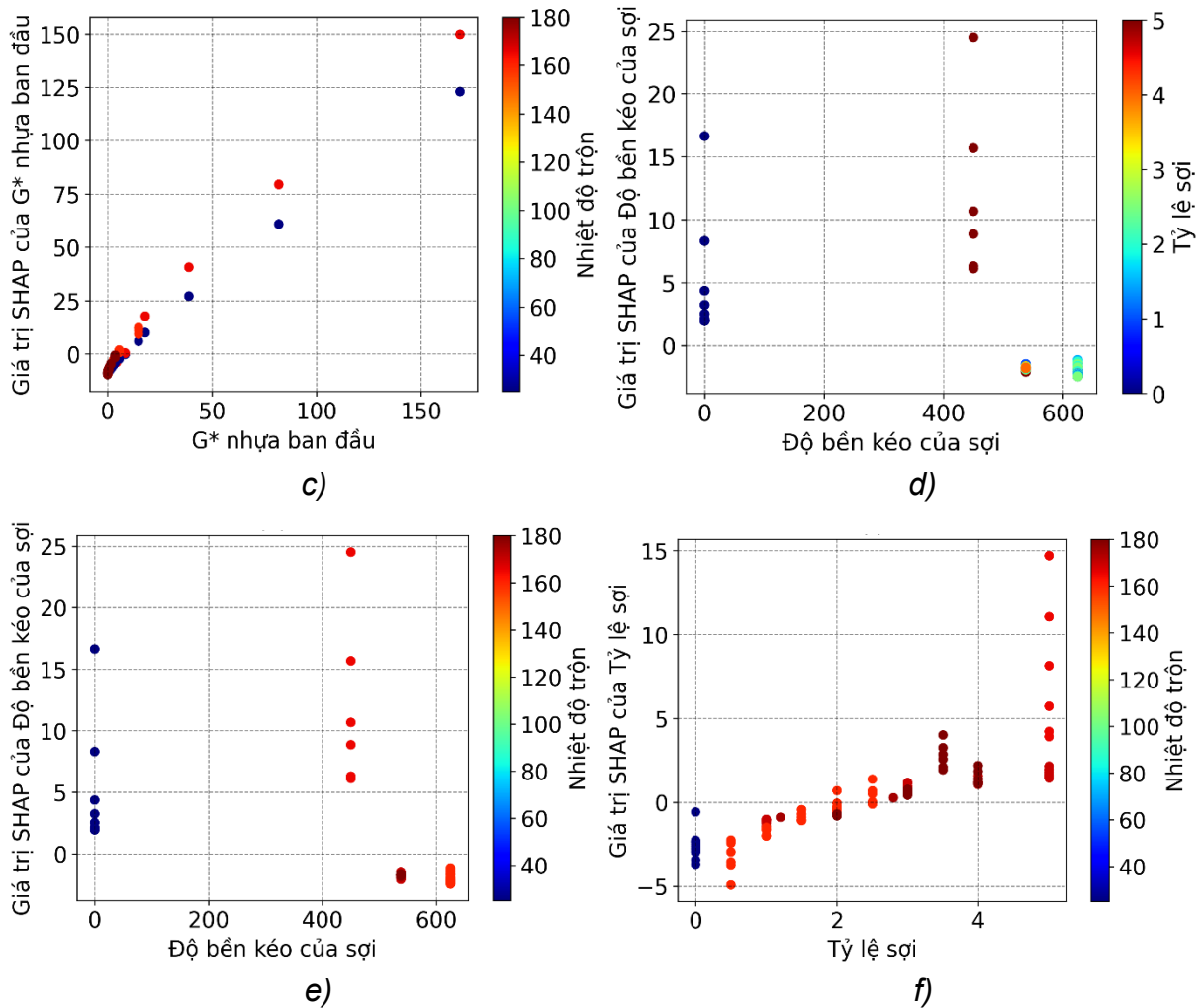
Hình 8b mô tả tương tác giữa G^* nhựa ban đầu (X7) và tỷ lệ sợi (X2). Xu hướng chung vẫn là SHAP tăng theo X7, tuy nhiên ảnh hưởng của X2 thể hiện rõ hơn so với X1. Cụ thể, tại cùng một giá trị X7, các điểm có X2 cao (màu đỏ) tương ứng với giá trị SHAP lớn hơn, khẳng định tỷ lệ sợi góp phần khuếch đại đáng kể tác động của X7 đến kết quả dự báo. Điều này khẳng định X2 là biến tương tác quan trọng thứ hai sau X7 trong mô hình.

Bên cạnh đó, ảnh hưởng của X4 thể hiện ở mức trung bình: tại cùng một giá trị X7, các điểm có nhiệt độ trộn cao (màu đỏ) có xu hướng cho giá trị SHAP lớn hơn so với nhiệt độ thấp (Hình 8c). Tuy nhiên, mức độ phân tách không quá rõ rệt như X2, chỉ ra X4 có tác động tương tác nhưng không phải yếu tố quyết định.

Cuối cùng, các Hình 8(d-f) cho thấy vai trò tương tác giữa đặc trưng của sợi và điều kiện công nghệ. Giá trị SHAP có xu hướng tăng khi tỷ lệ sợi (X2) tăng, cho thấy đây là biến tương tác quan trọng hơn so với độ bền kéo sợi (X1). Đồng thời, nhiệt độ trộn (X4) có xu hướng làm gia tăng giá trị SHAP tại cùng một mức X1 hoặc X2, tuy mức độ ảnh hưởng không quá rõ rệt. Nhìn chung, các kết quả này tiếp tục khẳng định X2 và X4 đóng vai trò khuếch đại tác động của các biến vật liệu đến đặc tính lưu biến của nhựa đường biến tính PP.

Như vậy, các kết quả phân tích SHAP cho thấy G^* nhựa ban đầu (X7) là yếu tố chi phối quyết định đến khả năng dự báo, trong khi tỷ lệ sợi (X2) và một số biến khuấy trộn như X1, X4 đóng vai trò tương tác và khuếch đại ảnh hưởng. Điều này khẳng định mô hình không chỉ phụ thuộc vào một biến đơn lẻ mà còn chịu tác động của các tương tác phi tuyến phức tạp giữa các biến đầu vào.





Hình 8. Biểu đồ SHAP tương tác giữa các biến đầu vào: (a) X7-X1; (b) X7-X2; (c) X7-X4

5. KẾT LUẬN

Nghiên cứu đã xây dựng thành công mô hình học máy nhằm phân tích và dự báo đặc tính lưu biến của nhựa đường biến tính sợi PP trên cơ sở bộ dữ liệu gồm 132 mẫu thí nghiệm được tổng hợp từ nhiều nguồn tài liệu uy tín. Trong số 5 thuật toán được khảo sát, mô hình CAT cho thấy hiệu suất vượt trội và ổn định nhất, đạt $R^2 = 0,995$ và $RMSE = 1,000$ trên tập kiểm tra, đồng thời duy trì khả năng tổng quát hóa tốt trên tập xác thực ($R^2 = 0,875$; $RMSE = 3,000$). Kết quả này khẳng định tính hiệu quả của phương pháp học máy trong mô hình hóa các mối quan hệ phi tuyến và phức tạp giữa các thông số vật liệu và công nghệ chế tạo.

Một đóng góp quan trọng của nghiên cứu là việc kết hợp mô hình học máy với phương pháp SHAP nhằm giải thích cơ chế ảnh hưởng của các biến đầu vào. Phân tích cho thấy mô-đun cắt phức tạp của nhựa ban đầu (X7) là yếu tố chi phối chính, trong khi tỷ lệ sợi (X2) cùng các thông số như độ bền kéo sợi (X1) và nhiệt độ trộn (X4) đóng vai trò tương tác và khuếch đại ảnh hưởng theo cơ chế phi tuyến. Điều này không chỉ nâng cao tính minh bạch của mô hình mà còn cung cấp bằng chứng định lượng về vai trò của từng yếu tố trong việc cải thiện đặc tính lưu biến của nhựa đường biến tính.

Về mặt ứng dụng, nghiên cứu đã cung cấp một công cụ dự báo có độ tin cậy cao, góp phần rút ngắn thời gian thí nghiệm, giảm chi phí khảo sát vật liệu và hỗ trợ tối ưu

hóa thiết kế cấp phối nhựa đường biến tính sợi PP. Đây là cơ sở khoa học quan trọng cho việc ứng dụng trí tuệ nhân tạo trong nghiên cứu vật liệu mặt đường, đồng thời mở ra tiềm năng mở rộng cho các loại phụ gia polymer khác trong tương lai.

TÀI LIỆU THAM KHẢO

- [1] B. Lei, W. Li, Z. Luo, V.W. Tam, W. Dong, K. Wang. Performance enhancement of permeable asphalt mixtures with recycled aggregate for concrete pavement application. *Frontiers in Materials*, vol. 7, 16 September 2020, doi.org/10.3389/fmats.2020.00253.
- [2] Z. Deng, W. Li, W. Dong, Z. Sun, J. Kodikara, D. Sheng. Multifunctional asphalt concrete pavement toward smart transport infrastructure: Design, performance and perspective. *Composites Part B: Engineering*, vol. 265, 110937, October 2023, doi.org/10.1016/j.compositesb.2023.110937.
- [3] C. Li, H. Liu, Y. Xiao, J. Li, T. Wang, L. Peng. Modification and enhancing contribution of fiber to asphalt binders and their corresponding mixtures: A study of viscoelastic properties. *Materials*, vol. 16, July 2023, doi.org/10.3390/ma16165727.
- [4] H. Xiaoming, I.B. Eldouma. Experimental study to determine the most preferred additive for improving asphalt performance using polypropylene, crumb rubber and tafpack super in medium and high-temperature range. *Applied Sciences*, vol. 9, April 2019, doi.org/10.3390/app9081567.
- [5] H.-G.T. Hoang, H.-L. Nguyen, N.H. Tran, H.-B. Ly. Evaluation of the influence of graphene oxide on asphalt binder physical and rheological properties. *Proceedings of the Institution of Mechanical Engineers, Part L: Journal of Materials: Design and Applications*, vol. 238, pp. 133-147, June 2024, doi.org/10.1177/14644207231186610.
- [6] H.-G.T. Hoang, H.-V.T. Mai, H.L. Nguyen, H.-B. Ly. Application of extreme gradient boosting in predicting the viscoelastic characteristics of graphene oxide modified asphalt at medium and high temperatures. *Front. Struct. Civ. Eng.*, vol. 18, pp. 899-917, June 2024, doi.org/10.1007/s11709-024-1025-y.
- [7] A. Behnood, E.M. Golafshani. Predicting the dynamic modulus of asphalt mixture using machine learning techniques: An application of multi biogeography-based programming. *Construction and Building Materials*, vol. 266, January 2021, doi.org/10.1016/j.conbuildmat.2020.120983.
- [8] P. Yeh, Y. Nien, J. Chen, W. Chen, J. Chen. Thermal and rheological properties of maleated polypropylene modified asphalt. *Polymer Engineering and Science*, vol. 45, pp. 1152-1158, August 2005, doi.org/10.1002/pen.20386.
- [9] L. Guo, W. Xu, Y. Zhang, W. Ji, S. Wu. Selecting the best performing modified asphalt based on rheological properties and microscopic analysis of RPP/SBS modified asphalt. *Materials*, vol. 15, December 2022, doi.org/10.3390/ma15238616.
- [10] Q. Zhang, D. Hou, Z. Li, H. Wang, S. Dong. Evaluation of the thermal stability and micro-modification mechanism of SBR/PP-modified asphalt. *Polymers*, vol. 16, February 2024, doi.org/10.3390/polym16040456.
- [11] P.M. Kathari. Rheological Properties of Polypropylene Reinforced Asphalt Binder. *Transp. Infrastruct. Geotech.* vol. 3, pp. 109-126, April 2016, doi.org/10.1007/s40515-016-0033-3.
- [12] H.-G.T. Hoang, T.-A. Nguyen, H.L. Nguyen, H.-B. Ly. Gradient boosting approach to predict complex modulus of GO-modified asphalt at low and medium temperature. *Innov. Infrastruct. Solut.*, vol. 9, April 2024, doi.org/10.1007/s41062-024-01466-w.