

Ứng dụng mạng nơ-ron sâu dự báo góc ma sát trong của đất loại sét lẫn dăm sạn

Application of deep neural networks for predicting internal friction angle of gravelly clay soils

> THS.NCS CAO TRỌNG CÔNG, PGS.TS NGUYỄN ĐỨC MẠNH*,
THS NGUYỄN ĐÌNH DŨNG, THS LÊ ANH ĐỨC

Trường Đại học Giao thông vận tải

*Email: nguyenducmanh@utc.edu.vn

TÓM TẮT

Góc ma sát trong (φ) là thông số sức kháng cắt cơ bản trong thiết kế và đánh giá ổn định công trình địa kỹ thuật. Nghiên cứu này ứng dụng mạng nơ-ron sâu (DNN) dự báo φ của đất loại sét lẫn dăm sạn dựa trên 8.353 mẫu thí nghiệm từ các tuyến giao thông miền núi phía Bắc Việt Nam. Mô hình sử dụng 10 thông số đầu vào gồm thành phần hạt (dăm cuội, sạn sỏi, cát, bụi, sét), giới hạn Atterberg (giới hạn chảy, giới hạn dẻo) và đặc trưng vật lý (khối lượng thể tích tự nhiên, khối lượng thể tích khô, tỷ trọng hạt). Dữ liệu chia theo tỷ lệ 70/30 cho huấn luyện và kiểm chứng. Kết quả đạt $R^2 = 0,958$ (huấn luyện) và $R^2 = 0,901$ (kiểm chứng), $MAE = 1,244^0$ và $RMSE = 1,771^0$. Phân tích SHAP xác nhận khối lượng thể tích khô là thông số chi phối dự báo (giá trị ảnh hưởng 3,58), phù hợp với lý thuyết cơ học đất về vai trò trạng thái chặt-rỗng đối với sức kháng cắt. Mô hình có thể ứng dụng làm công cụ ước lượng nhanh φ , hỗ trợ thiết kế sơ bộ và đánh giá ổn định công trình.

Từ khóa: Góc ma sát trong; mạng nơ-ron sâu; địa kỹ thuật; SHAP; học máy.

ABSTRACT

The internal friction angle (φ) is a fundamental parameter governing soil shear strength and geotechnical stability. This study developed a deep neural network (DNN) to predict φ of gravelly clay soils using 8,353 samples from northern mountainous Vietnam. Ten input parameters were employed: particle size fractions (gravel, pebble, sand, silt, clay), Atterberg limits (liquid and plastic limits) and physical properties (natural bulk density, dry bulk density, specific gravity). The 70/30 train-test split yielded $R^2 = 0.958$ (training) and $R^2 = 0.901$ (testing), with $MAE = 1.244^0$ and $RMSE = 1.771^0$. SHAP analysis identified dry bulk density as the most influential predictor (SHAP value = 3.58), followed by liquid limit (1.00), natural bulk density (0.90) and plastic limit (0.78), while particle size parameters showed minimal influence (≤ 0.32), consistent with geotechnical theory on soil state and shear resistance. The model offers a rapid estimation tool for preliminary design and stability assessment.

Keywords: Internal friction angle; deep neural network; geotechnical engineering; SHAP; machine learning; Vietnam.

1. ĐẶT VẤN ĐỀ

Góc ma sát trong (φ) là thông số sức kháng cắt cơ bản, quyết định sức kháng cắt của đất theo tiêu chuẩn Mohr-Coulomb và ảnh hưởng trực tiếp đến ổn định nền móng, mái dốc và tường chắn [1, 2]. Giá trị φ phụ thuộc vào kích thước hạt, trọng lượng riêng, độ rỗng và độ ẩm. Phương pháp xác định φ gồm cắt phẳng và nén ba trục; cắt phẳng phổ biến hơn do thiết bị đơn giản và chi phí thấp nhưng tồn tại hạn chế về phân bố ứng suất không đồng đều, trong khi nén ba trục cho kết quả chính xác hơn song đòi hỏi thời gian dài và chi phí cao.

Gần đây, các phương pháp trí tuệ nhân tạo (AI) được áp dụng rộng rãi để dự báo thông số cơ lý đất, bao gồm ANN [3], rừng ngẫu nhiên [4], mô hình lai tăng cường [5]. Zhu và cộng sự [6] sử dụng BPNN và SVR dự báo φ trên 83 mẫu, trong đó SVR đạt hiệu suất tốt

nhất. Anka và cộng sự [5] đạt $R^2 \approx 0,94$ với sai số (1-2)^o. Tại Việt Nam, Nguyễn Thùy Anh và cộng sự [3] xây dựng ANN dự báo φ đạt $R = 0,885$ trên 145 mẫu; Trần Văn Quân và cộng sự [7] áp dụng Random Forest trên 131 bộ dữ liệu; Phạm Tuấn Anh và cộng sự [8] chứng minh DNN kết hợp PSO cải thiện R^2 tăng 1,83% so với DNN thông thường.

Tuy nhiên, phần lớn nghiên cứu tập trung vào mô hình học máy truyền thống (ANN, SVM, Random Forest), trong khi khai thác kiến trúc học sâu cho dự báo thông số đất còn hạn chế [9]. Các mô hình nông khó biểu diễn mối quan hệ phi tuyến phức tạp giữa nhiều chỉ tiêu cơ lý đồng thời và nhiều nghiên cứu xây dựng trên tập dữ liệu nhỏ (vài trăm mẫu), hạn chế khả năng khái quát hóa.

Nghiên cứu này đề xuất mô hình DNN dự báo góc ma sát trong dựa trên bộ dữ liệu lớn (8.353 mẫu) phản ánh đa dạng điều kiện địa

chất. Kiến trúc DNN nhiều tầng cho phép tự động trích xuất đặc trưng tiềm ẩn và học các quan hệ phi tuyến bậc cao [10], kết hợp phân tích SHAP đánh giá mức độ ảnh hưởng của từng thông số đầu vào nhằm tăng tính minh bạch và khả năng diễn giải của mô hình [12].

2. CƠ SỞ DỮ LIỆU

Dữ liệu được thu thập từ các tuyến giao thông trọng điểm khu vực miền núi phía Bắc Việt Nam gồm các tuyến Hòa Bình - Sơn La, Nội Bài - Lào Cai, Tuyên Quang - Hà Giang, Hà Nội - Lạng Sơn, Đồng Đăng - Trà Lĩnh, Hạ Long - Vân Đồn cùng các mạng lưới đường khu vực liên tỉnh. Khu vực nghiên cứu có địa hình phân cắt mạnh, cấu trúc địa chất đa dạng, chịu ảnh hưởng khí hậu nhiệt đới gió mùa, dẫn đến sự thay đổi đáng kể về thành phần hạt và tính chất cơ học của đất [9].

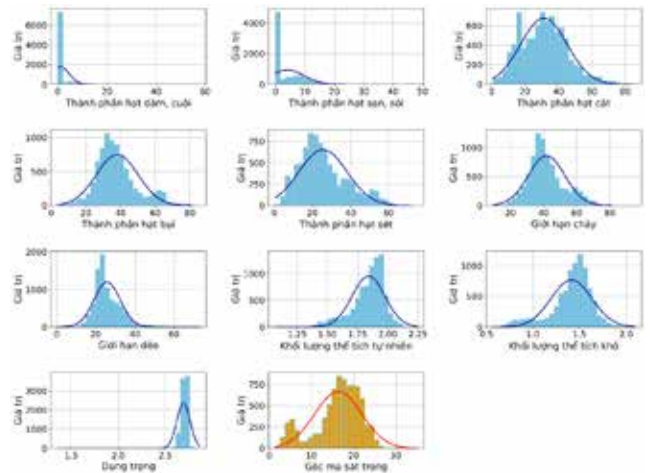
Tổng cộng 8.353 mẫu đất loại sét lẫn dăm sạn được thu thập và thí nghiệm. Bộ dữ liệu gồm 10 biến đầu vào: Thành phần hạt dăm cuội, sạn sỏi, cát, bụi, sét; giới hạn chảy; giới hạn dẻo; khối lượng thể tích tự nhiên; khối lượng thể tích khô; dung trọng. Biến đầu ra là góc ma sát trong được xác định qua thí nghiệm cắt trực tiếp theo Tiêu chuẩn TCVN 4199:2012. Bảng 1 trình bày thống kê mô tả các thông số.

Bảng 1. Phân tích thống kê các thông số đầu vào và đầu ra của mô hình - giữ nguyên

Thông số	Ký hiệu	Đơn vị	Giá trị trung bình	Giá trị nhỏ nhất	Giá trị lớn nhất	Độ lệch chuẩn
Thông số đầu vào						
Thành phần hạt dăm, cuội	-	%	0,995	0	57,5	3,57
Thành phần hạt sạn, sỏi	-	%	3,995	0	48	5,679
Thành phần hạt cát	-	%	31,34	1	85,3	13,835
Thành phần hạt bụi	-	%	38,101	3,9	85,2	12,03
Thành phần hạt sét	-	%	25,57	0,58	74,8	12,625
Giới hạn chảy	W_L	%	41,784	9,5	94,7	10,998
Giới hạn dẻo	W_p	%	25,786	0,617	72,75	6,702
Khối lượng thể tích tự nhiên	ρ	g/cm^3	1,841	1,07	2,24	0,136
Khối lượng thể tích khô	ρ_d	g/cm^3	1,414	0,571	2,09	0,219
Tỷ trọng hạt	Δ		2,694	1,356	2,86	0,071
Thông số đầu ra						
Góc ma sát trong	φ	($^\circ$)	16,371	1,433	35,033	5,591

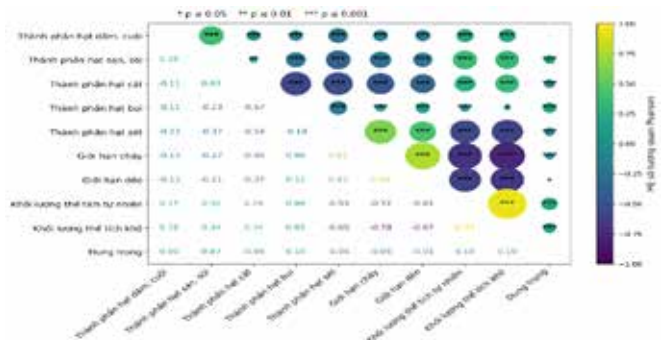
Phân tích thống kê cho thấy, nhóm hạt mịn (cát, bụi, sét) chiếm tỷ lệ chủ đạo với giá trị trung bình cao hơn so với các thành phần hạt thô, phù hợp với đặc điểm đất sét lẫn dăm sạn vùng miền núi. Các thông số khối lượng thể tích và tỷ trọng hạt có độ phân tán nhỏ (hệ số biến thiên thấp), phản ánh tính ổn định của đặc trưng vật lý. Cần lưu ý rằng, giá trị nhỏ nhất của tỷ trọng hạt ($\Delta = 1,356$) thấp hơn đáng kể so với phạm vi thông thường của khoáng vật đất (2,6 - 2,8), có thể liên quan đến mẫu đất chứa nhiều chất hữu cơ hoặc sai số thí nghiệm. Tuy nhiên, giá trị này đã được giữ lại trong bộ dữ liệu do nằm ngoài ngưỡng loại bỏ ngoại lai.

Phân bố thống kê của các thông số (Hình 1) cho thấy bộ dữ liệu có tính đại diện và đa dạng, đáp ứng yêu cầu xây dựng mô hình học máy.



Hình 1. Phân bố thống kê của các thông số đầu vào

Ma trận tương quan Pearson (Hình 2) cho thấy các thông số thành phần hạt có mức tương quan thấp đến trung bình với nhau, trong khi khối lượng thể tích tự nhiên (ρ) và khối lượng thể tích khô (ρ_d) thể hiện tương quan thuận mạnh - điều này phù hợp với bản chất vật lý vì cả hai đều phản ánh trạng thái chặt-rỗng của đất. Sự tương quan cao giữa hai thông số này cũng chứng tỏ tính nhất quán của dữ liệu thí nghiệm.



Hình 2. Ma trận tương quan Pearson của các thông số đầu vào

3. PHƯƠNG PHÁP NGHIÊN CỨU

3.1. Mạng nơ-ron sâu (DNN)

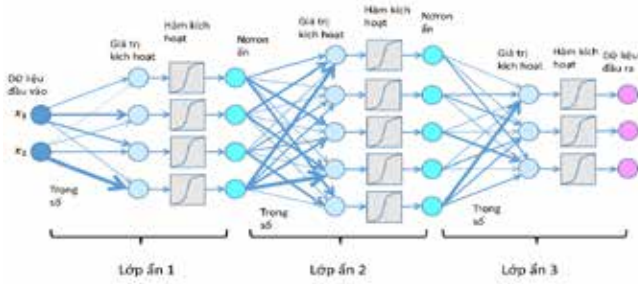
Mạng nơ-ron sâu (DNN) là kiến trúc mở rộng của mạng nơ-ron nhân tạo truyền thống, sử dụng nhiều lớp ẩn xếp chồng để học các biểu diễn phân cấp của dữ liệu [10]. Cụ thể, dữ liệu đầu vào được lan truyền tuần tự qua các lớp nơ-ron liên kết đầy đủ, tại mỗi lớp thực

hiện phép biến đổi affine (nhân ma trận trọng số và cộng vector bias), sau đó áp dụng hàm kích hoạt phi tuyến (Hình 3).

Quá trình lan truyền tiến tại lớp l được biểu diễn:

$$a^{(l)} = f(W^{(l)} \cdot a^{(l-1)} + b^{(l)}) \quad (1)$$

Trong đó: $a^{(l-1)}$ là đầu ra lớp trước; $W^{(l)}$ là ma trận trọng số; $b^{(l)}$ là vector bias; f là hàm kích hoạt phi tuyến. Nhờ cấu trúc nhiều lớp ẩn, DNN có khả năng trích xuất các đặc trưng tiềm ẩn theo cấu trúc phân cấp từ thấp đến cao [10].



Hình 3. Sơ đồ kiến trúc mạng nơ-ron sâu (DNN)

3.2. Chỉ số đánh giá

Hiệu suất dự báo của mô hình được đánh giá thông qua ba chỉ số [11]: Hệ số xác định (R^2), sai số tuyệt đối trung bình (MAE) và sai số bình phương trung bình (RMSE). Hệ số R^2 phản ánh tỷ lệ phương sai của biến phụ thuộc được giải thích bởi mô hình, nhận giá trị trong khoảng $[0, 1]$; R^2 càng gần 1 thì mô hình càng chính xác. Các chỉ số MAE và RMSE đo lường mức độ sai lệch trung bình giữa giá trị dự báo và giá trị thực đo, giá trị càng nhỏ phản ánh hiệu suất càng cao.

$$MAE = \frac{\sum_{i=1}^n |h_i - t_i|}{n} \quad (2)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (h_i - t_i)^2}{\sum_{i=1}^n (h_i - \bar{h})^2} \quad (3)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (h_i - t_i)^2}{n}} \quad (4)$$

Trong đó: n là tổng số dữ liệu; h_i là giá trị thực tế; t_i là giá trị dự báo; \bar{h} là giá trị trung bình thực tế.

3.3. Phương pháp SHAP

SHAP (SHapley Additive exPlanations) là phương pháp giải thích mô hình học máy dựa trên lý thuyết giá trị Shapley trong lý thuyết trò chơi hợp tác [12]. Giá trị SHAP của mỗi thông số phản ánh mức độ đóng góp biên (marginal contribution) của thông số đó đối với kết quả dự báo, đồng thời thể hiện chiều hướng tác động (tăng hoặc giảm giá trị dự báo). Ưu điểm nổi bật của SHAP so với các phương pháp đánh giá tầm quan trọng khác (như permutation importance) là khả năng phân tích cả ở mức cục bộ (từng mẫu) và tổng thể (toàn bộ tập dữ liệu), đảm bảo tính nhất quán toán học và tăng khả năng diễn giải kết quả mô hình.

3.4. Xử lý dữ liệu và huấn luyện

Các thông số đầu vào có đơn vị và phạm vi giá trị khác nhau, do đó dữ liệu được chuẩn hóa bằng phương pháp Min-Max scaling

về khoảng $[0, 1]$ nhằm đảm bảo các biến đóng góp đồng đều trong quá trình huấn luyện. Để kiểm soát hiện tượng quá khớp, kỹ thuật dừng sớm được áp dụng: Quá trình huấn luyện tự động dừng khi sai số trên tập kiểm chứng không cải thiện sau một số vòng lặp liên tiếp. Bộ dữ liệu được chia ngẫu nhiên theo tỷ lệ 70% cho huấn luyện (5.847 mẫu) và 30% cho kiểm chứng (2.506 mẫu).

4. KẾT QUẢ VÀ THẢO LUẬN

4.1. Kiến trúc và huấn luyện mô hình

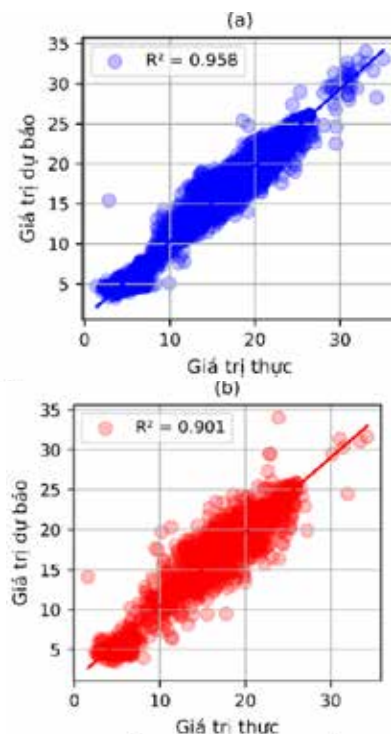
Mạng nơ-ron sâu được thiết kế gồm một lớp đầu vào (10 nơ-ron tương ứng với 10 thông số), ba lớp ẩn với số nơ-ron giảm dần và một lớp đầu ra (1 nơ-ron - giá trị φ dự báo). Tất cả các lớp ẩn sử dụng hàm kích hoạt ReLU (Rectified Linear Unit). Mô hình được huấn luyện bằng thuật toán tối ưu Adam với tốc độ học khởi tạo 0,001 và kích thước batch 32. Cơ chế điều chỉnh tốc độ học tự động (ReduceLROnPlateau) được tích hợp nhằm đảm bảo quá trình hội tụ ổn định khi giá trị hàm mất mát không tiếp tục giảm.

4.2. Hiệu suất dự báo

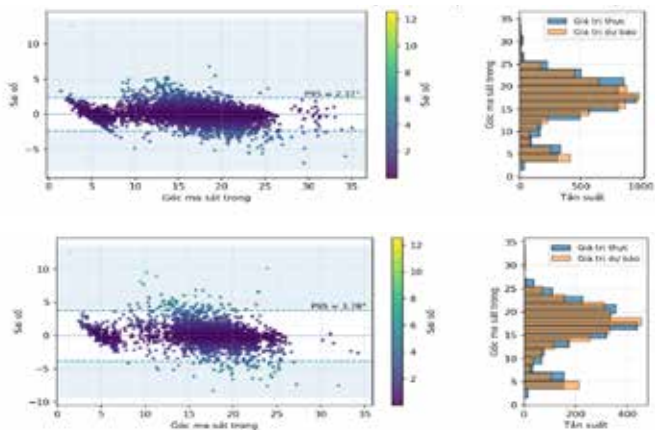
Kết quả đánh giá hiệu suất được trình bày trong Bảng 2. Mô hình DNN đạt $R^2 = 0,958$ trên tập huấn luyện và $R^2 = 0,901$ trên tập kiểm chứng. Chênh lệch R^2 giữa hai tập ($\Delta R^2 = 0,057$) cho thấy mô hình tồn tại hiện tượng quá khớp ở mức nhẹ, tuy nhiên giá trị R^2 trên tập kiểm chứng vẫn duy trì ở mức cao ($> 0,9$), phản ánh khả năng tổng quát hóa chấp nhận được. Sai số dự báo trên tập kiểm chứng đạt MAE = 1,244° và RMSE = 1,771°, mức sai số này nằm trong phạm vi chấp nhận cho ước lượng sơ bộ trong thực hành địa kỹ thuật (thường cho phép sai số 1 - 3° đối với góc ma sát trong).

Bảng 2. Hiệu suất dự báo của mô hình

Thông số	Tập huấn luyện	Tập kiểm chứng
R^2	0,958	0,901
MAE	0,805	1,244
RMSE	1,139	1,771



Hình 4. Hệ số xác định R^2 của mô hình trên tập huấn luyện và kiểm chứng

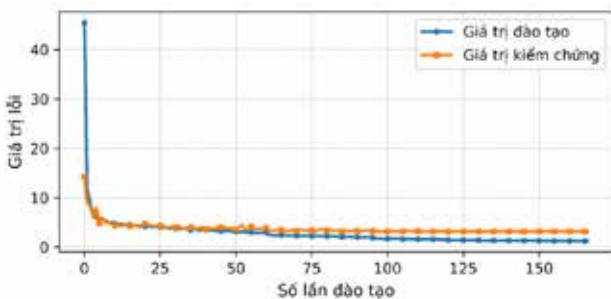


Hình 5. Phân bố sai số dự báo của mô hình DNN trên (a) tập huấn luyện và (b) tập kiểm chứng

Biểu đồ tương quan giữa giá trị thực đo và dự báo (Hình 4) cho thấy trên tập huấn luyện, các điểm dữ liệu phân bố tập trung sát đường hồi quy 1:1, phản ánh mô hình nắm bắt tốt xu thế dữ liệu. Trên tập kiểm chứng, mức độ phân tán tăng nhẹ nhưng các điểm vẫn bám sát đường hồi quy với xu thế tuyến tính rõ ràng, xác nhận khả năng dự báo ổn định của mô hình trên dữ liệu chưa được sử dụng trong huấn luyện.

Phân bố sai số trong Hình 5 cho thấy trên tập huấn luyện, sai số tập trung quanh 0 trong khoảng $\pm 2,37^\circ$, phản ánh độ chính xác cao. Trên tập kiểm chứng, khoảng sai số mở rộng lên $\pm 3,78$ nhưng phân bố vẫn đối xứng quanh 0, cho thấy mô hình không có xu hướng thiên lệch hệ thống (bias) trong dự báo.

Biểu đồ giá trị hàm mất mát theo số vòng lặp huấn luyện (Hình 6) cho thấy sai số trên cả hai tập giảm dần và ổn định, không xuất hiện hiện tượng phân kỳ giữa tập huấn luyện và kiểm chứng - xác nhận cơ chế dừng sớm hoạt động hiệu quả trong kiểm soát quá khớp.

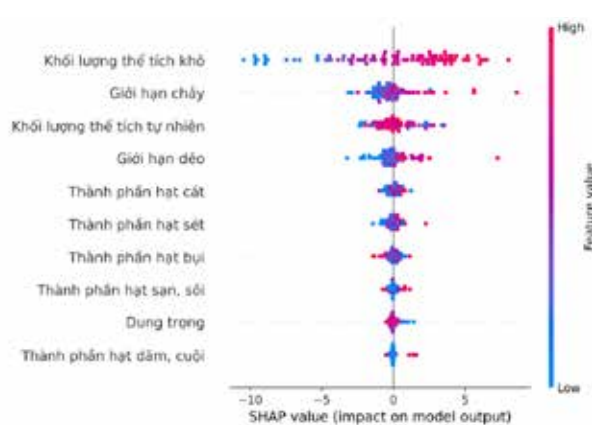


Hình 6. Giá trị lỗi theo số lần đào tạo đối với tập huấn luyện và tập kiểm chứng

4.3. Phân tích SHAP

Biểu đồ phụ thuộc SHAP trong Hình 7 cho thấy khối lượng thể tích khô là thông số có ảnh hưởng mạnh nhất, tiếp theo là giới hạn chảy, khối lượng thể tích tự nhiên và giới hạn dẻo. Các thông số này có miền giá trị SHAP rộng, phản ánh vai trò chi phối trong quá trình dự báo.

Kết quả mức độ quan trọng trong Hình 8 xác nhận khối lượng thể tích khô có ảnh hưởng cao nhất (3,58), tiếp theo là giới hạn chảy (1,00), khối lượng thể tích tự nhiên (0,90) và giới hạn dẻo (0,78). Các thông số liên quan đến thành phần hạt có giá trị ảnh hưởng nhỏ ($\leq 0,32$). Kết quả này phù hợp với lý thuyết địa kỹ thuật khi các đặc trưng trạng thái và tính chất cơ lý của đất đóng vai trò quyết định đến khả năng kháng cắt.



Hình 7. Biểu đồ phụ thuộc từng phần theo SHAP

Kết quả SHAP cho thấy khối lượng thể tích khô phản ánh trực tiếp trạng thái chặt-rỗng của đất: Đất có khối lượng thể tích khô cao tương ứng với hệ số rỗng thấp, khoảng cách giữa các hạt nhỏ, diện tích tiếp xúc giữa các hạt lớn, dẫn đến tăng ma sát giữa các hạt và tăng góc ma sát trong [1, 2]. Ảnh hưởng mạnh của giới hạn chảy và giới hạn dẻo phản ánh vai trò của khoáng vật sét: Đất có giới hạn Atterberg cao thường chứa nhiều khoáng sét hoạt tính (montmorillonite, illite), tạo lớp nước hấp phụ dày quanh hạt sét, làm giảm ma sát giữa các hạt và do đó giảm góc ma sát trong. Ngược lại, các thành phần hạt thô (dăm cuội, sạn sỏi, cát) mặc dù bản thân có ma sát cao, nhưng trong hệ đất sét lẫn dăm sạn, vai trò chủ đạo thuộc về pha sét liên kết giữa các hạt thô, giải thích tại sao tỷ lệ thành phần hạt thô có ảnh hưởng SHAP thấp.



Hình 8. Mức độ ảnh hưởng của các thông số theo giá trị SHAP

4.4. Thảo luận

Mô hình DNN đạt hiệu suất dự báo cao với $R^2 = 0,901$ trên tập kiểm chứng, vượt trội so với nhiều nghiên cứu trước đó sử dụng tập dữ liệu nhỏ hơn. So sánh với Nguyễn Thùy Anh và cộng sự [3] ($R = 0,885$, 145 mẫu) và Anka và cộng sự [5] ($R^2 \approx 0,94$, quy mô mẫu nhỏ), kết quả nghiên cứu này đáng tin cậy hơn nhờ bộ dữ liệu quy mô lớn (8.353 mẫu) phủ đa dạng điều kiện địa chất. Tuy nhiên, chênh lệch R^2 giữa tập huấn luyện và kiểm chứng ($\Delta R^2 = 0,057$) cho thấy mô hình vẫn tồn tại hiện tượng quá khớp nhẹ. Các nghiên cứu tiếp theo cần xem xét áp dụng kỹ thuật chính quy hóa (regularization) hoặc dropout để cải thiện khả năng tổng quát hóa.

Với mức sai số MAE = 1,244° và RMSE = 1,771° trên tập kiểm chứng, mô hình có thể đóng vai trò công cụ ước lượng nhanh góc ma sát trong dựa trên các thông tin khảo sát sẵn có, hỗ trợ kỹ sư định hướng giá trị tham khảo ở giai đoạn thiết kế sơ bộ. Các sai số lớn chỉ xuất hiện cục bộ tại miền giá trị biên (φ rất nhỏ hoặc rất lớn),

phù hợp với thực tiễn khi các giá trị cực trị thường cần được kiểm chứng bổ sung bằng thí nghiệm. Mô hình không nhằm thay thế các thí nghiệm truyền thống xác định sức kháng cắt, mà đóng vai trò công cụ hỗ trợ ra quyết định, góp phần tối ưu hóa công tác khảo sát và thiết kế địa kỹ thuật.

Cần lưu ý rằng nghiên cứu sử dụng phương pháp chia ngẫu nhiên tập dữ liệu, chưa xem xét tương quan không gian giữa các mẫu cùng tuyến đường hoặc cùng hố khoan. Ngoài ra, việc chưa đưa vào so sánh mô hình cơ sở (baseline) đơn giản như hồi quy tuyến tính bội là một hạn chế cần được bổ sung trong các nghiên cứu tiếp theo.

5. KẾT LUẬN

Nghiên cứu đã xây dựng và kiểm chứng mô hình mạng nơ-ron sâu (DNN) dự báo góc ma sát trong của đất loại sét lẫn dăm sạn dựa trên 8.353 mẫu thí nghiệm từ khu vực miền núi phía Bắc Việt Nam. Các kết luận chính bao gồm:

- Mô hình DNN đạt hiệu suất cao với $R^2 = 0,958$ trên tập huấn luyện và $R^2 = 0,901$ trên tập kiểm chứng, cùng sai số MAE = 1,244° và RMSE = 1,771°, phản ánh khả năng tổng quát hóa tốt trên bộ dữ liệu lớn và đa dạng địa chất.

- Phân tích SHAP cho thấy khối lượng thể tích khô là thông số chi phối kết quả dự báo (giá trị ảnh hưởng 3,58), phù hợp với lý thuyết cơ học đất về vai trò quyết định của trạng thái chặt-rỗng đối với sức kháng cắt.

- Mô hình có thể ứng dụng làm công cụ ước lượng nhanh góc ma sát trong, hỗ trợ thiết kế sơ bộ và đánh giá ổn định công trình địa kỹ thuật trong giai đoạn khảo sát ban đầu.

- Hướng phát triển tiếp theo bao gồm mở rộng nghiên cứu cho các loại đất khác nhau, tích hợp thêm các thông số môi trường (độ ẩm, lịch sử ứng suất), áp dụng kỹ thuật kiểm chứng chéo không gian (spatial cross-validation) và so sánh với mô hình cơ sở để nâng cao độ tin cậy và khả năng ứng dụng.

Lời cảm ơn: Nghiên cứu này được tài trợ bởi Đề tài nghiên cứu khoa học cấp Bộ Giáo dục và Đào tạo năm 2026, mã số B2026-GHA-01, do Trường Đại học Giao thông vận tải chủ trì.

TÀI LIỆU THAM KHẢO

[1] Hoàng Việt Hùng, Nguyễn Quang Tuấn, Ngô Văn Linh, Trịnh Minh Thu. Cơ học đất nâng cao. NXB. Bách Khoa Hà Nội, 2025.

[2] Ngô Thị Thanh Hương. Cơ học đất. NXB. Xây dựng, 2017.

[3] T.-A. Nguyen, H.-B. Ly, B.T. Pham. Backpropagation neural network-based machine learning model for prediction of soil friction angle. *Mathematical Problems in Engineering*, vol. 2020, no. 1, 8845768, 2020, doi: 10.1155/2020/8845768.

[4] M. Hiba, A.F. Ibrahim, S. Elkatatny, A. Ali. Prediction of cohesion and friction angle from well-logging data using decision tree and random forest. *Arabian Journal of Geosciences*, vol. 15, no. 1, 26, 2022, doi: 10.1007/s12517-021-09154-0.

[5] M.L.Y. Ankah, S. Adjei-Yeboah, Y.Y. Ziggah, E.N. Asare. Advanced hybrid machine learning models with explainable AI for predicting residual friction angle in clay soils. *Scientific Reports*, vol. 15, no. 1, 25868, 2025, doi: 10.1038/s41598-025-10069-3.

[6] L. Zhu, Q. Liao, Z. Wang, J. Chen, Z. Chen, Q. Bian, Q. Zhang. Prediction of soil shear strength parameters using combined data and different machine learning models. *Applied Sciences*, vol. 12, no. 10, 5100, 2022, doi: 10.3390/app12105100.

[7] V.Q. Tran, T.A. Pham. Investigation of input number effect on performance prediction of soil friction angle using random forest. Trong: *CIGOS 2021, Emerging Technologies and Applications for Green Infrastructure*, Springer, pp. 1859-1866, 2021, doi: 10.1007/978-981-16-7160-9_188.

[8] T.A. Pham, V.Q. Tran, H.-L.T. Vu. Evolution of deep neural network architecture using particle swarm optimization to improve the performance in determining the friction

angle of soil. *Mathematical Problems in Engineering*, vol. 2021, no. 1, 5570945, 2021, doi: 10.1155/2021/5570945.

[9] H. Liu, H. Su, L. Sun, D. Dias-da-Costa. State-of-the-art review on the use of AI-enhanced computational mechanics in geotechnical engineering. *Artificial Intelligence Review*, vol. 57, no. 8, 196, 2024, doi: 10.1007/s10462-024-10836-w.

[10] S. Wang, Q. Wang, N. Bailey, J. Zhao. Deep neural networks for choice analysis: A statistical learning theory perspective. *Transportation Research Part B: Methodological*, vol. 148, pp. 60-81, 2021, doi: 10.1016/j.trb.2021.03.011.

[11] D. Chicco, M.J. Warrens, G. Jurman. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, vol. 7, e623, 2021, doi: 10.7717/peerj-cs.623.

[12] G. Van den Broeck, A. Lykov, M. Schleich, D. Suciu. On the tractability of SHAP explanations. *Journal of Artificial Intelligence Research*, vol. 74, pp. 851-886, 2022, doi: 10.1613/jair.1.13283.