

ỨNG DỤNG TRÍ TUỆ NHÂN TẠO TRONG DỰ ĐOÁN SỨC CHỐNG CẮT CỦA ĐẤT SAU BIẾN DẠNG

Nguyễn Quang Hùng

Đại học Thủy lợi

Trần Văn Quân

Đại học Công nghệ Giao thông vận tải

Tóm tắt: Sau khi bị biến dạng do các tác động như xói mòn, sạt lở, đất sẽ thay đổi đáng kể sức chống cắt. Do vậy công tác dự báo suy giảm cường độ chống cắt của các lớp đất này từ đó giúp dự đoán được khả năng tái diễn sạt lở mất ổn định với các lớp đất bị biến dạng này là một công tác hết sức cần thiết. Trong bài báo này sẽ ứng dụng trí thông minh nhân tạo mà cụ thể là thuật toán rừng ngẫu nhiên (RF) để dự báo sức chống cắt còn lại của đất sau biến dạng. Để thực hiện việc mô phỏng, 131 dữ liệu thí nghiệm đã được thu thập từ công bố quốc tế. Bộ dữ liệu bao gồm bốn biến đầu vào là giới hạn chảy LL, chỉ số dẻo PI, độ lệch biểu đồ phân loại Casagrande ΔPI , hàm lượng sét CF. Việc đánh giá các mô hình được thực hiện và so sánh trên tập dữ liệu huấn luyện (70% dữ liệu) và tập dữ liệu kiểm chứng (30% dữ liệu còn lại) bằng các tiêu chí là hệ số tương quan Pearson R và sai số RMSE. Kết quả của nghiên cứu cho thấy mô hình rừng ngẫu nhiên khả thi trong việc xác định sức chống cắt còn lại của đất sau biến dạng của đất với hệ số tương quan cho mô hình huấn luyện là 0.97 và kiểm chứng là 0.78. Đồng thời, mô hình rừng cây ngẫu nhiên có thể chỉ ra tầm quan trọng của từng tính chất của đất đến sức chống cắt còn lại của đất biến dạng, lần lượt theo thứ tự là Giới hạn chảy > Độ lệch biểu đồ phân loại Casagrande ΔPI > Hàm lượng sét > Chỉ số dẻo.

Từ khóa: Trí tuệ nhân tạo (AI), rừng ngẫu nhiên (RF), sức chống cắt, góc ma sát, sạt lở

Summary: After being deformed due to disasters such as erosion, landslides, the soil will significantly change its shear strength. Therefore, it is necessary to forecast the reduction of the shear strength of these soils to predict the possibility of recurrence of unstable erosion with these deformed layers. In this paper, artificial intelligence (RF) will be applied to predict the remaining shear strength of soil after deformation. To perform the simulation, 131 experimental data were collected from literature. The data set consists of four input variables: LL liquid limit, PI plasticity index, Casagrande's classification deviation ΔPI , CF clay content. The evaluation of the models was made and compared on training data set (70% data) and control data set (30% remaining data) by criteria of Pearson correlation coefficient (R) and RMSE error. The results of the study showed that the random forest model is feasible in determining the remaining shear strength of soil after soil deformation with a correlation coefficient for the training model is 0.97 and verified as 0.78. At the same time, the random forest model can show the importance of each soil property to the remaining shear strength of deformed soil, respectively in the order of Liquid Limit > Casagrande classification deviation ΔPI > Clay Fraction > Plasticity index.

Keywords: Artificial Intelligence (AI), Random forest (RF), shear strength, friction angle, landslide.

1. ĐẶT VẤN ĐỀ

Độ ổn định của các sườn dốc tự nhiên hoặc các khu vực lở đất phụ thuộc vào các thông số cường độ chống cắt của lớp đất sét, mức cường độ này thay đổi một cách đáng kể tùy theo trạng

thái của lớp đất sét. Điều này đặc biệt đúng với đất sét có độ nhạy cảm cao, khi cường độ chống cắt của chúng khi ở trạng thái cực đại cao hơn đáng kể so với trạng thái sau biến dạng như là các hiện tượng sạt lở, mất ổn định bờ dốc, xói

Ngày nhận bài: 20/4/2020

Ngày thông qua phản biện: 25/5/2020

Ngày duyệt đăng: 05/6/2020

mòn do mưa. Ở trạng thái sau khi biến dạng, đất sét sẽ trải qua một biến dạng lớn và do đó sẽ không còn nguyên dạng như trước, với các liên kết trước đó bị phá vỡ và kết cấu khối đất sẽ bị phá hủy. Kết quả là, ở các trạng thái sau khi biến dạng, đất sét sẽ mất đi sự kết dính. Hơn nữa, góc ma sát (Φ_r) nhỏ hơn đáng kể so với góc ma sát ở trạng thái cực đại (Φ_p), dẫn đến cường độ chống cắt thấp hơn. Trong các vấn đề địa kỹ thuật liên quan đến sự biến dạng lớn như lở đất, cần sử dụng các tham số cường độ cắt sau biến dạng (lực dính $c_r = 0$ và góc ma sát Φ_r) trong quá trình tính toán, phân tích.

Đã có những nỗ lực để làm rõ mối tương quan giữa góc ma sát ở trạng thái sau biến dạng đất và các thuộc tính chỉ số như giới hạn Atterberg và hàm lượng sét (CF). Skempton [1] đã chỉ ra mối liên quan của giá trị Φ_r với hàm lượng sét. Φ_r giảm theo giới hạn chảy (LL). Mesri và Cepeda-Diaz [2] đã trình bày mối tương quan giữa Φ_r và LL. Colotta và cộng sự [3] đã đưa ra mối tương quan giữa Φ_r và một tham số là một hàm của LL, chỉ số dẻo (PI) và CF. Đối với đất trầm tích, Stark và Eid [4] quan sát thấy rằng các loại khoáng chất và hàm lượng sét chi phối giá trị của Φ_r . Sử dụng giá trị LL như một chỉ số của khoáng vật sét, họ đã đề xuất mối tương quan của Φ_r với LL cho các khoảng giá trị hàm lượng sét khác nhau. Wesley [5] nhận thấy rằng đối với đất sét ở mức nhiệt độ cao thì Φ_r có thể liên quan nhiều hơn tới ΔPI – độ lệch so với đường loại A trong biểu đồ phân loại Casagrande được đưa ra bởi công thức:

$$\Delta PI = PI - 0.73(LL - 20) \quad (1)$$

Sự ảnh hưởng của các yếu tố giới hạn chảy LL, chỉ số dẻo PI và hàm lượng sét CF đến góc ma sát Φ_r , Sridharan và Rao [6] thấy rằng hàm lượng sét ảnh hưởng lớn nhất. Sử dụng dữ liệu từ các mẫu đất được thu thập từ hơn 80 khu vực chịu các thảm họa sạt lở khác nhau, Tiwari và Marui [7] đã đề xuất một biểu đồ hình tam giác để dự đoán góc ma sát của đất sau biến dạng Φ_r , dựa trên các thành phần khoáng vật học. Mô

hình này đã đưa ra dự đoán tốt cho các mẫu mà họ đã thử nghiệm, nhưng đưa ra kết quả dự đoán thấp hơn cho 53 mẫu được thử nghiệm bởi Kaya và Kwong [8], các mẫu thí nghiệm được lấy từ các vụ lở đất diễn ra ở Hawaii. Việc dự đoán góc ma sát thấp hơn so với thực tế là do đối với các loại đất phù sa thường ở trạng thái vô định hình, do vậy chỉ có mối tương quan nhỏ giữa các chỉ số đặc tính của đất và góc ma sát sau biến dạng. Hơn nữa, rất khó để tìm ra diện tích bề mặt cụ thể của phần khoáng vật sét và các thành phần khoáng khác của các mẫu đất sét.

Từ các nghiên cứu trên có thể thấy rằng góc ma sát Φ_r là một hàm phụ thuộc vào các đặc tính của đất như giới hạn chảy LL, chỉ số dẻo PI, độ lệch biểu đồ phân loại Casagrande ΔPI , hàm lượng sét CF. Tuy nhiên, hầu hết các mối quan hệ được phát triển và biểu thị theo dạng biểu đồ. Mối tương quan được đưa ra bởi Tiwari và Marui [9] dựa trên hàm lượng khoáng vật của đất sét nhưng cũng được trình bày dưới dạng biểu đồ. Đối với một kỹ sư địa kỹ thuật, việc tìm ra hàm lượng khoáng vật không phải lúc nào cũng tiện dụng để thực hiện. Do đó, cần phải phát triển một mối tương quan hoặc phương pháp mà trong đó Φ_r có thể được biểu thị theo tất cả các biến có liên quan. Das và Basudhar [9] đã sử dụng mô hình trí tuệ nhân tạo để dự đoán góc ma sát sau biến dạng của đất sét với độ chính xác cao. Nhưng nghiên cứu trên chỉ giới hạn với đất ở một số khu vực cụ thể. Do đó, nhu cầu cấp thiết là phát triển một phương pháp phù hợp và hiệu quả để áp dụng cho các loại đất có nguồn gốc khác nhau.

Do vậy, trong bài báo này một thuật toán rất mạnh trong việc sử dụng trí thông minh nhân tạo AI là mô hình rừng ngẫu nhiên sẽ được ứng dụng để dự đoán góc ma sát của đất ở các trạng thái khác nhau dựa vào các 4 chỉ số của đất là giới hạn chảy LL, chỉ số dẻo PI, độ lệch biểu đồ phân loại Casagrande ΔPI , hàm lượng sét CF. Hệ số tương quan R và sai số RMSE được sử dụng để đánh giá độ chính xác khả năng dự báo của mô

hình. Ngoài ra kết quả của mô hình rừng ngẫu nhiên có thể giúp xác định thứ tự ảnh hưởng 4 chỉ số của đất tới góc ma sát.

2. THIẾT LẬP MÔ HÌNH DỰ BÁO

2.1. Mô hình rừng cây ngẫu nhiên (Random Forest)

Random Forests (RF) chỉ định một họ các phương thức học máy (Machine Learning), bao gồm các thuật toán khác nhau để tạo ra một tập hợp các cây quyết định, như thuật toán Breiman Forest được trình bày bởi Breiman [10] và thường được sử dụng trong tài liệu như một mô hình chuẩn. Bản chất của thuật toán rừng ngẫu nhiên đó là có thể kết hợp được nhiều cây quyết định thay vì chỉ đưa ra lựa chọn dựa trên quyết định của một cây. Từ đó, RF có thể giảm lỗi dự đoán và cải thiện hiệu suất dự đoán. Các bước học tập bao gồm xây dựng một tập hợp các cây quyết định, mỗi nhóm được điều khiển từ một tập hợp con 'bootstrap, từ tập học ban đầu, tức là sử dụng nguyên tắc đóng bao và sử dụng phương pháp cảm ứng cây gọi là cây ngẫu nhiên. Một thuật toán cảm ứng như vậy, thường dựa trên thuật toán cây phân loại và hồi quy [11].

Phương pháp rừng ngẫu nhiên cho phép học song song từ nhiều cây quyết định được xây dựng và huấn luyện ngẫu nhiên với nhiều tập con chứa các mẫu khác nhau. Mỗi cây trong rừng được huấn luyện bởi một tập hợp con với dữ liệu được phân phối ngẫu nhiên theo nguyên tắc đóng bao và cũng có các tính năng ngẫu nhiên. Các kết quả cuối cùng được đưa ra dưới dạng giá trị trung bình của mỗi cây quyết định cho bài toán hồi quy hoặc được xác định bằng kết quả đa số cho bài toán phân loại dữ liệu. Với nhiều ưu điểm của rừng ngẫu nhiên (RF), thuật toán này đã được áp dụng rộng rãi trong nhiều ứng dụng khác nhau [12].

2.2. Đánh giá khả năng dự báo của mô hình

Trong nghiên cứu này, hai tiêu chí được sử dụng là hệ số tương quan (R) (correlation coefficient) và sai số RMSE (Root Mean Square Error) để đánh giá độ chính xác của mô hình rừng ngẫu nhiên đã phát triển. Các giá trị của R và RMSE được ước tính bằng các phương trình sau [13]:

$$RMSE = \sqrt{\frac{1}{S} \sum_{j=1}^S (p_{0,j} - p_{t,j})^2} \quad (4)$$

$$R = \frac{\sum_{j=1}^S (p_{0,j} - \bar{p}_0)(p_{t,j} - \bar{p}_t)}{\sqrt{\sum_{j=1}^S (p_{0,j} - \bar{p}_0)^2 \sum_{j=1}^S (p_{t,j} - \bar{p}_t)^2}} \quad (5)$$

Trong đó: S là số lượng mẫu, p_0 và \bar{p}_0 là giá trị thí nghiệm thực tế và giá trị thí nghiệm thực tế trung bình, p_t và \bar{p}_t là giá trị dự đoán và giá trị dự đoán trung bình, được tính theo mô hình dự báo.

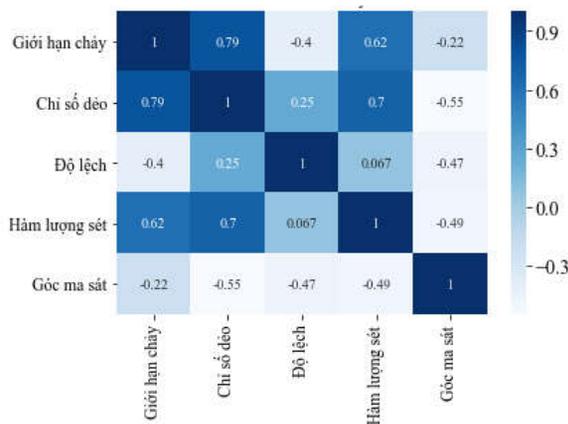
2.3. Thu thập dữ liệu

Trong nghiên cứu này, 131 dữ liệu thực tế về dầm góc ma sát đã được thu thập từ tài liệu công bố tại các tạp chí uy tín trên thế giới [14], [15]. Mô hình rừng ngẫu nhiên sử dụng 4 biến đầu vào bao gồm: (1) giới hạn chảy LL, (2) chỉ số dẻo PI, (3) độ lệch biểu đồ phân loại Casagrande ΔPI , (4) hàm lượng sét CF. Tham số đầu ra được xét tới là góc ma sát của đất. Bộ dữ liệu được sử dụng trong công việc này được chia ngẫu nhiên thành hai bộ dữ liệu con bằng cách sử dụng phân phối thống nhất, trong đó 70% dữ liệu được sử dụng để đào tạo các mô hình rừng ngẫu nhiên và 30% dữ liệu còn lại được dùng trong việc kiểm chứng mô hình. Mô tả về dữ liệu được ghi lại trong bảng 1, giá trị trung bình, độ lệch chuẩn, các giá trị nhỏ nhất, giá trị tại góc phần tư thứ nhất, góc phần tư thứ hai, góc phần tư thứ ba và giá trị lớn nhất. Các mẫu thí nghiệm được tập hợp từ nhiều loại đất chịu các điều kiện thảm họa khác nhau từ các vùng khác nhau nên có sự biến thiên rất lớn về giá trị góc ma sát 5.5° đến 39.0°

Bảng 1: Thống kê dữ liệu đầu vào và đầu ra cho mô hình rừng ngẫu nhiên

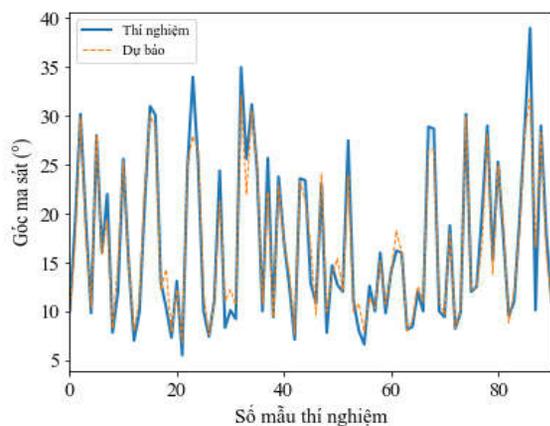
Biến	Giới hạn chảy	Chỉ số dẻo	Độ lệch	Hàm lượng sét	Góc ma sát
------	---------------	------------	---------	---------------	------------

Đơn vị	%	%	%	%	°
Vai trò	Đầu vào	Đầu vào	Đầu vào	Đầu vào	Kết quả
Số mẫu	131.00	131.00	131.00	131.00	131.00
Trung bình	68.79	34.21	-1.41	30.98	16.47
Độ lệch chuẩn	28.78	19.88	13.41	19.45	8.05
Giá trị nhỏ nhất	22.00	4.50	-94.89	0.40	5.50
Giá trị lớn nhất 25%	54.00	19.00	-4.94	17.00	10.00
Giá trị lớn nhất 50%	65.00	33.00	0.53	28.00	13.50
Giá trị lớn nhất 75%	82.00	44.45	5.88	46.00	23.35
Giá trị lớn nhất	213.00	132.00	29.07	91.00	39.00



Hình 1: Quan hệ giữa các biến số đầu vào và ra với nhau

Cũng như tương quan giữa các biến đầu vào của

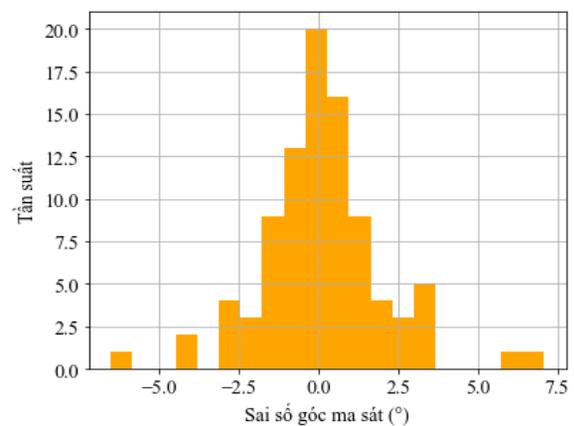


Hình 2: Góc ma sát dự đoán cho phân huấn luyện bởi mô hình rừng cây ngẫu nhiên RF

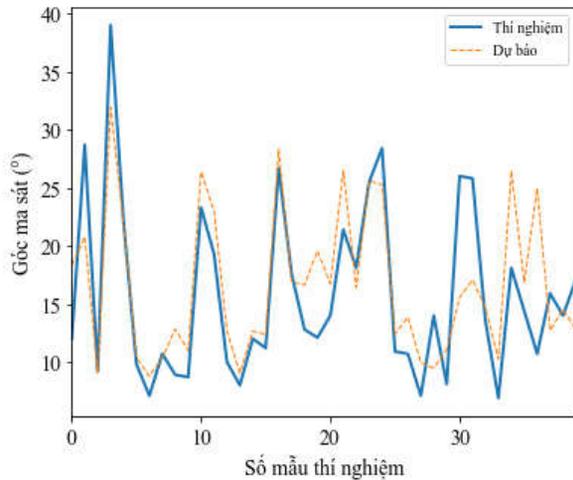
mô hình được thể hiện trong hình 1 dưới đây. Có thể thấy các biến đầu vào đều có các tương quan với nhau, không có biến nào có thể được thể được biểu diễn độc lập thông qua biến khác.

3. KẾT QUẢ VÀ THẢO LUẬN

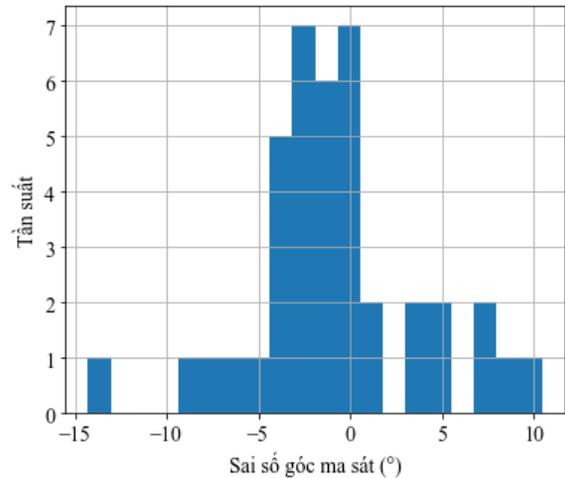
Hình 2 và hình 3 lần lượt là kết quả của góc ma sát dự đoán cho phân huấn luyện và tần suất sai số tương ứng giữa giá trị dự đoán của mô hình và giá trị thực. Hình 3 và hình 4 lần lượt là kết quả của góc ma sát dự đoán cho phân kiểm chứng và tần suất sai số tương ứng giữa giá trị dự đoán của mô hình RF cho kiểm chứng và giá trị thực.



Hình 3: Tần suất sai số giữa góc ma sát dự đoán bởi RF và giá trị thực tế cho phân huấn luyện



Hình 4: Góc ma sát dự đoán cho phần kiểm chứng bởi mô hình rừng cây ngẫu nhiên RF

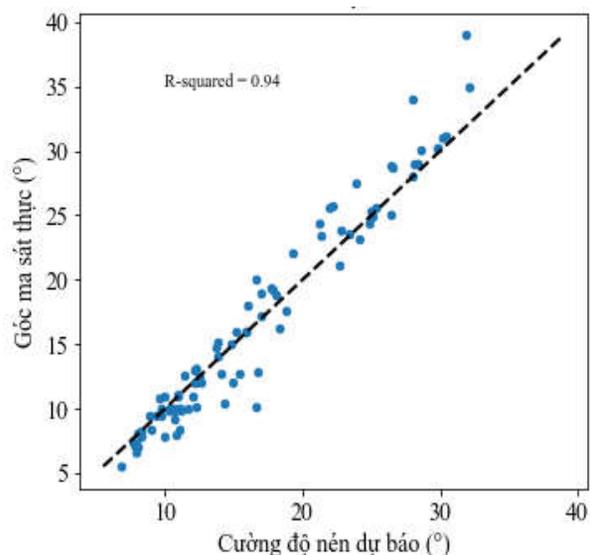


Hình 5: Tần suất sai số giữa góc ma sát dự đoán bởi RF và giá trị thực tế cho phần kiểm chứng

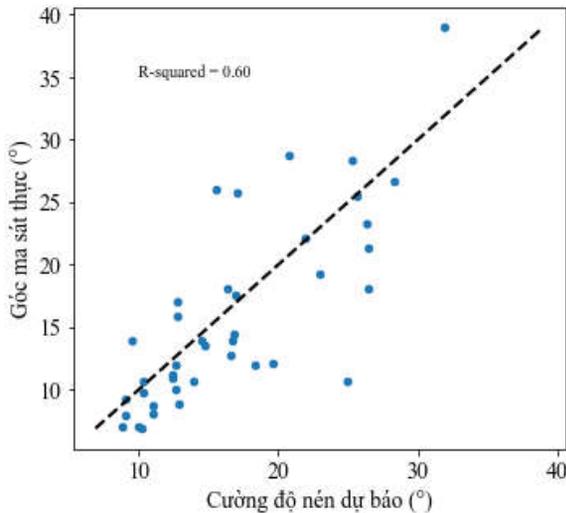
Kết quả cho thấy mô hình rừng cây ngẫu nhiên dự đoán góc ma sát cho khoảng 92 mẫu đất với độ chính xác tương đối cao cho phần huấn luyện. Sự sai lệch giá trị dự đoán là rất nhỏ $\pm 2.5^\circ$, và sai số xấp xỉ 0 MPa với khoảng hơn 20 mẫu, đây là một kết quả rất tốt cho phần huấn luyện của mô hình rừng cây ngẫu nhiên với sai số RMSE = 1.97° . Do vậy, mô hình RF sau khi đã huấn luyện thành công được kiểm chứng bởi khoảng hơn 39 mẫu góc ma sát còn lại. Kết quả trong hình 4, 5 cho thấy mô hình RF sau khi đã được huấn luyện cho kết quả tương đối tốt với khoảng 39 mẫu góc ma sát còn lại. Sai số trong mô hình kiểm chứng tập chung chủ yếu trong khoảng $\pm 10^\circ$ với sai số RMSE = 4.71°

Mô hình hồi quy cho 2 phần huấn luyện và kiểm chứng được thể hiện ở hình 6 và hình 7. Từ hình trên ta thấy khả năng dự báo của mô hình là tương đối cao khá sát với góc ma sát thực tế, tuy nhiên vẫn có sự sai số với các góc ma sát lớn. Giá trị tương quan R thu được cho phần huấn luyện là $R=0.97$ còn phần kiểm chứng là $R=0.78$. Điều này cho thấy việc áp dụng mô hình rừng cây ngẫu nhiên RF cho việc dự báo góc ma sát của đất là rất khả thi. Tuy nhiên, hệ số tương quan R dành cho phần kiểm chứng chưa thực sự cao như phần huấn luyện.

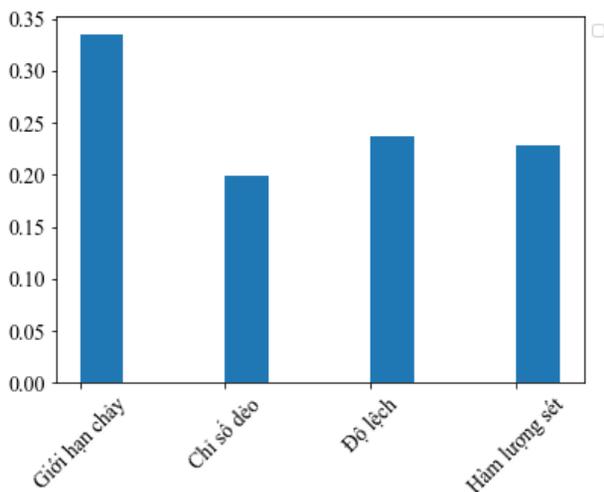
Đặc biệt bằng vào mô hình rừng cây ngẫu nhiên RF, vai trò quan trọng của từng thành phần đến góc ma sát của đất cũng được biểu diễn trong hình 8. Kết quả biểu diễn trong hình 8 cho thấy rằng yếu tố ảnh hưởng lớn nhất đến góc ma sát là giới hạn chảy, sau đó là độ lệch so với đường loại A trong biểu đồ phân loại Casagrande, theo sau là hàm lượng sét và chỉ số dẻo là yếu tố ảnh hưởng nhỏ nhất đến góc ma sát của đất



Hình 6: Kết quả hồi quy mô hình rừng cây ngẫu nhiên cho phần huấn luyện



Hình 7: Kết quả hồi quy mô hình rừng cây ngẫu nhiên RF cho phần kiểm chứng



Hình 8: Các yếu tố ảnh hưởng tới góc ma sát của đất phân tích bằng mô hình rừng cây ngẫu nhiên

4. KẾT LUẬN VÀ KIẾN NGHỊ

Trong nghiên cứu này, khả năng của kỹ thuật trí tuệ nhân tạo (AI) trong dự đoán góc ma sát của đất đã được kiểm tra. Số lượng dữ liệu dùng cho mô phỏng được tập hợp từ kết quả thí nghiệm đã được công bố trong các tạp chí uy tín trên thế giới. Các mẫu đất thí nghiệm là tập hợp của nhiều loại đất khác nhau từ nhiều vùng khác nhau nên có sự biến thiên lớn. Do vậy việc ứng dụng được mô hình rừng ngẫu nhiên một thuật toán của trí thông minh nhân tạo AI vào dự đoán và nghiên cứu được góc ma sát của đất là hết sức có ý nghĩa. Góp phần có thể dự đoán được các nguy cơ xảy ra tai biến thiên nhiên của đất khi biết được các thông số đầu vào như giới hạn chảy, chỉ số dẻo, độ lệch biểu đồ phân loại Casagrande ΔPI và hàm lượng sét của đất. Kết quả chỉ ra mô hình rừng cây ngẫu nhiên khả thi trong việc xác định sức chống cắt còn lại của đất biến dạng của đất với hệ số tương quan cho mô hình huấn luyện là 0.97 và kiểm chứng là 0.78. Sai số của mô hình huấn luyện và áp dụng cho kiểm chứng là rất nhỏ chủ yếu tập chung vào khoảng sai số từ $\pm 2.5^\circ$. Ngoài ra, mô hình rừng cây ngẫu nhiên có thể chỉ ra tầm quan trọng của từng tính chất của đất đến góc ma sát của đất biến dạng này, lần lượt theo thứ tự là Giới hạn chảy > Độ lệch biểu đồ phân loại Casagrande ΔPI > Hàm lượng sét > Chỉ số dẻo.

TÀI LIỆU THAM KHẢO

- [1] ‘Skempton A.W., The long term stability of clay slopes. Geotechnique, 1964, 14, 77-101’.
- [2] ‘Mesri G., Cepeda-Diaz A.F., Residual strength of clays and shales. Geotechnique, 1986, 36, 269-274’.
- [3] ‘Colotta T., Cantoni R., Pavesi U., Robert E., Moretti P.C., A correlation between residual friction angle, gradation and index properties of cohesive soil. Geotechnique, 1989, 39, 343-346’.
- [4] ‘Stark T.D., Eid H.T., Drained residual strength of cohesive soils. J. Geotech. Geoenviron. Eng., 1994, 120, 856-871’.

- [5] ‘Wesley L.D., Residual strength of clays and correlations using Atterberg limit. *Geotechnique*, 2003, 53, 669-672’.
- [6] ‘Sridharan A., Rao P.R., Discussion: Residual strength of clays and correlation using Atterberg limits. *Geotechnique*, 2004, 54, 503-504’.
- [7] ‘Tiwari B., Marui H., A new method for the correlation of residual shear strength of the soil with mineralogical composition. *J. Geotech. Geoenviron. Eng.*, 2005, 131, 1139-1150’.
- [8] ‘Kaya A., Kwong J.K.P., Evaluation of common practice empirical procedures for residual friction angle of soils: Hawaiian amorphous material rich colluvial soil case study. *Eng. Geol.*, 2007, 92, 49–58’.
- [9] ‘Das S.K., Basudhar P.K., Prediction of residual friction angle of clays using artificial neural network. *Eng. Geol.*, 2008, 100, 142-145’.
- [10] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and Regression Trees*. Taylor & Francis, 1984.
- [11] L. Breiman, *Classification and Regression Trees*. Routledge, 2017.
- [12] A.-L. Boulesteix, S. Janitza, J. Kruppa, and I. R. König, ‘Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics’, *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, 2012, doi: 10.1002/widm.1072.
- [13] M. Shariati *et al.*, ‘Application of a hybrid artificial neural network-particle swarm optimization (ANN-PSO) model in behavior prediction of channel shear connectors embedded in normal and high-strength concrete’, *Applied Sciences*, vol. 9, no. 24, p. 5534, 2019.
- [14] S. K. Das, P. Samui, S. Z. Khan, and N. Sivakugan, ‘Machine learning techniques applied to prediction of residual strength of clay’, *cent.eur.j.geo.*, vol. 3, no. 4, pp. 449–461, Dec. 2011, doi: 10.2478/s13533-011-0043-1.
- [15] S. Z. Khan, S. Suman, M. Pavani, and S. K. Das, ‘Prediction of the residual strength of clay using functional networks’, *Geoscience Frontiers*, vol. 7, no. 1, pp. 67–74, Jan. 2016, doi: 10.1016/j.gsf.2014.12.008.