



Original Article

Application of Artificial Intelligence (AI) for Predicting Fracture Intervals and Fracture Density from Well Log Data

Nguyen Hoang Anh^{1,*}, Nguyen Thanh Tung¹, Nguyen Thanh Sang¹,
Nguyen Trung Son¹, Nguyen Phuong Thuy²

¹Vietnam Petroleum Institute, 167 Trung Kinh, Yen Hoa, Hanoi, Vietnam

²Thang Long Joint Operating Company, 3C Ton Duc Thang, Sai Gon, Ho Chi Minh, Vietnam

Received 28th November 2025

Revised 10th March 2026; Accepted 17th April 2026

Abstract: Fractured granitoid basement plays a critical role in hydrocarbon production in Vietnam, where fracture interval, fracture density and their characteristics are key parameters for reservoir evaluation, reserve estimation, and field development. However, data obtained from core analysis and borehole image (BHI) logs are usually scarce and expensive to acquire and process, limiting their availability for reservoir characterization. This study develops supervised machine learning (ML) models to predict fracture zones and fracture density using well log data from the SA field, Cuu Long Basin. Input data include DCALI, GR, LLD, NPHI, RHOB, DTC, DTS..., combined with hydrocarbon indicator, BHI-derived fracture labels for model training. LightGBM, an efficient gradient boosting algorithm, was selected for its speed and performance. The model achieved recall values exceeding 90% across training, validation, and testing datasets. Blind tests on well XA-2X confirmed the model's robustness, with prediction accuracy of 82% respectively, and strong agreement between predicted and BHI-derived fracture densities. Fracture porosity and permeability were also calculated from predicted fracture density and aperture (available from two BHI-analyzed wells), yielding values consistent with fractured basement reservoir behavior. The resulting ML-based prediction modules can be applied to wells lacking BHI data, supporting real-time fracture identification, reservoir characterization, and reserve estimation in fractured basement reservoirs. This approach significantly reduces dependency on high-cost imaging tools and paves the way for broader artificial intelligence into digital reservoir characterization and management workflows.

Keywords: Fracture interval, fracture density, well log, BHI, machine learning.

* Corresponding author.

E-mail address: anhnh@vpi.pvn.vn

<https://doi.org/10.25073/2588-1094/vnuces.5459>

Ứng dụng trí tuệ nhân tạo (AI) trong việc dự báo khoảng nứt nẻ và mật độ nứt nẻ từ tài liệu địa vật lý giếng khoan

Nguyễn Hoàng Anh^{1,*}, Nguyễn Thanh Tùng¹, Nguyễn Thanh Sang¹,
Nguyễn Trung Sơn¹, Nguyễn Phương Thủy²

¹*Viện Dầu khí Việt Nam, 167 Trung Kính, Phường Yên Hoà, Hà Nội, Việt Nam*

²*Công ty điều hành chung Thăng Long, 3C Tôn Đức Thắng, Sài Gòn, Hồ Chí Minh, Việt Nam*

Nhận ngày 28 tháng 11 năm 2025

Chỉnh sửa ngày 10 tháng 3 năm 2026; Chấp nhận đăng ngày 17 tháng 4 năm 2026

Tóm tắt: Đá móng granitoid nứt nẻ là đối tượng chứa dầu khí quan trọng tại Việt Nam, trong đó khoảng nứt nẻ, mật độ và đặc tính nứt nẻ đóng vai trò then chốt trong đánh giá đặc tính vỉa chứa, tính toán tài nguyên và phát triển mỏ. Tuy nhiên, dữ liệu phân tích mẫu lõi và ảnh thành hệ (BHI) cho đối tượng này thường thiếu hụt và chi phí tốn kém. Nghiên cứu này phát triển các mô hình học máy có giám sát để dự báo khoảng nứt nẻ và mật độ nứt nẻ dựa trên tài liệu địa vật lý giếng khoan (ĐVLGK) tại mỏ SA, bể Cửu Long. Bộ dữ liệu đầu vào bao gồm các đường cong DCALI, GR, LLD, NPHI, RHOB, DTC, DTS,... kết hợp với dữ liệu phân tích BHI và các biểu hiện dầu khí để xây dựng nhân huấn luyện. Thuật toán LightGBM được lựa chọn nhờ khả năng xử lý dữ liệu hiệu quả và cho kết quả chính xác cao. Mô hình đạt recall >90% ở cả tập huấn luyện, kiểm tra và kiểm thử. Kiểm chứng trên giếng XA-2X cho thấy kết quả đạt độ chính xác 82%, đồng thời mật độ nứt nẻ dự báo tương đồng với kết quả phân tích BHI. Dựa trên mật độ và độ mở khe nứt, độ rỗng và độ thấm nứt nẻ được tính toán và cho kết quả phù hợp với các đặc tính đá móng thực tế. Mô hình dự báo tối ưu có thể triển khai mở rộng cho các giếng mới chưa có tài liệu BHI, hỗ trợ hiệu quả cho công tác nghiên cứu đặc tính tầng chứa, đánh giá tiềm năng trữ lượng và mô hình hóa vỉa chứa móng nứt nẻ.

Từ khóa: Khoảng nứt nẻ, mật độ nứt nẻ, ĐVLGK, BHI, học máy.

1. Mở đầu

Đá chứa móng granitoid nứt nẻ là đối tượng chứa dầu khí đặc biệt và có ý nghĩa quan trọng trong khai thác dầu khí tại các mỏ thuộc thềm lục địa Việt Nam, đặc biệt tại bể Cửu Long. Việc nhận diện khoảng nứt nẻ, mật độ nứt nẻ và đặc tính nứt nẻ (độ mở, độ rỗng, độ thấm) là bước then chốt trong nghiên cứu tính chất đá chứa móng nhằm phục vụ tính toán dầu khí tại chỗ ban đầu và trữ lượng dầu khí, xây dựng mô hình địa

chất, mô hình khai thác mỏ. Tuy nhiên, do hạn chế về số lượng mẫu lõi cho đối tượng đá móng, việc xác định các tham số nứt nẻ hiện nay chủ yếu vẫn dựa vào tài liệu ảnh thành hệ (Borehole Imaging – BHI) thường chỉ được thực hiện tại một số giếng khoan do chi phí cao, dẫn đến hạn chế trong nghiên cứu và đánh giá tính phân dị của hệ thống nứt nẻ ở quy mô mỏ.

Hệ thống nứt nẻ được hình thành bởi trường ứng suất kiến tạo, chủ yếu là ứng suất nén ép và co ngót magma và bị hòa tan, rửa lữa bởi các hoạt

* Tác giả liên hệ.

Địa chỉ email: anhnh@vpi.pvn.vn

<https://doi.org/10.25073/2588-1094/vnuees.5459>

động thủy nhiệt xảy ra sau hàng chục triệu năm làm mở rộng, liên thông và tăng khả năng thẩm chứa của đá móng. Các đường đo ĐVLGK cung cấp chỉ báo quan trọng để nhận diện và dự báo nứt nẻ trong đá móng [1]. Sóng dọc (DTC) và sóng ngang (DTS) phản ánh thời gian truyền sóng qua thành hệ, tăng lên ở các đới nứt nẻ – đứt gãy do mật độ khung đá giảm. Đường cong neutron (NPHI) đo chỉ số hydro, có giá trị cao hơn ở vỉa chứa nứt nẻ, hang hốc chứa chất lưu (dầu khí, nước, dung dịch khoan) và thường dùng để dự báo sự tồn tại nứt nẻ. Đường cong mật độ (RHOB) thể hiện mật độ khối thành hệ, giảm đáng kể tại đới nứt nẻ – đập vỡ so với khoảng chặt xít. Đường cong điện trở suất nông và sâu (LLS, LLD) giảm khi có nứt nẻ, đứt gãy do chất lưu xâm nhập, thể hiện rõ ở đá móng có điện trở thực cao. Ngoài ra, đới nứt nẻ đập vỡ chịu ảnh hưởng thủy nhiệt và khoáng vật thứ sinh, hạt mịn và uranium cao, đặc trưng bởi giá trị Gamma Ray (GR) tăng và đường kính giếng DCALI lớn, dễ sập lở. Ngoài ra các hiện tượng như mất dung dịch khoan cũng thể hiện các khoảng nứt nẻ, đứt gãy hoặc các biểu hiện dầu khí cũng minh chứng sự có mặt của dầu khí trong đới nứt nẻ, đập vỡ [2]. Do vậy, việc kết hợp tổ hợp các tài liệu ĐVLGK với chi phí thấp hơn thường được sử dụng để dự báo khoảng nứt nẻ và mật độ nứt nẻ nhằm đánh giá tính chất vỉa chứa [3, 4].

Với sự phát triển của các giải thuật học máy ngày nay cho phép liên kết các số liệu một cách rất hiệu quả giữa các số liệu không liên kết vật lý với nhau cho phép dự báo các tham số một cách hiệu quả và đạt được những thành tựu đáng kể. Ứng dụng các mô hình học máy khác nhau để dự báo dự báo mật độ nứt nẻ của đá chứa trên cơ sở tài liệu ĐVLGK đã được thực hiện như các nghiên cứu của [1, 5-9] sử dụng mô hình hồi quy tuyến tính trong việc dự báo mật độ nứt nẻ từ các đường cong ĐVLGK và kết quả minh giải ảnh thành hệ (BHI). Dong và nnk, 2020 sử dụng thuật toán học máy nửa giám sát (semi-supervised machine learning) cho các đường cong đo ĐVLGK cơ bản để xác định mật độ nứt nẻ cho đối tượng đá chứa cát kết chặt xít, thuộc bể Ordos, Trung Quốc. Li và nnk, 2018 dự báo

mật độ nứt nẻ đã sử dụng thuật toán máy vector trợ giúp (support vector machine) với đầu vào là các đường cong đo siêu âm [7]. Tokhmechi và nnk, 2009a&b, 2010 sử dụng biến đổi wavelet (wavelet transform) trong việc xác định mật độ nứt nẻ của đá chứa [1, 8, 9].

Bên cạnh các nghiên cứu quốc tế, tại Việt Nam cũng đã có một số công trình bước đầu ứng dụng AI/ML trong lĩnh vực tìm kiếm–thăm dò và đánh giá đá chứa. Tiêu biểu, Nguyễn Thu Huyền và nnk, 2019 đã áp dụng mạng neuron nhân tạo (ANN) trong điều kiện số lượng giếng khoan hạn chế, kết hợp diễn giải địa chấn, phân tích ĐVLGK và dữ liệu mẫu để dự báo phân bố và chất lượng đá chứa carbonate Miocene bể Phú Khánh, cho thấy tính khả thi và hiệu quả của AI trong bài toán dự báo đặc tính vỉa chứa khi dữ liệu hạn chế [10]. Ở hướng mô tả tương/môi trường trầm tích phục vụ mô hình địa chất và đặc trưng hóa vỉa chứa, một nghiên cứu trình bày tại OTC Asia đã sử dụng AI để dự báo tương từ dữ liệu giếng khoan tại một mỏ ở bể Nam Côn Sơn [11]. Trong lĩnh vực khoan và địa cơ học phục vụ thiết kế giếng, ANN cũng đã được áp dụng để dự báo áp suất nứt vỉa, góp phần hỗ trợ ra quyết định kỹ thuật khi điều kiện địa chất phức tạp [12]. Gần đây, các nghiên cứu trong nước tiếp tục mở rộng sang các ứng dụng ML trên dữ liệu giếng khoan nhằm bổ sung/dự báo một số đường cong ĐVLGK còn thiếu phục vụ minh giải thông số [13]. Tuy nhiên, đa số các nghiên cứu trước đây chủ yếu tập trung vào đá trầm tích/carbonate, trong khi các nghiên cứu AI/ML cho đối tượng đá móng granitoid nứt nẻ vẫn còn hạn chế.

Ngoài ra, một số kênh thông tin khoa học–công nghệ trong nước giai đoạn 2022–2025 cũng ghi nhận xu hướng gia tăng ứng dụng AI trong ngành dầu khí, phản ánh nhu cầu chuyển đổi số và tối ưu hóa quy trình kỹ thuật [14, 15]. Nhìn chung, các công bố trong nước cho thấy tiềm năng của AI/ML trong nhiều bài toán dầu khí, đồng thời đặt ra nhu cầu phát triển các chu trình nghiên cứu chuyên biệt hơn cho từng đối tượng địa chất–vỉa chứa và đặc điểm dữ liệu tại Việt Nam [16].

Trong nghiên cứu này, nhóm tác giả tập trung xây dựng mô hình trí tuệ nhân tạo (AI) và

học máy (ML) để dự báo khoảng nứt nẻ, mật độ nứt nẻ từ đó làm cơ sở xác định các đặc tính rỗng – thấm của đá móng nứt nẻ từ tài liệu ĐVLGK kết hợp kết quả phân tích BHI mỏ SA, bể Cửu Long. Kết quả của nghiên cứu được kỳ vọng cho phép dự báo phân bố và đặc trưng nứt nẻ một cách định lượng, làm rõ đặc điểm đá chứa móng nứt nẻ ở các giếng khoan không có tài liệu BHI, qua đó góp phần giảm chi phí đầu tư và nâng cao hiệu quả công tác tìm kiếm – thăm dò và khai thác dầu khí.

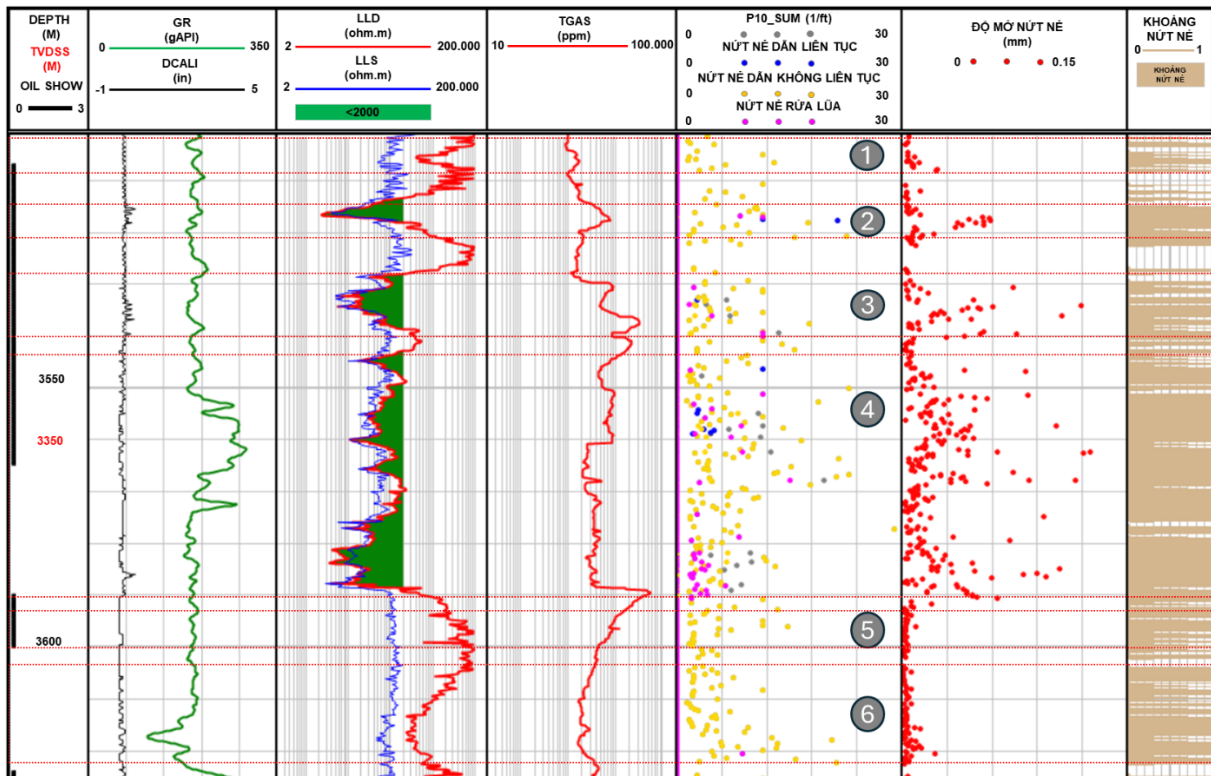
2. Dữ liệu và phương pháp nghiên cứu

2.1. Dữ liệu nghiên cứu

Nghiên cứu sử dụng bộ dữ liệu tổng hợp từ 12 giếng khoan tại một khu vực mỏ, bể Cửu Long được trình bày trong Bảng 1. Tài liệu đầu vào bao gồm các đường cong ĐVLGK tiêu chuẩn như sóng dọc (DTC), sóng ngang (DTS),

mật độ (RHOB), neutron (NPHI), điện trở suất sâu và nông (LLD, LLS), gamma ray (GR), caliper (DCALI), ảnh thành hệ (BHI) và cùng với thành phần khí (Total Gas) và các biểu hiện dầu khí (Oil Show) kèm theo. Như đã trình bày ở phần mở đầu, các tài liệu này đều có mối liên hệ chặt chẽ với sự phát triển nứt nẻ và đới dập vỡ của đá móng, vì vậy được lựa chọn làm dữ liệu đầu vào để xây dựng mô hình dự báo [1-4].

Dữ liệu ảnh thành hệ (BHI) bao gồm phân tích định lượng mật độ nứt nẻ, loại nứt nẻ (dẫn liên tục, dẫn không liên tục, nứt nẻ rửa lỗ) và độ mở khe nứt tại các giếng có dữ liệu (giếng khoan 1 và 2) được sử dụng làm cơ sở dán nhãn khoảng nứt nẻ huấn luyện mô hình. Các khoảng nứt nẻ được xác định thông qua kết quả phân tích mật độ nứt nẻ bao gồm tổng của 3 loại nứt nẻ trên, áp dụng kỹ thuật cửa sổ trượt 1m và bước nhảy 0.1524 m kết hợp tài liệu ĐVLGK và biểu hiện dầu khí tương ứng để gán nhãn cho các khoảng nứt nẻ (Hình 1).



Hình 1. Dán nhãn khoảng nứt nẻ (track 7) dựa trên tài liệu BHI và ĐVLGK giếng khoan 1.

Bảng 1. Tổng hợp tài liệu giếng khoan khu vực nghiên cứu

Giếng khoan	Nóc móng (m)		Tài liệu ĐVLGK								
	MD	TVDSS	GR	CALI	LLD /LLS	RHOB	NPHI	DTC	DTS	BHI	Ghichú
1	3366	3225.5	+	+ 8.5	+	-	-	-	-	+	Schlumberger
2	3249.7	3216	+	+ 8.5,6	+	+	+	+	+	+	Baker Hughes
3	3558	3234.9	+	+ 8.5	+	+	+	+	+	+	Schlumberger
4	3385	3355.9	+	+ 8.5	+	+	+	+	-	+	Schlumberger
5	3611	3565.2	+	+ 6	+	+	+	+	+	+	Schlumberger
6	3927	3272.6	+	+ 6	+	+	+	+	+	+	Schlumberger
7	3459	3228.8	+	+ 8.5	+	+	+	+	+	+	Schlumberger
8	3367	3239.1	+	+ 8.5	+	+	+	+	+	+	Schlumberger
9	3615	3211.3	+	-	+	-	-	-	-	+	Schlumberger
10	3614	3365	+	-	+	-	-	-	-	+	Schlumberger
11	4256	3505.9	+	+ 8.5	+	+	+	+	+	+	Schlumberger
12	3484	3297.7	+	-	+	-	-	-	-	+	Schlumberger

+ Có - Không có

2.2. Phương pháp nghiên cứu

Trí tuệ nhân tạo (AI) là lĩnh vực tích hợp sức mạnh tính toán của máy tính với trí thông minh của con người nhằm mang lại những giải pháp thông minh và đáng tin cậy đối với các bài toán phi tuyến và phức tạp. AI là một lĩnh vực khoa học nghiên cứu phương pháp cho phép máy tính suy nghĩ và đưa ra các quyết định. Học máy (ML) là một bộ phận của AI, cung cấp công cụ thống kê để khai phá và phân tích lượng dữ liệu. Học máy được chia thành các nhánh chính như học có giám sát, học không giám sát, học giám sát một phần, học sâu, học tăng cường và học tự giám sát, cụ thể:

- Học có giám sát: là kỹ thuật học máy được sử dụng với dữ liệu đã được dán nhãn, người dùng sử dụng dữ liệu đã được phân loại để huấn luyện mô hình;

- Học không giám sát: là kỹ thuật học máy sử dụng với các dữ liệu chưa được dán nhãn, mô hình sẽ sử dụng dữ liệu đầu vào để phân nhóm, phân loại dữ liệu;

- Học giám sát một phần: là kỹ thuật học máy, hỗn hợp của giám sát và không giám sát, sử dụng một phần dữ liệu có dán nhãn và một phần dữ liệu không dán nhãn để huấn luyện mô hình;

- Học sâu: là một nhánh của học máy, sử dụng mô hình mạng nơron nhân tạo với nhiều lớp ẩn giúp tìm ra các qui luật phức tạp, ẩn sâu trong dữ liệu;

- Học tăng cường: là một nhánh của học máy, phương pháp này sử dụng kỹ thuật thử và sai số để huấn luyện AI bằng các hàm thưởng phạt khi thực hiện thành công hay thất bại một nhiệm vụ trong một môi trường cụ thể;

- Học tự giám sát: là kỹ thuật học máy mô hình sẽ tự huấn luyện để nhận diện một phần dữ liệu từ một phần khác của dữ liệu.

Trong nghiên cứu này, dữ liệu đã được phân loại và dán nhãn dựa trên tài liệu BHI, ĐVLGK và các biểu hiện dầu khí vì vậy, các kỹ thuật học có giám sát được sử dụng để xây dựng mô hình dự báo cho các giếng khoan mới. Các thuật toán học có giám sát phổ biến gồm hồi quy tuyến tính, SVM, KNN, cây quyết định, mạng nơron nhiều lớp, và đặc biệt là các phương pháp cây tích hợp (Random Forest, Gradient Boosting,...). Các mô hình cây tích hợp được đánh giá cao nhờ khả năng khớp dữ liệu phi tuyến, ít bị nhiễu và cho độ chính xác cao trên dữ liệu dạng bảng.

Nhiều nghiên cứu đã chứng minh điều này: Syed và nnk, 2022 đã so sánh SVM, KNN và Random Forest (RF) khi dự báo đường cong

ĐVLGK còn thiếu như DTS, RHOB và nhận thấy rằng RF chính xác nhất [17]; Grinsztajn và nnk, 2022 đã làm một so sánh giữa mô hình học máy cây tích hợp và mô hình học sâu thực hiện trên 45 tập dữ liệu dạng bảng khác nhau và nhận thấy rằng các mô hình học máy cây tích hợp cho kết quả tốt hơn hẳn so với các mô hình học sâu nhờ khả năng khớp dữ liệu phi tuyến và không có qui luật rõ ràng cũng như ít bị ảnh hưởng bởi các dữ liệu ít có quan hệ với đối tượng cần dự đoán [18].

Các mô hình học máy cây tích hợp đã thể hiện được tính ưu việt hơn hẳn các phương pháp khác đối với dữ liệu dạng bảng vì vậy, trong nghiên cứu này, nhóm tác giả sẽ kế thừa điều này nhằm giảm thời gian xây dựng và hiệu chỉnh mô hình. Một số thuật toán chính sử dụng trong nghiên cứu bao gồm:

- XGBoost: thuật toán gradient boosting cực đại, tăng tốc và mở rộng tốt, xử lý dữ liệu lớn.

- LightGBM: boosting tối ưu hóa chia tập dữ liệu, tính toán nhanh, hiệu quả.

- CatBoost: gradient boosting cho dữ liệu phân loại, xử lý giá trị khuyết, không cần one-hot, dễ dùng và chính xác cao.

Nghiên cứu này huấn luyện mô hình học máy phân loại nhị phân (binary classification) với hai lớp: nút nẻ (1) và không nút nẻ (0). Để bảo đảm tính cân bằng và đánh giá chính xác hiệu quả của mô hình, các chỉ số đánh giá mô hình bao gồm độ chính xác (Accuracy), độ thu hồi (Recall), độ chính xác dự báo (Precision), được tính toán trên cả tập dữ liệu kiểm chứng và thử nghiệm blind test từ các giếng chưa từng được đưa vào huấn luyện. Ma trận nhầm lẫn (Bảng 2) là một công cụ để đo lường kết quả của một mô hình phân loại, trong đó mỗi hàng và cột tương ứng với một lớp dự báo và thực tế, và giá trị trong ô tương ứng là số lượng mẫu được phân loại đúng hoặc sai. Cụ thể, ma trận nhầm lẫn có 4 giá trị chính:

- True Positives (TP): số mẫu dự báo đúng thuộc lớp positive.

- False Positives (FP): số mẫu dự báo sai thuộc lớp positive.

- True Negatives (TN): số mẫu dự báo đúng thuộc lớp negative.

- False Negatives (FN): số mẫu dự báo sai thuộc lớp negative.

Recall và Precision được tính toán từ ma trận nhầm lẫn (confusion matrix).

- Recall là tỉ lệ số mẫu thực tế positive được dự báo đúng trên tổng số mẫu thực tế positive; $Recall = TP / (TP + FN)$.

- Precision là tỉ lệ số mẫu dự báo đúng trong lớp positive trên tổng số mẫu dự báo thuộc lớp positive; $Precision = TP / (TP + FP)$.

Bảng 2. Ma trận so sánh kết quả giữa dữ liệu đầu vào và dự báo

		Giá trị dự báo	
		1	0
Giá trị thực	1	TP	FN
	0	FP	TN

2.2.1. Phân tích dữ liệu đầu vào

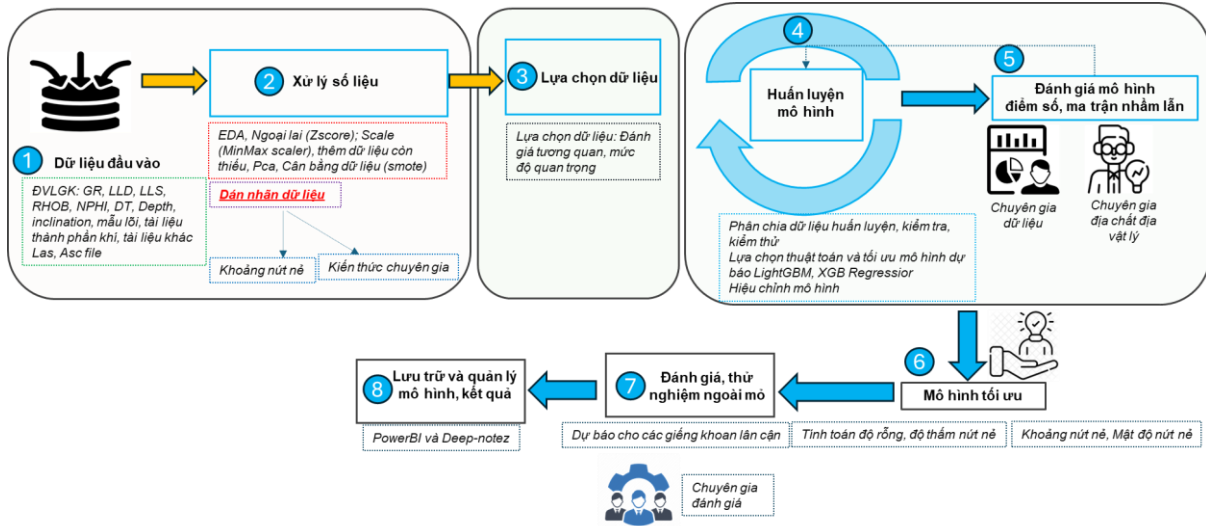
Quy trình xây dựng mô hình dự báo được thực hiện theo Hình 2. Dữ liệu đầu vào của 12 giếng được gom nhóm thành một tệp .csv chung. Phân tích và mô tả dữ liệu được thực hiện bằng công cụ DataPrep trên bộ dữ liệu 14 biến với 108.908 dòng (1.524.712 ô), trong đó 203.762 ô thiếu dữ liệu (13,4%) và không có dòng trùng lặp. Đường DTS có tỷ lệ thiếu cao nhất (45.050 ô), tiếp theo là RHOB, NPFI, DTC, DCALI_FINAL (từ 33.043 đến 36.685 ô), tập trung tại các giếng 1, 9, 10, 12; các đường còn lại như INCL, AZIM_TN, TVD thiếu ít hơn từ 4.574 đến 6.762 ô. Một số giếng xiên dài như giếng khoan 3 có gần 20.000 điểm; dữ liệu tập trung nhiều ở giếng khai thác (P) hơn giếng thăm dò (X).

Công tác chọn, hiệu chỉnh dữ liệu, phục hồi dữ liệu còn thiếu và loại bỏ dị thường được tiến hành, tuy nhiên các khoảng dữ liệu phản ánh đặc điểm địa chất vẫn được giữ nguyên trên cơ sở trao đổi với nhóm chuyên gia để đảm bảo độ chính xác và tính đặc trưng. Tỷ lệ nhãn nút nẻ/không nút nẻ lần lượt là 36,24% và 63,76%, tương đối cân bằng, giúp mô hình học máy không bị lệch nhãn; nếu lệch cần lấy mẫu tăng cường (oversampling), lấy mẫu giảm bớt (undersampling) hoặc điều chỉnh tập dữ liệu. Dữ liệu có nhãn sau khi tích hợp được chia thành ba tập: huấn luyện (70%), kiểm tra (20%) và kiểm thử (10%). Tập huấn luyện dùng để xây dựng mô

hình, tập kiểm thử để đánh giá chính xác sau huấn luyện, tập kiểm tra để đánh giá tính ổn định khi dự đoán dữ liệu mới.

Ngoài ra các bước khai phá dữ liệu cũng được các tác giả tiến hành trong trình thực hiện công việc để đảm bảo dữ liệu được quan sát đầy

đủ dưới góc độ của khoa học dữ liệu. Giúp người thực hiện có cái nhìn khách quan nhất dữ liệu đang có, từ đó có những đánh giá chính xác nhất dưới góc độ dữ liệu, đồng thời dưới góc độ chuyên môn để hiểu rõ được những giới hạn của mô hình học máy khi sử dụng những tài liệu hiện có.



Hình 2. Quy trình thực hiện xây dựng mô hình dự báo.

2.2.2. Lựa chọn dữ liệu đầu vào cho mô hình

Nhằm đảm bảo mô hình học máy mang tính đại diện cho toàn bộ dữ liệu được huấn luyện cũng như có khả năng dự báo cho các đối tượng tương tự trong tương lai, các yếu tố đầu vào sẽ được sàng lọc sơ bộ trước khi đưa vào huấn luyện và tinh chỉnh. Sử dụng mô hình học máy sơ khai (các thông số mô hình là mặc định) kết hợp với SHAP để đánh giá ảnh hưởng của từng yếu tố lên kết quả đầu ra ở từng giếng theo các bước sau:

- Khớp một model với các thông số mặc định cho từng giếng;
- Tính giá trị SHAP cho các yếu tố đầu vào của từng giếng và xếp hạng;
- Lấy trung bình cho toàn bộ tập dữ liệu huấn luyện và xếp hạng theo thứ tự từ cao đến thấp (1 là cao nhất).

Bằng cách này, các yếu tố quan trọng của các giếng được đánh giá ở mức tương đương nhau mà ko bị chi phối bởi độ lớn của dữ liệu của từng giếng qua đó tăng mức độ đại diện của mô hình.

SHAP (14-Shap-github) (SHapley Additive exPlanations) [19] là một thuật toán mới để giải thích các dự đoán của mô hình Machine Learning. Nó tạo ra một giải thích dựa trên nguyên tắc của lý thuyết Shapley để phân tích sự đóng góp của từng đặc trưng cho dự đoán của mô hình. SHAP cung cấp một cách dễ hiểu để giải thích dự đoán của mô hình và có thể áp dụng cho tất cả các loại mô hình học máy, bao gồm cả mô hình dựa trên điểm số (scored models) và mô hình phân loại (classification models). Giá trị SHAP có thể được sử dụng để xác định đặc trưng nào là quan trọng nhất cho dự đoán và có thể được sử dụng để kiểm tra tính đối xứng và tính đúng đắn của mô hình.

Do các giếng 2, 6, 7, và 9 thiếu hầu hết các đường cong cơ bản, giá trị phải ngoại suy từ giếng khác nên không đủ cơ sở đánh giá độc lập. Kết quả SHAP cho thấy các đường AZIM, INCL, LLD, RHOB, NPFI, GR, DTC, LLS, DTS và DCALI có mức độ quan trọng thấp đối với mô hình (DTS còn thiếu ở nhiều giếng) nên cũng bị loại khỏi tập dữ liệu huấn luyện (Hình 3).

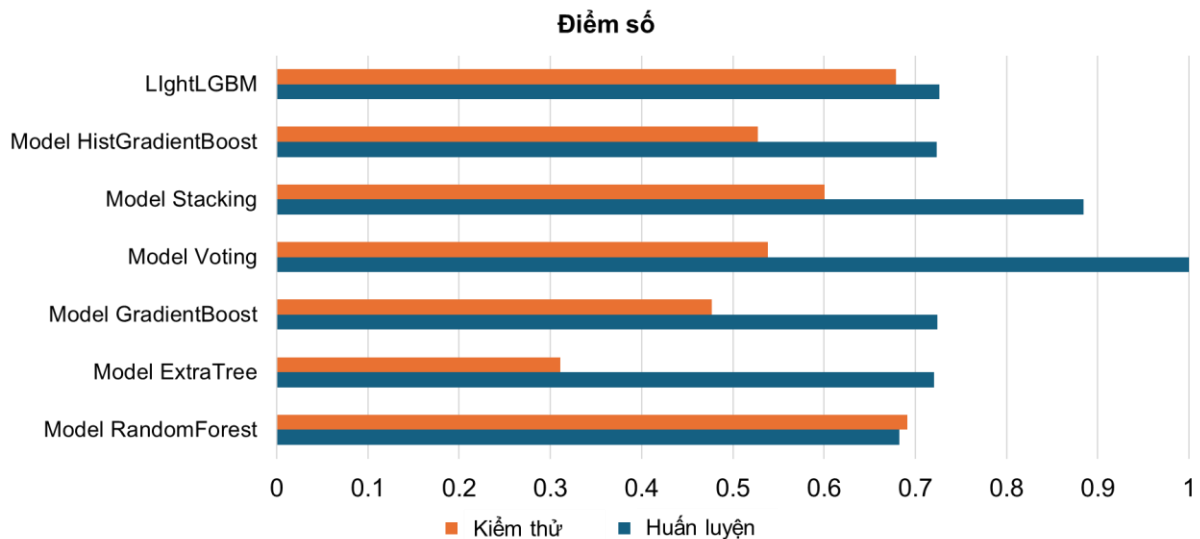
Giếng khoan/ Dữ liệu đầu vào	DCALI	GR	LLD	LLS	NPHI	RHOB	DTC	DTS	INCL	AZIM_TN	Mức độ quan trọng
1	6	7	5	10	4	3	8	9	2	1	1
2	Không đủ dữ liệu để đánh giá										2
3	8	7	5	10	4	3	9	6	2	1	3
4	7	4	3	10	9	5	6	8	2	1	4
5	9	5	2	10	4	8	6	7	3	1	5
6	Không đủ dữ liệu để đánh giá										6
7	Không đủ dữ liệu để đánh giá										7
8	10	9	5	6	4	3	8	7	2	1	8
9	Không đủ dữ liệu để đánh giá										9
10	9	4	2	10	8	7	5	6	3	1	10
11	10	6	2	4	9	5	8	7	3	1	
12	10	9	2	4	5	6	7	8	3	1	
Trung bình	8.6	6.4	3.3	8.0	5.9	5.0	7.1	7.3	2.5	1.0	

Hình 3. Tổng hợp kết quả phân tích SHAP ở các giếng khoan.

2.2.3. Huấn luyện mô hình và kiểm tra

Mô hình huấn luyện được xây dựng bằng ngôn ngữ lập trình Python, triển khai trên nền tảng điện toán đám mây Deepnote, và sử dụng các thư viện mã nguồn mở phổ biến trong lĩnh vực học máy như scikit-learn, LightGBM, XGBoost, Pandas, NumPy và Matplotlib để xử lý, huấn luyện và trực quan hóa dữ liệu. Toàn bộ tham số huấn luyện, siêu tham số

(hyperparameters), kết quả đánh giá mô hình, cùng biểu đồ huấn luyện – kiểm thử được quản lý và theo dõi tự động trên nền tảng Neptune.ai, giúp nhóm nghiên cứu dễ dàng giám sát, kiểm chứng và tái lập kết quả. Hệ thống này không lưu trữ dữ liệu gốc mà chỉ lưu thông tin huấn luyện, bảo đảm tính bảo mật và tuân thủ quy trình quản lý dữ liệu nghiên cứu. Người dùng có thể thay đổi API_Token để lưu trữ kết quả cá nhân.



Hình 4. Thử nghiệm với các thuật toán khác nhau.

Nghiên cứu đã thử nghiệm thêm nhiều thuật toán (Random Forest, Extra Trees, XGBoost...) nhưng duy nhất chỉ có thuật toán LightGBM là

đáp ứng yêu cầu đặt ra là Recall của dữ liệu huấn luyện và kiểm thử => 0,7 và không bị quá khớp (overfitting), đồng thời có sự chênh lệch giữa chỉ

số huấn luyện và kiểm thử nhỏ nhất (Hình 4). Dựa trên kết quả đã nhận được, thuật toán LightGBM cho thấy ưu thế về xử lý dữ liệu đầu vào, tốc độ huấn luyện và sự phù hợp nhất về dữ liệu hiện có. Vì vậy, nhóm tác giả đã lựa chọn sử dụng thuật toán này để phục vụ việc xây dựng mô hình học máy dự báo các khoảng nứt nẻ từ tài liệu ĐVLGK. Thuật toán LightGBM giúp giảm chi phí tìm cây quyết định tối ưu, tối ưu hóa việc chia nhánh, giảm số lượng cây con. Mô hình được huấn luyện với 200 vòng lặp, tối ưu tham số để đạt hiệu quả tốt nhất. Các thông số sử dụng để huấn luyện mô hình và thông số mô hình sau khi tinh chỉnh được trình bày ở Bảng 3, 4.

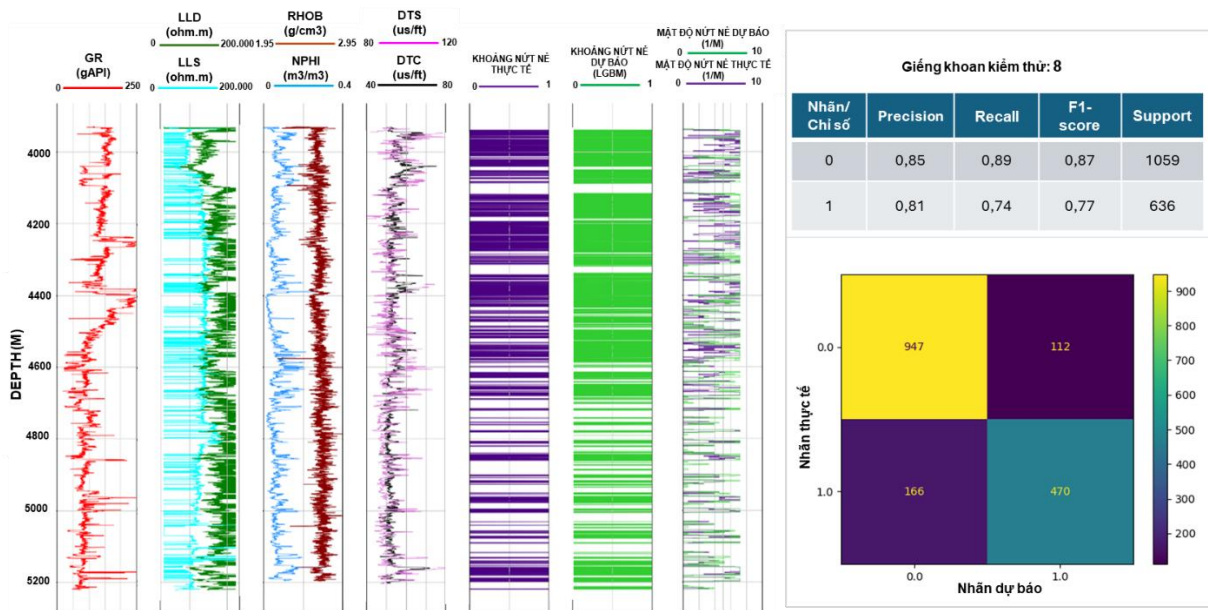
Việc đánh giá mô hình bằng ma trận nhầm lẫn (Confusion Matrix) với các chỉ số chính: Recall, Precision và F1-score để kiểm tra chất lượng mô hình. Lựa chọn hệ số recall để ưu tiên việc xác định chính xác giá trị nứt nẻ (positive), và thông thường được áp dụng trong các bài toán phân loại với số lượng giá trị các khoảng đá nứt nẻ (positive chiếm 36,24%) ít hơn số lượng giá trị các khoảng đá chặt xít (63,76%), dùng chỉ số “recall” sẽ đảm bảo tính chính xác cao trong việc dự đoán các khoảng đá có nứt nẻ.

Bảng 3. Các thông số (hyperparameters) sử dụng để huấn luyện mô hình

Thông số	Giới hạn dưới	Giới hạn trên
is_unbalance	True	
boosting_type	gbdt	
learning_rate	1e-3	0,5
max_bin	32	256
max_depth	3	20
num_interations	100	2000
num_leaves	32	256
objective	cross_entropy	
random_state	42	

Bảng 4. Thông số mô hình sau khi tinh chỉnh

Thông số	Giá trị
is_unbalance	True
boosting_type	gbdt
learning_rate	0,5
max_bin	256
max_depth	10
num_interations	1000
num_leaves	256
objective	cross_entropy
random_state	42



Hình 5. Kết quả dự báo của mô hình cho tập kiểm thử (testing) giếng khoan 8.

Bảng 5. Thông số mô hình sau khi tinh chỉnh

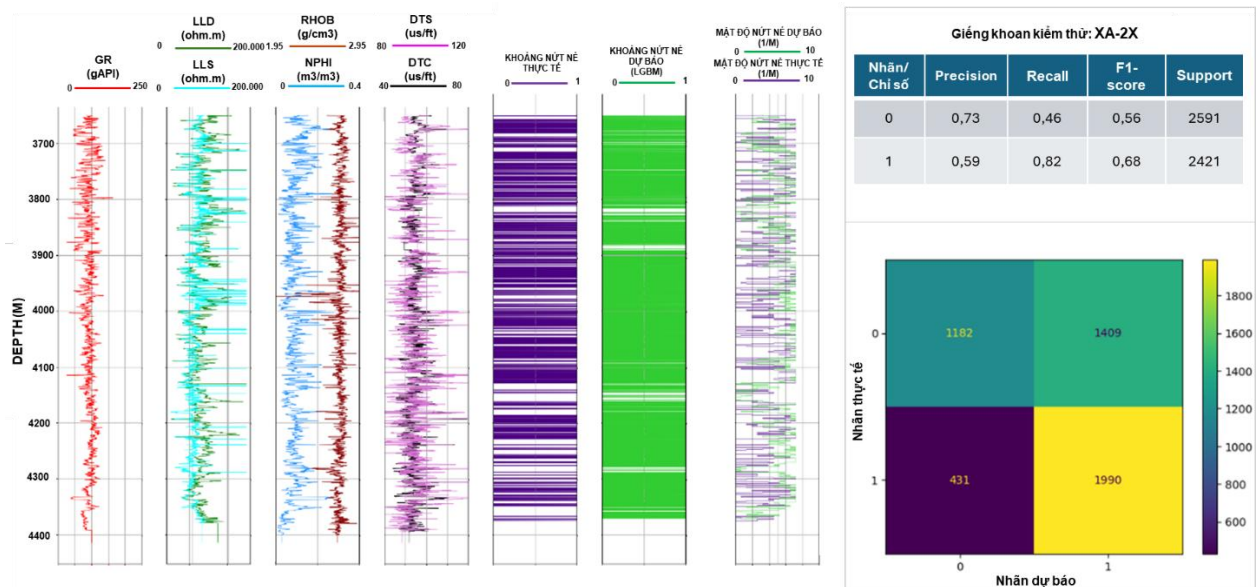
	Huấn luyện	Kiểm tra	Kiểm thử
Recall	0,997	0,909	0,909
Precision	0,997	0,908	0,908
Accuracy	0,997	0,909	0,909

Để có cái nhìn toàn diện về kết quả dự báo, nhóm tác giả đã tiến hành dự báo cho từng giếng riêng lẻ và trình bày kết quả phân loại cùng ma trận nhầm lẫn (confusion matrix) của mô hình sau khi huấn luyện cho từng giếng. Kết quả quan sát cho thấy tất cả các giếng đều đạt Recall trên 70%, chứng tỏ mô hình học máy được xây dựng hoàn toàn đáp ứng các yêu cầu đề ra và có độ tin cậy (Bảng 5); tập kiểm thử (testing) cũng đạt độ chính xác >0,7 với hầu hết giếng (ví dụ ở Hình 5, giếng khoan 8 đạt 0,74), ngoại trừ giếng khoan số 9 chỉ đạt 0,67 do khác biệt vị trí và mất cân bằng dữ liệu, nhưng hiện tượng này chỉ xảy ra duy nhất ở giếng này nên tính phù hợp tổng thể của mô hình vẫn được đảm bảo. Sau huấn luyện, mô hình tối ưu được lưu dưới dạng .json để dự đoán cho các giếng chưa dán nhãn; toàn bộ thông số được lưu trên Neptune để chuyên gia giám sát và kiểm tra quá trình xây dựng mô hình.

3. Kết quả đạt được và thảo luận

3.1. Khoảng nứt nẻ và mật độ nứt nẻ

Để đánh giá và kiểm chứng độ tin cậy của mô hình dự báo nứt nẻ, nhóm tác giả đã lựa chọn một giếng khoan XA-2X thuộc khu vực mỏ lân cận. Phân bố dữ liệu ĐVLGK tại giếng khoan XA-2X nằm trong vùng dữ liệu huấn luyện của mô hình và có đầy đủ tất cả các đường cong ĐVLGK. Kết quả dự báo từ mô hình tối ưu cho độ chính xác đạt 82%. Điều này chứng tỏ mô hình đạt độ tin cậy cao trong kiểm chứng blind test và có thể áp dụng cho các tập dữ liệu mở rộng. Sau đó, mật độ nứt nẻ dự báo được tính ngược lại từ kết quả dự báo khoảng nứt nẻ bằng cách áp dụng kỹ thuật cửa sổ trượt với kích thước 1 m và bước nhảy 0,1524 m, tương tự như quy trình dán nhãn dữ liệu. Kết quả cho thấy, mật độ nứt nẻ dự báo từ mô hình thể hiện sự tương đồng đáng kể với mật độ nứt nẻ xác định từ tài liệu BHI (Hình 6). Những kết quả này cho thấy mô hình dự báo khoảng nứt nẻ và mật độ nứt nẻ từ tài liệu ĐVLGK kết hợp dán nhãn bằng tài liệu BHI, đã được kiểm chứng với độ chính xác cao, ổn định và hoàn toàn đáp ứng yêu cầu thực tế.



Hình 6. Kết quả dự báo của mô hình cho giếng khoan kiểm thử XA-2X.

3.2. Đặc tính độ rỗng, độ thấm của nứt nẻ

Ngoài việc xác định khoảng và mật độ nứt nẻ, các đặc tính nứt nẻ như độ rỗng và độ thấm cũng cần được tính toán để phục vụ cho công tác tính toán tài nguyên dầu khí và xây dựng mô hình mỏ. Việc xác định độ rỗng nứt nẻ từ mật độ nứt nẻ dự báo đòi hỏi thông số độ mở khe nứt (aperture). Tuy nhiên, trong khu vực nghiên cứu chỉ có hai giếng (giếng khoan 1 và 2) có dữ liệu độ mở chiếm khoảng 10% trên tổng số điểm dữ liệu. Ngoài ra, giếng khoan 1 thiếu nhiều đường cong cơ bản (RHOB, NPHI, DTC), do đó chỉ còn giếng khoan 2 có dữ liệu đầy đủ. Khối lượng dữ liệu hạn chế này không đủ để xây dựng mô hình học máy dự báo độ mở dựa trên tài liệu ĐVLGK, vì vậy, việc tính toán độ rỗng nứt nẻ từ mật độ nứt nẻ được tính toán theo công thức thực nghiệm [20] cho 2 giếng khoan 1 và 2.

$$\text{PHI}_{fr} = 0.001 * \text{Frac_Aperture} * \text{Frac_Density} * \text{KF1}$$

$$\text{Kfrac} = \frac{833 * 1011 * \text{PHI}_{fr}}{3 / (\text{Frac_Density}^2 * \text{KF1}^2)}$$

Trong đó:

- PHI_fr: độ rỗng nứt nẻ (%);
- Kfrac: độ thấm nứt nẻ (mD);
- Frac_Aperture (FVA): độ mở khe nứt (mm);
- Frac_Density: mật độ nứt nẻ (số nứt nẻ/m);
- KF1 = số hướng chính của khe nứt; =1 đối với khe nứt gần nằm ngang hoặc vuông góc, =3 đối với khe nứt nằm lộn xộn, đối đập vỡ.

Kết quả cho thấy, độ rỗng nứt nẻ trong khu vực nghiên cứu nhìn chung rất thấp và phân tán mạnh. Kết quả thống kê cho thấy giá trị trung bình chỉ khoảng $6,55 \times 10^{-6}$ (tương đương 0,000655 %), với phần lớn số điểm bằng 0 (chặt xít, không có nứt nẻ hiệu dụng). Giá trị lớn nhất ghi nhận được là 0,001246 ($\approx 0,1246$ %), trong khi phân vị 95 % chỉ đạt $2,52 \times 10^{-5}$ ($\approx 0,0025$ %). Phân bố này có độ lệch phải rất cao (Skewness $\approx 11,8$; Kurtosis ≈ 210), phản ánh đặc điểm phổ biến của đá móng nứt nẻ: phần lớn thể tích đá gần như chặt xít, chỉ có số khoảng xuất hiện các khe nứt mở tạo nên độ rỗng nứt nẻ hiệu dụng. Độ thấm nứt nẻ thể hiện sự biến thiên lớn và phân bố lệch phải mạnh. Kết quả thống kê cho thấy giá trị trung bình khoảng 19,73 mD, nhưng phần lớn các giá trị rất thấp, median chỉ 0,3246 mD.

Phân vị 5% chỉ 0,001314 mD, trong khi phân vị 95% đạt 67,44 mD. Giá trị cực đại lên tới 1988,36 mD, độ lệch chuẩn lớn (112,25 mD) cùng skewness $\approx 12,26$ và kurtosis $\approx 176,9$ phản ánh phân bố có đuôi dài: đa số điểm có độ thấm rất nhỏ, nhưng tồn tại một số ít khoảng/nứt nẻ có độ thấm cực cao.

3.3. Thảo luận

Những kết quả trên đóng góp những điểm mới quan trọng khi lần đầu tiên xây dựng mô hình học máy có khả năng dự báo đồng thời khoảng nứt nẻ và mật độ nứt nẻ của tầng móng granitoid dựa trên các đường cong ĐVLGK thông thường, khắc phục hạn chế lớn về thiếu dữ liệu BHI. Quy trình được đề xuất kết hợp tối ưu giữa LightGBM và phân tích SHAP nhằm giải thích hóa mô hình, bảo đảm tính phù hợp địa chất và khả năng tổng quát hóa cao. Bộ dữ liệu chuẩn hóa cùng hệ thống nhãn nứt nẻ được xây dựng từ nguồn BHI, cho phép mô hình áp dụng linh hoạt cho các giếng chưa đo BHI và có khả năng mở rộng sang các khu vực khác trong bể Cửu Long.

Đây là điểm đặc biệt quan trọng bởi tại nhiều khu vực trong bể Cửu Long, chỉ một số ít giếng được đo ảnh thành hệ do chi phí cao và yêu cầu kỹ thuật phức tạp. Điều này không chỉ giảm đáng kể thời gian xử lý – minh giải truyền thống, mà còn tiết kiệm chi phí đo đạc, cho phép mở rộng ứng dụng sang nhiều khu vực và các mỏ có điều kiện địa chất tương tự. Mô hình vì vậy đóng vai trò như một công cụ hỗ trợ hiệu quả, linh hoạt và có tính ứng dụng cao trong công tác đánh giá đặc tính nứt nẻ của tầng móng ở quy mô mỏ, bể trầm tích.

Về ý nghĩa thực tiễn, mô hình tối ưu đề xuất cung cấp một công cụ định lượng giúp chuẩn hóa và tự động hóa quá trình nhận diện nhanh khoảng nứt nẻ và mật độ nứt nẻ, đánh giá via, hỗ trợ xây dựng mô hình via chứa và tối ưu kế hoạch khai thác trong tầng móng nứt nẻ. Các kết quả này có thể được xem là nền tảng để mở rộng sang các bài toán liên quan như phân vùng rủi ro khoan, dự báo khoảng chiều sâu mất dung dịch khoan và đánh giá tiềm năng dầu khí bị bỏ sót ở các mỏ có điều kiện tương tự.

Trong các nghiên cứu tiếp theo, mô hình có thể tiếp tục được cải thiện khi bổ sung dữ liệu

huấn luyện và dữ liệu gán nhãn, đồng thời xem xét đầy đủ hơn các yếu tố ảnh hưởng như điều kiện địa chất, đặc điểm thạch học–biến đổi thứ sinh và điều kiện dung dịch khoan. Việc tăng cường kiểm chứng độc lập bằng các nguồn dữ liệu khác (tài liệu khai thác, thử vỉa DST/PLT, mất dung dịch, biểu hiện dầu khí) sẽ giúp nâng cao độ tin cậy và đánh giá khả năng tổng quát hóa theo từng khu vực/mỏ. Ngoài ra, có thể tích hợp thêm các đặc trưng địa cơ học và thông tin khoan để cải thiện độ ổn định dự báo trong trường hợp thiếu một số đường cong đầu vào, cũng như phát triển mô hình theo hướng cập nhật liên tục khi có thêm dữ liệu.

4. Kết luận

Nghiên cứu đã thu thập và chuẩn hóa toàn bộ tài liệu ĐVLGK, phân tích ảnh thành hệ, biểu hiện dầu khí tại các giếng khoan mỏ SA, bể Cửu Long phục vụ xây dựng mô hình dự báo khoảng nứt nẻ và mật độ nứt nẻ ứng dụng trí tuệ nhân tạo và học máy. Hệ phương pháp và thông số mô hình đã được tối ưu hóa trên toàn bộ dữ liệu giếng khoan hiện có, đạt độ chính xác cao và tạo tiền đề áp dụng cho các khu vực khác của bể Cửu Long. Kết quả nghiên cứu cho thấy chất lượng và mức độ đầy đủ của dữ liệu đầu vào là yếu tố then chốt quyết định chất lượng dự báo. Thuật toán LightGBM cho kết quả tốt hơn các thuật toán khác, phù hợp với dữ liệu và bài toán nghiên cứu và được lựa chọn để xây dựng mô hình. Kết quả dự báo khoảng nứt nẻ từ tài liệu ĐVLGK đạt độ chính xác cao trên tập kiểm thử (>70%) và giếng khoan kiểm thử (82%) và mật độ nứt nẻ dự báo có độ tương đồng cao với mật độ nứt nẻ thực tế, chứng minh sự phù hợp và độ tin cậy của mô hình tối ưu. Theo đó độ rỗng và độ thấm nứt nẻ có thể tính từ mật độ và độ mở nứt nẻ và cần tiếp tục mở rộng cơ sở dữ liệu để hoàn thiện mô hình học máy cho các tham số này.

Trên cơ sở đó, nhóm tác giả kiến nghị mở rộng kết quả nghiên cứu áp dụng cho các đối tượng khác nhau (đá trầm tích lục nguyên, carbonat) và phạm vi lớn hơn (các mỏ, các bể trầm tích). Bên cạnh đó, cần bổ sung và mở rộng thêm tập dữ liệu đầu vào bao gồm khoảng nứt nẻ,

mật độ nứt nẻ và độ mở nứt nẻ từ tài liệu BHI ở các khu vực mỏ lân cận ở bể Cửu Long, đồng thời áp dụng các kỹ thuật xử lý, cân bằng số liệu và các mô hình học sâu để nâng cao độ tin cậy dự báo. Việc phát triển các mô hình dự báo dựa trên dữ liệu đã dán nhãn từ nghiên cứu này có thể tiếp tục phát triển và mở rộng sang các bài toán dự báo thạch học và đặc tính vỉa chứa phục vụ tìm kiếm – khai thác dầu khí.

Lời cảm ơn

Nhóm tác giả xin chân thành cảm ơn Viện Dầu khí Việt Nam đã giao và hỗ trợ kinh phí thực hiện nghiên cứu này, cũng như trân trọng ghi nhận sự đóng góp tích cực của các tác giả và sự hỗ trợ chuyên môn quý báu từ các chuyên gia trong lĩnh vực Tìm kiếm – Thăm dò và Khoa học Dữ liệu.

Tài liệu tham khảo

- [1] B. Tokhmechi, H. Memarian, H. A. Noubari, B. Moshiri, A Novel Approach Proposed for Fractured Zone Detection Using Petrophysical Logs, *Journal of Geophysics and Engineering*, Vol. 6, No. 4, 2009, pp. 365-373, <https://doi.org/10.1088/1742-2132/6/4/004>.
- [2] H. H. Alkinani, A. T. T. A. Hameedi, W. H. A. Bazzaz, S. D. Norman, A. A. A. Khateeb, Using Artificial Neural Networks to Estimate Mud Losses Prior to Drilling for Natural Fractures Formations, 2019 AADE National Technical Conference and Exhibition, Denver, USA, 2019, https://www.aade.org/application/files/3415/7132/0376/AADE-19-NTCE-022_-_Alkinani.pdf (accessed on: September 1st, 2025).
- [3] G. Aghli, B. Soleimani, R. M. Harami, R. Mohammadian, Fractured Zones Detection Using Conventional Petrophysical Logs by Differentiation Method and its Correlation with Image Logs, *Journal of Petroleum Science and Engineering*, Vol. 142, 2016, pp. 152-162, <https://doi.org/10.1016/j.petrol.2016.02.002>.
- [4] G. Aghli, R. M. Harami, S. Mortazavi, R. Mohammadian, Evaluation of New Method for Estimation of Fracture Parameters Using Conventional Petrophysical Logs and ANFIS in the Carbonate Heterogeneous Reservoirs, *Journal*

- of Petroleum Science and Engineering, Vol. 172, 2019, pp. 1092-1102, <https://doi.org/10.1016/j.petrol.2018.09.017>.
- [5] R. Taherdangkoo, M. Abdideh, Fracture Density Estimation from Well Logs Data Using Regression Analysis: Validation Based on Image Logs (Case study: South West Iran), *International Journal of Petroleum Engineering*, Vol. 2, No. 4, 2016, pp. 289-301, <https://doi.org/10.1504/IJPE.2016.084117>.
- [6] S. Dong, L. Zeng, W. Lyu, D. Xia, G. Liu, Y. Wu, X. Du, Fracture Identification and Evaluation Using Conventional Logs in Tight Sandstones: A Case Study in the Ordos Basin, China, *Energy Geoscience*, Vol. 1, No. 3-4, 2020, pp. 115-123, <https://doi.org/10.1016/j.engeos.2020.06.003>.
- [7] T. Li, R. Wang, Z. Wang, M. Zhao, L. Li, Prediction of Fracture Density Using Genetic Algorithm Support Vector Machine Based on Acoustic Logging Data, *Geophysics*, Vol. 83, No. 2, 2018, pp. D49-D60, <https://doi.org/10.1190/geo2017-0229.1>.
- [8] B. Tokhmechi, H. Memarian, V. Rasouli, H. A. Noubari, B. Moshiri, Fracture Detection from Water Saturation Log Data Using A Fourier-wavelet Approach, *Journal of Petroleum Science and Engineering*, Vol. 69, No. 1-2, 2009, pp. 129-138, <https://doi.org/10.1016/j.petrol.2009.08.005>.
- [9] B. Tokhmechi, H. Memarian, M. R. Rezaee, Estimation of the Fracture Density in Fractured Zones Using Petrophysical Logs, *Journal of Petroleum Science and Engineering*, Vol. 72, No. 1-2, 2010, pp. 206-213, <https://doi.org/10.1016/j.petrol.2010.03.018>.
- [10] N. T. Huyen, T. D. Cuong, T. X. Cuong, N. T. Hieu, P. T. Hong, N. T. M. Hong, L. H. An, H. A. Tuan, Using Artificial Neural Networks (ANN) to Predict the Distribution Characteristics and Reservoir Quality of Miocene Carbonate Reservoirs in the Phu Khanh Sedimentary Basin, *Petrovietnam Journal*, No. 5, 2019, pp. 25-31.
- [11] T. V. Tran, H. H. Ngo, S. K. Hoang, H. N. Tran, J. J. Lambiasi, Depositional Facies Prediction Using Artificial Intelligence to Improve Reservoir Characterization in a Mature Field of Nam Con Son Basin, Offshore Vietnam, *Offshore Technology Conference Asia (OTC Asia)*, Paper OTC-30086-MS, 2020, <https://doi.org/10.4043/30086-MS>.
- [12] N. V. Hung, D. H. Minh, Application of ANN in Predicting Formation Fracture Pressure, *Petrovietnam Journal*, No. 3, 2019, pp. 32-41, <https://www.tapchidaukhi.vn/index.php/TCDK/article/download/245/220> (accessed on: September 1st, 2025).
- [13] D. V. Long, A. R. Joseph, D. D. Phan, G. Hepler, L. C. Trung, B. H. Phuoc, N. T. Bao, D. T. Huy, N. K. Xa, N. H. An, T. V. Lam, N. D. Hung, Application of AI to Predict Perforation Intervals under Limited Well Log Data Conditions in the Te Giac Trang Field, *Petrovietnam Journal*, No. 3, 2025, pp. 5-12, <https://doi.org/10.47800/PVSI.2025.03-01>.
- [14] Nature & Environment, Application of Artificial Intelligence in Oil and Gas Exploration, 2022, <https://thiennhienmoitruong.vn/ung-dung-tri-tue-nhan-tao-trong-tim-kiem-va-tham-do-dau-khi.html>, 2022 (accessed on: September 1st, 2025).
- [15] Industry and Trade Magazine, Vietnam Petroleum Institute Uses Artificial Intelligence to Predict Fractured Basement Reservoirs, <https://congthuong.vn/vien-dau-khi-viet-nam-sung-dung-tri-tue-nhan-tao-de-du-bao-da-mong-nut-ne-218499.html>, 2022 (accessed on: September 1st, 2025).
- [16] N. T. Yen, Artificial Intelligence in the Oil and Gas Industry and Its Application Potential in Vietnam, *Petrovietnam Journal*, No. 12, 2020, pp. 57-64, <https://doi.org/10.47800/PVJ.2020.12-07>.
- [17] S. A. Ahmed, MonaLisa, M. Hussain, Z. U. Khan, Supervised Machine Learning for Predicting Shear Sonic Log (DTS) and Volumes of Petrophysical and Elastic Attributes, Kadanwari Gas Field, Pakistan, *Frontiers in Earth Science*, Vol. 10, 2022, Article 919130, <https://doi.org/10.3389/feart.2022.919130>.
- [18] L. Grinsztajn, E. Oyallon, G. Varoquaux, Why do Tree-based Models Still Outperform Deep Learning on Tabular Data?, *arXiv preprint*, 2022, arXiv:2207.08815, <https://doi.org/10.48550/arXiv.2207.08815>.
- [19] S. M. Lundberg, S. I. Lee, A Unified Approach to Interpreting Model Predictions, *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 30, 2017, pp. 4765-4774, https://papers.nips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html.
- [20] Crain, *Crain's Petrophysical Handbook: Porosity Fracture Aperture Model*, <https://www.spec2000.net/crain.htm> (accessed on: September 1st, 2025).