



## Journal of Mining and Earth Sciences

Website: <https://jmes.humg.edu.vn>

# A review of research on reservoir porosity prediction by machine learning based on real-time drilling data



Duyen Quang Le \*

Hanoi University of Mining and Geology, Hanoi, Vietnam

### ARTICLE INFO

#### Article history:

Received 21<sup>st</sup> May 2024

Revised 11<sup>th</sup> Aug. 2024

Accepted 23<sup>rd</sup> Aug. 2024

#### Keywords:

Drilling data,  
Formation,  
Machine learning,  
Porosity,  
Well log.

### ABSTRACT

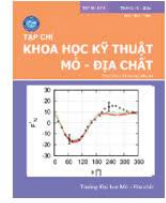
*To exploit oil effectively, it is necessary to determine the reservoir parameters of the formation. One of the important parameters that needs to be determined is porosity. Porosity prediction helps evaluate the reservoir's production performance, select the location of the production well, design enhanced oil recovery and evaluate the reservoir's economic feasibility. Normally, the value of porosity is determined directly by various laboratory core sample tests or indirectly based on the results of interpreting well geophysical measurement documents, and well logs. These traditional identification methods are often time-consuming and expensive. Laboratory testing methods are highly accurate but often require available core samples, require a lot of auxiliary measuring equipment, and sometimes require additional core sample measurement results, which consumes time and sampling costs. Moreover, well logging measurements are not always performed in all production wells. By applying different machine learning techniques such as Artificial Neural Network, Decision Tree, Random Forest, etc., the porosity value is also predicted. These techniques often use input parameters as data from well log curves or drilling data. However, using well geophysical measurement curve data as input parameters for machine learning models often faces limitations from the availability of data sources. Meanwhile, drilling parameters such as rate of penetration, weight on bit, drillstring or drillpipe rotation speed measured in revolutions/minute, torque, pumping rate of the circulation of the drilling fluid measured in gallons/minute, and the resulting standpipe pressure are continuously collected from measured while drilling sensors. The article focuses on evaluating and analyzing scientific works that have been researched on the application of machine learning techniques to predict formation porosity values, based on real-time drilling data.*

Copyright © 2024 Hanoi University of Mining and Geology. All rights reserved.

\*Corresponding author

E - mail: [lequangduyen@humg.edu.vn](mailto:lequangduyen@humg.edu.vn)

DOI: 10.46326/JMES.2024.65(5).02



## Tạp chí Khoa học Kỹ thuật Mỏ - Địa chất

Trang điện tử: <https://tapchi.humg.edu.vn>

# Tổng quan dự báo độ rỗng thành hệ bằng kỹ thuật máy học dựa trên dữ liệu khoan thực

Lê Quang Duyệt \*

Trường Đại học Mỏ - Địa chất, Hà Nội, Việt Nam

### THÔNG TIN BÀI BÁO

### TÓM TẮT

#### Quá trình:

Nhận bài 21/5/2024

Sửa xong 11/8/2024

Chấp nhận đăng 23/8/2024

#### Từ khóa:

Dữ liệu khoan,

Độ rỗng,

Đường cong đo địa vật lý

giếng khoan,

Máy học,

Thành hệ.

Để khai thác dầu hiệu quả, cần thiết phải xác định các thông số vỉa của thành hệ. Một trong các thông số quan trọng cần phải được xác định đó là độ rỗng. Việc dự báo độ rỗng giúp đánh giá khả năng khai thác của vỉa chứa, lựa chọn vị trí giếng khai thác, thiết kế thu hồi dầu tăng cường và đánh giá tính khả thi về kinh tế. Thông thường, giá trị của độ rỗng được xác định trực tiếp bằng các phép thí nghiệm mẫu lõi trong phòng khác nhau hoặc một cách gián tiếp dựa trên kết quả minh giải tài liệu đo địa vật lý giếng khoan. Các phương pháp xác định truyền thống này thường tốn nhiều thời gian và gây tốn kém. Phương pháp thí nghiệm trong phòng có độ chính xác cao nhưng thường yêu cầu có sẵn mẫu thí nghiệm, đòi hỏi nhiều thiết bị đo phụ trợ và đôi khi yêu cầu kết quả đo mẫu lõi bổ sung tiêu tốn thời gian và chi phí lấy mẫu. Phương pháp đo địa vật lý giếng khoan không phải lúc nào cũng được thực hiện ở tất cả các giếng khoan. Bằng cách ứng dụng các kỹ thuật máy học khác nhau như Mạng nơ-ron nhân tạo, Cây quyết định, Rừng ngẫu nhiên,... giá trị độ rỗng cũng được dự báo. Các kỹ thuật này thường sử dụng thông số đầu vào là dữ liệu từ đường cong đo địa vật lý giếng khoan hoặc dữ liệu khoan. Tuy nhiên, việc sử dụng dữ liệu đường cong đo địa vật lý giếng khoan làm thông số đầu vào cho mô hình máy học thường gặp phải hạn chế từ tính sẵn có của nguồn dữ liệu. Trong khi đó các thông số khoan như vận tốc cơ học khoan, lưu lượng bơm của dung dịch khoan, tốc độ vòng quay trên phút, áp suất trong cần dựng, mô-men xoắn và tải trọng chiều trục được thu thập một cách liên tục từ các cảm biến dữ liệu trong quá trình khoan. Bài báo tập trung đánh giá, phân tích những công trình khoa học đã được nghiên cứu về việc ứng dụng kỹ thuật máy học để dự báo giá trị độ rỗng của thành hệ, dựa trên dữ liệu khoan theo thời gian thực.

© 2024 Trường Đại học Mỏ - Địa chất. Tất cả các quyền được bảo đảm.

\*Tác giả liên hệ

E - mail: [lequangduyen@humg.edu.vn](mailto:lequangduyen@humg.edu.vn)

DOI: 10.46326/JMES.2024.65(5).02

## 1. Mở đầu

Trong việc xác định đặc tính vỉa chứa và thăm dò hydrocarbon thì việc xác định độ rỗng là rất quan trọng (Denney, 2003; Lucia và nnk., 2003). Giá trị độ rỗng được sử dụng để đánh giá lượng

dầu và/hoặc khí có trong vỉa chứa, lượng hydrocarbon có thể thu hồi được, thiết kế khai thác và dự báo khai thác trong tương lai (Olatunji và nnk., 2011).

Độ rỗng của vỉa chứa được định nghĩa là phần không gian trống trong thân đá và được tính bằng tỷ số giữa thể tích lỗ rỗng của không gian trống với thể tích khối của thân đá. Độ rỗng là thông số xác định không gian lỗ rỗng trong thành hệ và ảnh hưởng trực tiếp đến khả năng lưu trữ chất lưu trong vỉa chứa.

Việc dự báo giá trị độ rỗng chính xác giúp đưa ra quyết định lựa chọn vị trí giếng và thiết kế thu hồi dầu tăng cường. Thông thường, độ rỗng được xác định thông qua các thí nghiệm mẫu lõi trong phòng. Độ rỗng được đo trực tiếp bằng các kỹ thuật khác nhau như đo độ rỗng bằng helium, đo độ rỗng xâm nhập thủy ngân hoặc phương pháp gián nở khí (Ramana và Venkatanarayana, 1971). Mặc dù phương pháp này được cho là chính xác nhất nhưng thường gây tăng chi phí, tốn nhiều công sức, thời gian và đòi hỏi nhiều thiết bị đo phụ trợ. Ngoài ra, các phép đo này yêu cầu có sẵn mẫu thí nghiệm và đôi khi yêu cầu kết quả đo mẫu lõi bổ sung, gây tiêu tốn thời gian và chi phí lấy mẫu.

Các đường cong đo địa vật lý giếng khoan, chẳng hạn như đường cong đo độ rỗng neutron, đường cong đo mật độ và đường cong đo sóng âm, có thể được sử dụng để dự báo độ rỗng. Những đường cong đo địa vật lý giếng khoan này đo lường các đặc tính vật lý khác nhau của thành hệ và giúp xác định gián tiếp giá trị độ rỗng (Ellis và nnk., 2004; Fischetti và Andrade, 2002).

Các kỹ thuật địa chấn, chẳng hạn như trở kháng âm nghịch, có thể được sử dụng để dự báo độ rỗng một cách gián tiếp. Bằng cách phân tích phản xạ sóng địa chấn của vỉa chứa, thông tin về độ rỗng có thể được suy luận dựa trên tính chất đá và hàm lượng chất lỏng (Angeleri và Carpi, 1982; Doyen, 1988).

Việc dự đoán chính xác độ rỗng là một thách thức do tính không đồng nhất vốn có của các thành hệ dưới bề mặt. Các phương pháp truyền thống, chẳng hạn như thí nghiệm mẫu lõi, đo địa vật lý giếng khoan và phương pháp địa chấn, tốn nhiều thời gian và tốn kém, khiến việc đo đạc ở mọi vị trí trong vỉa chứa là không thực tế (Hassaan và nnk., 2024)

Bên cạnh đó, giá trị độ rỗng thu được từ các dụng cụ đo địa vật lý giếng khoan cũng có sai sót

do lỗi hiệu chỉnh và thiết kế công cụ đo không đúng cũng như thu thập dữ liệu không đúng cách (Kane và Jennings, 2005), do ảnh hưởng của giếng khoan hoặc lượng dung dịch khoan tới đường cong đo (Bonnecaze và nnk., 2002). Các đường cong đo độ rỗng không đầy đủ hoặc bị thiếu hoàn toàn từ các hoạt động khoan là phổ biến. Điều này có thể là do các sự cố gặp phải với máy ghi đường cong đo, chẳng hạn như cảm biến bị lỗi, điều kiện ghi đường cong đo kém và thiết bị bị hỏng. (Sun và nnk., 2021).

Giá trị độ rỗng cũng có thể được dự báo thông qua thuật toán trí tuệ nhân tạo dựa trên dữ liệu đầu vào là các đường cong đo địa vật lý giếng khoan như đường cong gamma, điện trở suất), độ rỗng (mật độ, neutron và log âm) và độ bão hòa chất lỏng (Alyafei, 2021).

Ngoài ra, độ rỗng còn được dự báo bằng kỹ thuật máy học dựa trên các dữ liệu khoan theo thời gian thực như vận tốc cơ học khoan (ROP), lưu lượng bơm của dung dịch khoan (GPM), tốc độ vòng quay trên phút (RPM), áp suất trong cần dụng (SPP), mô-men xoắn (T) và tải trọng chiều trục (WOB) (Hassaan và nnk., 2024).

Việc đánh giá độ rỗng theo thời gian thực đóng vai trò then chốt trong quản lý vỉa và thăm dò hydrocarbon. Khả năng xác định nhanh chóng và chính xác độ rỗng rất cần thiết cho các quyết định kịp thời, sáng suốt, ảnh hưởng trực tiếp đến việc khai thác vỉa chứa và tính khả thi về mặt kinh tế. Việc đánh giá liên tục độ rỗng theo thời gian thực hỗ trợ tối ưu hóa vị trí giếng và đưa ra các giải pháp thu hồi dầu nâng cao hiệu quả một cách linh hoạt. Việc này cũng cho phép nhà thầu vận hành nhanh chóng điều chỉnh các phương pháp khoan và khai thác, phù hợp với điều kiện vỉa chứa. Sự linh hoạt trong việc ra quyết định này góp phần tối ưu hóa việc thu hồi hydrocarbon, giảm thiểu rủi ro vận hành và đảm bảo thành công chung của phát triển mỏ. Hơn nữa, việc đánh giá theo thời gian thực về độ rỗng là điều cần thiết để nâng cao hiệu quả quản lý vỉa chứa, tạo điều kiện thuận lợi cho việc xác định ngay lập tức những thay đổi về đặc tính vỉa chứa, cho phép chủ động sửa đổi các hoạt động khoan, hoàn thiện giếng và chiến lược khai thác (Hassaan và nnk., 2024)

Dữ liệu khoan dễ dàng thu thập được một cách đầy đủ từ cảm biến đo theo thời gian thực trên bề mặt. Các dữ liệu này, khi kết hợp với các thuật toán máy học thích hợp, có thể đóng vai trò

là đầu vào để dự đoán độ rỗng. Việc sử dụng kỹ thuật máy học dựa trên dữ liệu có sẵn để tạo ra một mô hình dự đoán độ rỗng này không cần thêm chi phí hoặc tiến hành phép đo nào trong giếng khoan.

## 2. Kỹ thuật máy học

Máy học (ML) là một nhánh của trí tuệ nhân tạo sử dụng dữ liệu đầu vào để huấn luyện mô hình và tạo ra dự đoán, tiết kiệm chi phí và thời gian. Các thuật toán máy học (ML) đưa ra các giải pháp đầy hứa hẹn cho các vấn đề khác nhau bằng cách phân tích lượng dữ liệu khổng lồ và xác định các mô hình và mối quan hệ phức tạp mà các nhà phân tích có thể không rõ ràng (Gurina và nnk., 2020; Noshi và Schubert, 2018; Sircar và nnk., 2021).

Bằng cách đào tạo các mô hình trên một tập dữ liệu bao gồm cả tham số đầu vào và phép đo đầu ra tương ứng, các thuật toán có thể tìm hiểu các mẫu cơ bản và tạo ra các mô hình dự đoán có khả năng dự báo các đặc tính hình thành.

Trong số các kỹ thuật máy học, phương pháp thường được sử dụng để dự báo độ rỗng của thành hệ là mạng nơ-ron nhân tạo (ANN), Cây quyết định (DT), Rừng ngẫu nhiên (RF) và Máy vector hỗ trợ (SVM).

*Mạng nơ-ron nhân tạo (ANN):* ANN là một kỹ thuật máy học cố gắng bắt chước các đặc tính hoạt động của hệ thần kinh sinh học (Nakamoto, 2018). Kỹ thuật này có khả năng cao để giải quyết các vấn đề kỹ thuật phức tạp mà không thể giải quyết được bằng cách sử dụng các phép tính toán học cổ điển với các quy trình truyền thống (Andagoya Carrillo và nnk., 2015). Bất kỳ ANN nào cũng bao gồm các phần tử cơ bản được gọi là nơ-ron và ANN được hình thành bởi một tập hợp các nơ-ron được sắp xếp với các hình dạng cụ thể và các kết nối giữa các nơ-ron này có liên quan đến trọng số và độ lệch (Hinton và nnk., 2006). Việc xác định số lượng nơ-ron cùng với trọng số và độ lệch là rất quan trọng trong việc tối ưu hóa hiệu suất của ANN. Kiến trúc của ANN bao gồm tối thiểu ba lớp: lớp đầu vào, lớp ẩn và lớp đầu ra và các lớp này được kết nối bằng các hàm truyền (Lippmann, 1987).

*Cây quyết định (DT):* DT là một thuật toán máy học được giám sát, xây dựng mô hình dạng cây để đưa ra quyết định hoặc dự đoán dựa trên các tính năng đầu vào. phân vùng dữ liệu dựa trên các giá trị thuộc tính khác nhau và tạo ra cấu trúc

quyết định phân cấp. Mỗi nút bên trong thể hiện một quyết định dựa trên một tính năng cụ thể và mỗi nút lá thể hiện một quyết định hoặc dự đoán cuối cùng. Cây quyết định rất phổ biến do tính dễ hiểu và khả năng xử lý cả dữ liệu số và dữ liệu phân loại. Chúng có thể được sử dụng cho các nhiệm vụ phân loại và hồi quy. DT đã được áp dụng trong các bài toán máy học có giám sát khác nhau (Barjouei và nnk., 2021).

*Rừng ngẫu nhiên (RF):* Thuật toán hồi quy rừng ngẫu nhiên (RFR) đánh giá một tập hợp cây quyết định và ban đầu được đề xuất bởi Breiman (2001). RFR dựa trên khái niệm rằng mỗi cây quyết định độc lập riêng biệt sử dụng một bộ dự đoán (các biến đầu vào có sẵn) có khả năng dự báo biến phụ thuộc có phần không chính xác ở một mức độ nào đó. Bằng cách kết hợp kết quả của nhiều cây bằng cách sử dụng các tổ hợp dự đoán khác nhau và xem xét kết quả của tập hợp các cây, độ chính xác của dự đoán thường được cải thiện. Các dự đoán được cải thiện có thể được tạo ra bằng cách kết hợp các cấu trúc cây khác nhau với tính năng đóng bao (tập hợp bootstrap) và lựa chọn tính năng ngẫu nhiên để xây dựng một tập hợp cây quyết định ngẫu nhiên. RFR đưa ra quyết định tại mỗi nút của cây quyết định bằng cách sử dụng một tập hợp con dự đoán ngẫu nhiên, có tác dụng giảm thiểu việc khái quát hóa các lỗi trên nhiều cây (Chen và nnk., 2020). Hiệu suất RF có thể được cải thiện bằng cách tối ưu hóa các siêu tham số. Rừng ngẫu nhiên được biết đến với tính mạnh mẽ vì làm giảm việc trang bị quá mức bằng cách kết hợp các mô hình đa dạng. có thể xử lý dữ liệu nhiều chiều và cung cấp thứ hạng tầm quan trọng của tính năng, giúp hữu ích cho các vấn đề phân loại và hồi quy. Rừng ngẫu nhiên được sử dụng rộng rãi do tính linh hoạt và khả năng xử lý các tập dữ liệu lớn (Fu và nnk., 2021; Nallathambi và Ramasamy, 2017).

*Máy vector hỗ trợ (SVM):* SVM là một thuật toán máy học có giám sát mạnh mẽ được sử dụng cho các nhiệm vụ phân loại và hồi quy (Ma và Guo, 2014). SVM nhằm mục đích tìm ra một siêu phẳng tối ưu để phân tách các điểm dữ liệu của các lớp khác nhau hoặc dự đoán một biến mục tiêu liên tục. Thuật toán tìm cách tối đa hóa lề giữa siêu phẳng và các điểm dữ liệu gần nhất, được gọi là vector hỗ trợ. SVM có thể xử lý cả nhiệm vụ phân loại tuyến tính và phi tuyến thông qua việc sử dụng các hàm kernel, giúp chuyển đổi dữ liệu

thành không gian có chiều cao hơn (Patle và Chouhan, 2013). Điều này cho phép SVM phân loại một cách hiệu quả dữ liệu có thể không thể tách tuyến tính khỏi không gian đặc trưng ban đầu. SVM được biết đến với khả năng xử lý dữ liệu nhiều chiều và hiệu suất tổng quát hóa tốt. SVM mang đến sự linh hoạt, mạnh mẽ trước các ngoại lệ và khả năng nắm bắt các ranh giới quyết định phức tạp. Để tối ưu hóa các công cụ này, các siêu tham số khác nhau đã được điều chỉnh cho từng kỹ thuật ML để cải thiện hiệu suất của mô hình (Ghorbani và nnk., 2020).

### 3. Dự báo độ rỗng dựa trên thông số địa vật lý giếng khoan

Các nghiên cứu gần đây triển khai các ưu điểm của máy học với dữ liệu đầu khí để dự đoán độ rỗng của thành hệ từ dữ liệu đường cong đo địa vật lý thông qua việc xây dựng các mô hình khác nhau. Đường cong đo địa vật lý giếng khoan và dữ liệu địa tầng được sử dụng làm dữ liệu đầu vào để dự báo độ rỗng bằng mô hình máy học (Elkatatny và nnk., 2018; Wood, 2020; Zerrouki và nnk., 2014).

Dựa trên thông số đầu vào từ đường cong đo địa vật lý giếng khoan (log), nhiều mô hình máy học khác nhau bao gồm mạng lưới thần kinh nhân tạo, thuật toán di truyền, cây quyết định mờ, thuật toán cạnh tranh (ICA), tối ưu hóa bầy hạt (PSO) và các phương pháp kết hợp,... được sử dụng để dự đoán độ rỗng trong các vỉa chứa dầu (Ahmadi và Chen, 2019; Sun và nnk., 2021; Wood, 2020). Tuy nhiên, kỹ thuật này yêu cầu dữ liệu được đo tốt, tốn kém, mất thời gian và không theo thời gian thực trong quá trình khoan.

Kết quả tính toán độ rỗng từ tài liệu địa vật lý giếng khoan của 4 giếng đã khoan được sử dụng làm tham số đầu vào để xây dựng mô hình dự báo độ rỗng từ tài liệu địa chấn trong trầm tích Miocen lô 103, bắc Bể Sông Hồng. Xác định mối quan hệ giữa độ rỗng từ tài liệu ĐVLGK và địa chấn dựa trên phương pháp hồi quy bội và mạng nơ-ron MLFN và PNN để xây dựng mô hình dự báo độ rỗng từ thuộc tính địa chấn. Với mỗi phương pháp, các mô hình đơn lẻ thu được đều có thể sử dụng để tính toán độ rỗng, tuy nhiên, để giảm thiểu yếu tố không chắc chắn cho kết quả dự báo độ rỗng từ tài liệu địa chấn, các tác giả đã đề xuất sử dụng mô hình tích hợp trung bình kết hợp các mô hình hồi quy bội và mạng nơ-ron (Nguyễn và Lê, 2014).

Nyein và Ali Hamada áp dụng hai mạng ANN đã được huấn luyện dự đoán độ rỗng sử dụng năm dữ liệu đường cong đo đầu vào: (Gamma Ray)GR, (Laterlog Deep)LLD, (mật độ) RHOB, (Neutron) NPHI và hệ số quang điện (PEF) với dữ liệu mẫu lõi chỉ từ hai trên tổng số bốn giếng. Dữ liệu đầu vào được phân bổ cho ba bộ đào tạo, xác nhận và kiểm tra: lần lượt là 70%, 15% và 15%. (Nyein & Ali Hamada, 2018).

Elkatatny và nnk.(2018) đã đánh giá các công cụ máy học như mạng nơ-ron nhân tạo (ANN), máy vectơ hỗ trợ (SVM) và hệ thống suy luận mờ thần kinh thích ứng (ANFIS) để dự đoán độ rỗng của vỉa chứa dựa trên dữ liệu đường cong đo địa vật lý giếng khoan. Hơn 1700 phép đo độ rỗng tại hiện trường với dữ liệu đường cong đo đã được sử dụng để đào tạo và thử nghiệm các kỹ thuật máy học. Kết quả thu được cho thấy ANN và ANFIS có thể được sử dụng để dự báo độ rỗng của vỉa dựa trên dữ liệu đường cong đo có hệ số tương quan cao và sai số phần trăm tuyệt đối trung bình thấp. Các thông số đầu vào chính cần thiết để dự báo độ rỗng là mật độ khối, độ rỗng neutron và thời gian truyền sóng. Phương trình toán học được phát triển dựa trên trọng số và độ lệch của mô hình ANN được sử dụng để dự đoán độ rỗng của vỉa chứa dựa trên dữ liệu đường cong đo với hệ số tương quan là 0,98 và AAPE nhỏ hơn 8% (Elkatatny và nnk., 2018).

Hamada và Elshafei xác nhận mô hình ANN có thể được sử dụng để dự đoán độ rỗng dựa trên dữ liệu NMR và đường cong đo địa vật lý giếng khoan thông thường với độ chính xác cao khi so với kết quả đo từ dữ liệu mẫu lõi (Hamada và Elshafei, 2009).

Yazmyradova và đồng nghiệp sử dụng dữ liệu đo địa chấn và đường cong đo địa vật lý giếng khoan làm thông số đầu vào cho mô hình ANN để đánh giá các đặc tính vật lý dầu ở mỏ X, bồn trũng Nam Caspian. Bộ dữ liệu đường cong đo có sẵn bao gồm đường cong đo tia gamma, độ rỗng neutron, mật độ, điện trở suất, tốc độ sóng dọc và đường cong đo tốc độ sóng ngang (Yazmyradova và nnk., 2022).

Việc sử dụng dữ liệu đường cong đo địa vật lý giếng khoan làm thông số đầu vào cho mô hình máy học thường gặp phải hạn chế từ tính sẵn có của nguồn dữ liệu. Những đường cong đo này đôi khi không có sẵn cho tất cả các giếng và yêu cầu các hoạt động đo bổ sung để thu được dữ liệu

đường cong đo. Điều này ảnh hưởng tới chi phí vận hành trong quá trình khoan.

#### 4. Dự báo độ rỗng dựa trên dữ liệu khoan theo thời gian thực

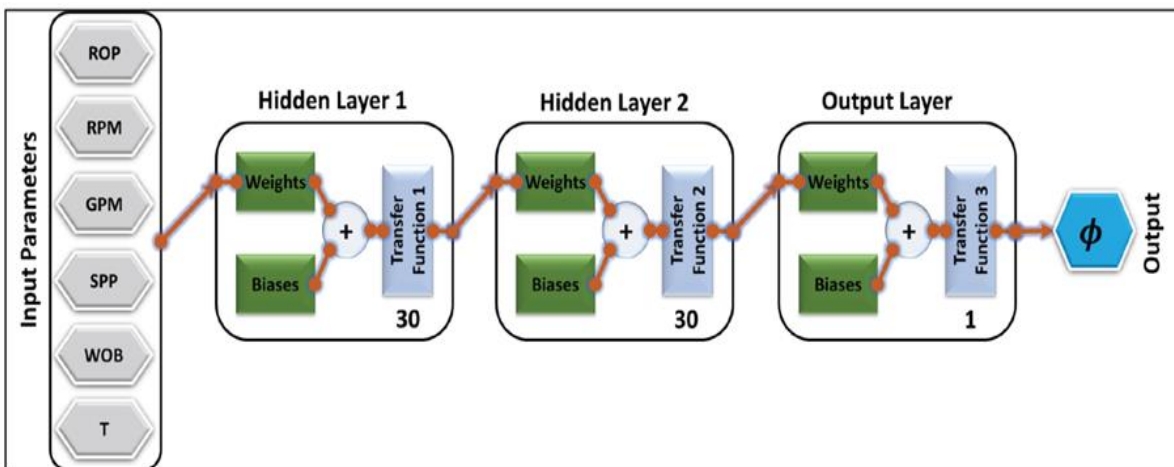
dụng kỹ thuật máy học trong xác định giá trị độ rỗng dựa trên dữ liệu khoan theo thời gian thực. Các dữ liệu khoan thường được sử dụng là vận tốc cơ học khoan (ROP), lưu lượng bơm của dung dịch khoan (GPM), tốc độ vòng quay trên phút (RPM), áp suất trong cần dụng (SPP), mô-men xoắn (T) và tải trọng chiều trục (WOB). Nguồn dữ liệu này luôn có sẵn, dễ dàng và nhanh chóng được thu thập hơn so với đường cong đo địa vật lý giếng khoan và mẫu lõi thông thường.

Al-Abduljabbar và nnk.(2020) sử dụng dữ liệu khoan để xây dựng mô hình dự đoán độ rỗng trong quá trình khoan ngang qua vỉa cacbonat. Tác giả sử dụng mô hình ANN với sai số thấp nhất khi sử dụng hai lớp; với ba lớp trở lên không cải thiện kết quả ngay cả với số lượng tế bào thần kinh khác nhau. Đối với số lượng nơ-ron, phân tích độ nhạy cho thấy 30 nơ-ron cho kết quả tốt nhất cho tập dữ liệu huấn luyện. Hàm truyền được sử dụng là tan-sigmoid (tansig) cho cả hai lớp, cho thấy sự cải thiện đáng kể qua việc sử dụng hàm truyền tuyến tính thuần túy, khi được sử dụng một mình hay kết hợp với hàm truyền khác. Mô hình ANN được lựa chọn sau cùng dựa trên phân tích độ nhạy bao gồm hai lớp ẩn với 30 nơ-ron mỗi lớp với trainlm là hàm huấn luyện và tan-sigmoid là hàm truyền cho cả hai lớp. Kết quả huấn luyện cho hệ số tương

quan lần lượt là 0,96 và 0,94 với tập huấn luyện và kiểm tra. Kiến trúc mạng ANN được sử dụng biểu thị dưới dạng sơ đồ trong Hình 1 (Al-Abduljabbar và nnk., 2020).

Gamal và Elkatatny (2021) phát triển mô hình dựa trên mạng ANN chỉ dựa trên các dữ liệu khoan để dự đoán độ rỗng thành hệ trong quá trình khoan quá các thành hệ cát kết, đá phiến và cacbonat. Nghiên cứu sử dụng hai bộ dữ liệu khác nhau: bộ thứ nhất để xây dựng mô hình (3767 điểm dữ liệu) và bộ thứ hai để phát triển mô hình ANN (1676 điểm dữ liệu). Mô hình ANN được xây dựng và tối ưu hóa với khả năng phân tích độ nhạy cho các tham số của mô hình ANN để đạt được kết quả dự đoán cao. Mô hình ANN được phát triển với hệ số tương quan (R) giữa giá trị độ rỗng dự đoán và thực tế lần lượt là 0,97 và 0,92; với sai số phần trăm tuyệt đối trung bình (AAPE) lần lượt là 6,2 và 9,3% cho tập đào tạo và kiểm tra. Việc hiệu chuẩn mô hình đã nâng cao hiệu suất dự đoán cao khi ANN đạt được R là 0,95 và AAPE là 8,5% (Gamal và Elkatatny, 2021).

Al-Sabaa và nnk. (2021) đã sử dụng mô hình ANN có 1 lớp ẩn với 25 nơ-ron, hàm huấn luyện lan truyền ngược chính quy hóa Bayes, hàm truyền sigmoid tiếp tuyến Hyperbolic nhằm đạt mức độ chính xác cao nhất khi dự báo độ rỗng của thành hệ. Kết quả cho thấy sự chênh lệch giữa giá trị thực tế của độ rỗng so với giá trị dự đoán từ mô hình máy học với độ chính xác lần lượt là 0,93 và 0,91 đối với giá trị R trong quá trình đào tạo và kiểm tra với sai số (AAPE) dưới 6,1 %. (Al-Sabaa và nnk., 2021).



Hình 1. Kiến trúc mô hình mạng ANN dự báo độ rỗng trong thành hệ cacbonat (Al-Abduljabbar và nnk., 2020).

Năm 2021, Gamal và nnk. (2021) đã sử dụng ba phương pháp máy học (máy vectơ hỗ trợ (SVM), hệ thống suy luận mờ dựa trên mạng thích ứng (ANFIS) và mạng chức năng (FN)) để phát triển các mô hình dự đoán độ rỗng của thành hệ cát kết, cacbonat và phiến sét. Dữ liệu khoan đầu vào bao gồm ROP, GPM, RPM, WOB, SPP và T với 3817 phép đo để phát triển mô hình. Ba mô hình đã được tối ưu hóa để nâng cao mức độ chính xác tốt hơn và các mô hình được đánh giá bằng cách tính hệ số tương quan (R), phương sai (VAF), sai số phần trăm tuyệt đối trung bình (AAPE) và sai số bình phương trung bình gốc (RMSE) giữa các phép đo độ rỗng thực tế và các phép đo độ rỗng dự đoán. Giá trị độ rỗng dự đoán thu được từ ba mô hình được phát triển có độ chính xác cao với R lớn hơn 0,92, VAF cao hơn 0,85, AAPE nhỏ hơn 8,34% và RMSE dưới 0,02 (Gamal và nnk., 2021)

Alhowaish, Mezghani và Shakriov sử dụng phương pháp máy học có giám sát với thuật toán tăng cường độ dốc. Các thông số đầu vào chỉ bao gồm các dữ liệu khoan. Trong số 10 giếng, 9 giếng được sử dụng để đào tạo, được chia thành tỷ lệ 9:1 cho tập đào tạo và kiểm tra, và một giếng khác để kiểm chứng mô hình. Mức độ chính xác của mô hình được đo bằng sai số bình phương trung bình gốc (RMSE) và sai số trung bình (MAE). Sau đó, mô hình được đánh giá bằng cách sử dụng giếng hiệu chuẩn dựa trên chỉ số RMSE và MAE. Kết quả hiệu chuẩn cho thấy tính chính xác của mô hình với giá trị sai số RMSE và MAE đều rất thấp, lần lượt là 0.03 và 0.02 (Alhowaish và nnk., 2023).

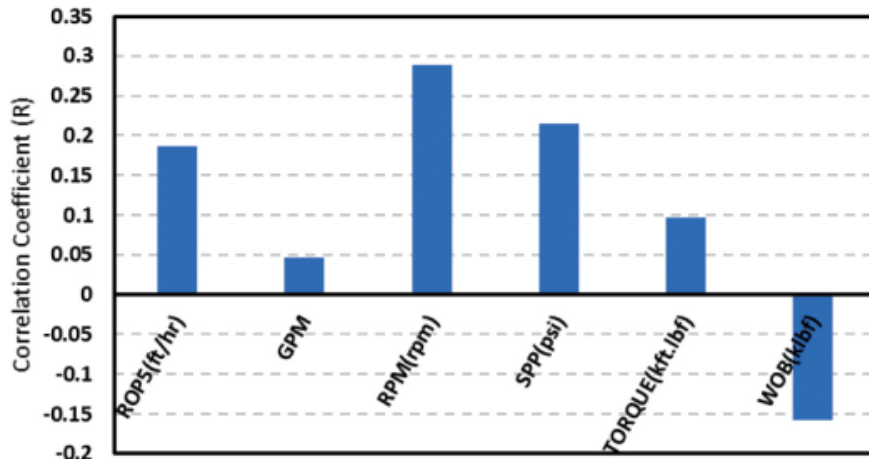
Năm 2024, Hassan và đồng nghiệp sử dụng ba mô hình máy học, cụ thể là cây quyết định (DT), rừng ngẫu nhiên (RF) và máy vectơ hỗ trợ (SVM) để dự đoán độ rỗng của thành hệ. Cả 3 mô hình đều đạt hệ số tương quan (R) cao hơn 0,90 trong dự đoán độ rỗng. Mô hình rừng ngẫu nhiên thể hiện dự đoán chính xác về độ rỗng, đạt giá trị R là 0,93 trong bộ dữ liệu thử nghiệm, 0,94 trong bộ dữ liệu xác thực. Trong khi mô hình DT cho hiệu suất thấp hơn một chút, với giá trị R giảm xuống 0,91 trong tập dữ liệu thử nghiệm và 0,93 trong tập dữ liệu xác thực, mô hình SVM bị trang bị quá mức, với giá trị R giảm xuống 0,92 trong tập dữ liệu thử nghiệm và 0,91 trong bộ dữ liệu xác thực (Hassan và nnk., 2024).

## 5. Mối tương quan giữa dữ liệu khoan và độ rỗng

Khi dự báo giá trị độ rỗng dựa trên dữ liệu khoan theo thời gian thực, quá trình xác định mối tương quan giữa các thông số dữ liệu khoan và độ rỗng có vai trò quan trọng. Việc này giúp đánh giá mức độ ảnh hưởng của từng thông số tới giá trị độ rỗng, từ đó người điều hành có định hướng xác định độ rỗng nhanh chóng nhằm đưa ra phương án khoan, khai thác và phát triển mỏ hợp lý.

Các dữ liệu thông số khoan được thu thập từ các cảm biến dữ liệu trong quá trình khoan qua các thành hệ thạch học khác nhau. Dữ liệu thông số khoan thường được thu thập là vận tốc cơ học khoan (ROP), lưu lượng bơm của dung dịch khoan (GPM), tốc độ vòng quay (RPM), áp suất trong cần dụng (SPP), mô-men xoắn (T) và tải trọng chiều trục (WOB). Dữ liệu này cần phải được xử lý trước để đảm bảo chất lượng dữ liệu sẽ cung cấp cho mô hình. Việc loại bỏ dữ liệu bị thiếu và dữ liệu ngoại lệ khỏi dữ liệu đã thu thập là điều rất nên làm và ảnh hưởng đến độ chính xác của mô hình máy học. Công tác thống kê dữ liệu cho dữ liệu sạch là một bước quan trọng để nghiên cứu phạm vi dữ liệu cho từng tham số bên cạnh các phân tích thống kê khác. Ngoài ra, việc nghiên cứu hệ số tương quan (R) giữa các dữ liệu cho thấy tầm quan trọng sự liên quan của từng thông số đối với đầu ra của mô hình.

Al-AbdulJabba và nnk.(2020) chỉ ra mối tương quan giữa thông số khoan và độ rỗng, được biểu diễn trên Hình 2. Mặc dù các số liệu không cho thấy mối tương quan cao giữa độ rỗng và một số dữ liệu khoan đầu vào, chẳng hạn như lưu lượng bơm của dung dịch khoan và mô-men xoắn, nhưng không có thông số nào trong số này có thể bỏ qua khỏi các thông số đầu vào. Lý do cho điều này là ROP được kết nối với năm dữ liệu khoan: WOB, T, GPM, SPP và RPM. Mối quan hệ giữa độ rỗng và các dữ liệu khoan xuất hiện khi hệ tầng có độ rỗng cao sẽ mang lại ROP cao hơn nhiều so với hệ tầng chặt, do chúng giống nhau về thạch học và tổng chiều sâu thẳng đứng (TVD). ROP tăng đột ngột có thể là do sự gia tăng độ rỗng của thành hệ, điều này không phải lúc nào cũng đúng vì cũng có thể do RPM hoặc WOB gây ra. Do đó, chỉ sử dụng ROP làm yếu tố dự đoán độ rỗng hình thành sẽ không tương quan hoàn toàn đến độ rỗng. Vì vậy, trên thực tế, có hai nhóm tham số: trực tiếp và gián tiếp. Các thông số liên quan trực tiếp đến độ rỗng



Hình 2. Hệ số tương quan giữa các thông số đầu vào và độ rỗng (Al-Abduljabbar và nnk., 2020).

là ROP, RPM và WOB. Các thông số liên quan gián tiếp đến độ rỗng là mô-men xoắn, SPP và GPM. Đối với các thông số trực tiếp, ROP thay đổi theo độ rỗng do thể tích vùng khoan qua. Đối với RPM, giá trị độ rỗng nhỏ thì độ mài mòn và độ cứng của thành hệ càng cao, do đó giá trị RPM cao sẽ gây ra rung động. Độ rỗng càng cao thì càng có thể áp dụng giá trị RPM cao để khoan và phá hủy đá mà không lãng phí năng lượng. Cuối cùng, cần giá trị WOB nhỏ hơn để phá hủy các vùng có độ rỗng cao hơn vì tồn tại ít vùng phá hủy hơn và ngược lại (Al-Abduljabbar và nnk., 2020).

Trong nghiên cứu của Al-Sabaa và nnk. (2021) cho thấy tốc độ vòng quay có hệ số tương quan nhỏ nhất với 0,22 và lưu lượng bơm của dung dịch khoan đạt giá trị lớn nhất là 0,441 khi so với độ rỗng của đá thành hệ, thông số đầu ra của mô hình. (Al-Sabaa và nnk., 2021).

Năm 2021, Gamal và nnk.(2021) chỉ ra mối quan hệ giữa độ rỗng của thành hệ (đầu ra của mô hình) và các dữ liệu khoan (đầu vào của mô hình) như Q, RPM, WOB, ROP và T có hệ số tương quan (R) thấp lần lượt là 0,299, 0,233, 0,151, 0,144 và 0,086. Trong khi đó, mối quan hệ giữa độ rỗng và SPP rất yếu với R là -0,003. Điều đó, cho thấy mối quan hệ giữa độ rỗng và các dữ liệu khoan có thể là mối quan hệ phi tuyến tính (Gamal và Elkatatny, 2021).

Gamal, Elkatatny và Mahmoud (2021) cũng chỉ ra hệ số tương quan (R) giữa dữ liệu thông số khoan và độ rỗng thể hiện mối quan hệ tuyến tính yếu khi giá trị R nằm trong khoảng từ - 0,07 đến 0,26 và do đó, có khả năng xảy ra loại mối quan hệ phi tuyến tính. Mặc dù R giữa các thông số khoan

đầu vào và độ rỗng là thấp, nhưng việc bỏ qua bất kỳ thông số đầu vào nào trong số này có thể làm giảm độ chính xác của các mô hình được tối ưu hóa trong dự báo giá trị độ rỗng và đây là lý do để xem xét tất cả các đầu vào này trong việc đào tạo các mô hình máy học (Gamal và nnk., 2021).

Trong công bố của Alhowaish và nnk.(2023) cho thấy tầm quan trọng của từng tham số đầu vào (thông số khoan) có thể được đo lường và xếp hạng bằng cách sử dụng phân tích tính năng hoán vị. Phân tích tầm quan trọng của tính năng hoán vị đã tiết lộ các biến đầu vào đóng góp nhiều nhất vào kết quả đầu ra. Các tham số đầu vào được xử lý trước khi lọc và phân cụm để nâng cao chất lượng mô hình. Tầm quan trọng của tính năng hoán vị cho thấy tốc độ vòng quay (RPM), mô-men xoắn (T) và áp suất trong cần dụng (SPP) đóng góp nhiều hơn vào dự đoán độ rỗng so với các tính năng đầu vào khác (Alhowaish và nnk., 2023).

## 6. Kết luận

Đánh giá tổng quan về việc áp dụng kỹ thuật máy học trong dự báo giá trị độ rỗng đã được trình bày. Việc dự báo độ rỗng theo thời gian thực bằng mô hình máy học dựa trên dữ liệu thông số khoan sẽ tiết kiệm chi phí và thời gian cho việc xác định độ rỗng trong thực tế bằng cách sử dụng các phép đo trong phòng thí nghiệm hoặc quá trình đo địa vật lý giếng khoan. Việc ứng dụng thành công mô hình máy học vào dự đoán thời gian thực về độ rỗng mang lại giải pháp thiết thực và hiệu quả cho ngành dầu khí.

Điểm yếu của mô hình được phát triển là dữ liệu phạm vi cho các tham số đầu vào. Do đó, mô



hình chỉ có thể sử dụng để dự đoán độ rỗng với độ chính xác cao cho tập dữ liệu mới có cùng dải dữ liệu với tham số đầu vào.

Các thông số khoan theo thời gian thực có mối quan hệ phi tuyến tính với độ rỗng. Để tăng độ chính xác của mô hình máy học nên sử dụng đồng thời các thông số khoan làm thông số đầu vào cho mô hình. Cần có thêm các nghiên cứu khác về sử dụng các mô hình khác trong dự báo độ rỗng của thành hệ theo thời gian thực.

### Lời cảm ơn

Tác giả xin chân thành cảm ơn nhóm nghiên cứu IPR Khoa dầu khí và năng lượng, Trường Đại học Mỏ - Địa chất.

### Đóng góp của tác giả

Lê Quang Duyệt - xây dựng ý tưởng nghiên cứu, tập hợp tài liệu, phân tích, đánh giá và viết bản thảo, chỉnh sửa bài báo.

### Tài liệu tham khảo

- Ahmadi, M. A., & Chen, Z. (2019). Comparison of machine learning methods for estimating permeability and porosity of oil reservoirs via petro-physical logs. *Petroleum*, 5(3), 271–284.
- Al-Abduljabbar, A., Al-Azani, K., & Elkatatny, S. (2020). Estimation of Reservoir Porosity From Drilling Parameters Using Artificial Neural Networks. *Petrophysics – The SPWLA Journal of Formation Evaluation and Reservoir Description*, 61(3), 318–330.
- Alhawaish, H. A., Mezghani, M. M., & Shakriov, A. (2023, October 2). *Predicting Porosity from Drilling Data Using Machine Learning – Challenges and Solutions*. ADIPEC.
- Al-Sabaa, A., Gamal, H., & Elkatatny, S. (2021, October 18). *Generation of a Complete Profile for Porosity Log While Drilling Complex Lithology by Employing the Artificial Intelligence*. SPE Symposium: Artificial Intelligence - Towards a Resilient and Efficient Energy Industry.
- Alyafei, N. (2021). *Fundamentals of Reservoir Rock Properties—2nd edition*. QScience.com. [https://doi.org/10.5339/Fundamentals\\_of\\_Reservoir\\_Rock\\_Properties\\_2ndEdition](https://doi.org/10.5339/Fundamentals_of_Reservoir_Rock_Properties_2ndEdition)

Andagoya Carrillo, K. I., Avellán, F. J., & Camacho, G. (2015, November 18). *ECD and Downhole Pressure Monitoring While Drilling at Ecuador Operations*. SPE Latin American and Caribbean Petroleum Engineering Conference.

Angeleri, G. P., & Carpi, R. (1982). Porosity Prediction from Seismic Data. *Geophysical Prospecting*, 30(5), 580–607.

Barjoui, H. S., Ghorbani, H., Mohamadian, N., Wood, D. A., Davoodi, S., Moghadasi, J., & Saberi, H. (2021). Prediction performance advantages of deep machine learning algorithms for two-phase flow rates through wellhead chokes. *Journal of Petroleum Exploration and Production*, 11(3), 1233–1261.

Bonnecaze, R. T., Sharma, M. M., Butler, J. E., & Arboleda, G. (2002, September 29). *High Resolution Downhole Measurements of Porosity and Fluid Saturation While Core Drilling*. SPE Annual Technical Conference and Exhibition.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32.

Chen, T., Zhu, L., Niu, R., Trinder, C. J., Peng, L., & Lei, T. (2020). Mapping landslide susceptibility at the Three Gorges Reservoir, China, using gradient boosting decision tree, random forest and information value models. *Journal of Mountain Science*, 17(3), 670–685.

Denney, D. (2003). A New Approach for Reservoir Characterization. *Journal of Petroleum Technology*, 55(09), 62–63.

Doyen, P. M. (1988). Porosity from seismic data: A geostatistical approach. *GEOPHYSICS*, 53(10), 1263–1275.

Elkatatny, S., Tariq, Z., Mahmoud, M., & Abdulraheem, A. (2018). New insights into porosity determination using artificial intelligence techniques for carbonate reservoirs. *Petroleum*, 4(4), 408–418.

Ellis, D. V., Case, C. R., & Chiaramonte, J. M. (2004). Porosity from Neutron Logs II: Interpretation. *Petrophysics - The SPWLA Journal of Formation Evaluation and Reservoir Description*, 45(01).

Fischetti, A. I., & Andrade, A. (2002). Porosity images from well logs. *Journal of Petroleum Science and Engineering*, 36(3), 149–158.

- Fu, B., Liu, M., He, H., Lan, F., He, X., Liu, L., Huang, L., Fan, D., Zhao, M., & Jia, Z. (2021). Comparison of optimized object-based RF-DT algorithm and SegNet algorithm for classifying Karst wetland vegetation communities using ultra-high spatial resolution UAV data. *International Journal of Applied Earth Observation and Geoinformation*, 104, 102553.
- Gamal, H., & Elkatatny, S. (2021). Prediction Model Based on an Artificial Neural Network for Rock Porosity. *Arabian Journal for Science and Engineering*, 47(9), 11211–11221.
- Gamal, H., Elkatatny, S., & Mahmoud, A. A. (2021). Machine learning models for generating the drilled porosity log for composite formations. *Arabian Journal of Geosciences*, 14(23), 2700.
- Ghorbani, H., Wood, D. A., Choubineh, A., Tatar, A., Abarghoyi, P. G., Madani, M., & Mohamadian, N. (2020). Prediction of oil flow rate through an orifice flow meter: Artificial intelligence alternatives compared. *Petroleum*, 6(4), 404–414.
- Gurina, E., Klyuchnikov, N., Zaytsev, A., Romanenkova, E., Antipova, K., Simon, I., Makarov, V., & Koroteev, D. (2020). Application of machine learning to accidents detection at directional drilling. *Journal of Petroleum Science and Engineering*, 184, 106519.
- Hamada, G. M., & Elshafei, M. A. (2009, May 9). *Neural Network Prediction of Porosity and Permeability of Heterogeneous Gas Sand Reservoirs*. SPE Saudi Arabia Section Technical Symposium.
- Hassaan, S., Mohamed, A., Ibrahim, A. F., & Elkatatny, S. (2024). Real-Time Prediction of Petrophysical Properties Using Machine Learning Based on Drilling Parameters. *ACS Omega*, acsomega.3c08795.
- Hinton, G. E., Osindero, S., & Teh, Y.-W. (2006). A Fast Learning Algorithm for Deep Belief Nets. *Neural Computation*, 18(7), 1527–1554.
- Kane, J. A., & Jennings, J. W. (2005, October 9). *A Method to Normalize Log Data by Calibration to Large-Scale Data Trends*. SPE Annual Technical Conference and Exhibition.
- Lippmann, R. (1987). An introduction to computing with neural nets. *IEEE ASSP Magazine*, 4(2), 4–22.
- Lucia, F. J., Kerans, C., & Jennings, J. W., Jr. (2003). Carbonate Reservoir Characterization. *Journal of Petroleum Technology*, 55(06), 70–72.
- Ma, Y., & Guo, G. (2014). *Support Vector Machines Applications*. Springer Science & Business Media.
- Nakamoto, P. (2018). *Neural Networks and Deep Learning: Neural Networks and Deep Learning, Deep Learning Explained to Your Granny*. CreateSpace Independent Publishing Platform.
- Nallathambi, S., & Ramasamy, K. (2017). Prediction of electricity consumption based on DT and RF: An application on USA country power consumption. *2017 IEEE International Conference on Electrical, Instrumentation and Communication Engineering (ICEICE)*, 1–7.
- Nguyễn, T. M. H., và Lê, H. A. (2014). Dự báo độ rỗng trầm tích Miocen khu vực lô 103. *Tạp chí Khoa học Kỹ thuật Mỏ - Địa chất*.
- Noshi, C. I., & Schubert, J. J. (2018, October 7). *The Role of Machine Learning in Drilling Operations: A Review*. SPE/AAPG Eastern Regional Meeting.
- Nyein, C. Y., & Ali Hamada, G. M. M. (2018). Artificial Neural Network (ANN) Prediction of Porosity and Water Saturation of Shaly Sandstone Reservoirs. *2018 AAPG/EAGE/MGS Myanmar Oil and Gas Conference: A Global Oil and Gas Hotspot: Unleashing the Petroleum Systems Potential*. 2018 AAPG/EAGE/MGS Myanmar Oil and Gas Conference: A Global Oil and Gas Hotspot: Unleashing the Petroleum Systems Potential, Yangon, Myanmar.
- Olatunji, S. O., Selamat, A., & Abdulraheem, A. (2011). Modeling the permeability of carbonate reservoir using type-2 fuzzy logic systems. *Computers in Industry*, 62(2), 147–163.
- Patle, A., & Chouhan, D. S. (2013). SVM kernel functions for classification. *2013 International Conference on Advances in Technology and Engineering (ICATE)*, 1–9.

- Ramana, Y. V., & Venkatanarayana, B. (1971). An air porosimeter for the porosity of rocks. *International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts*, 8(1), 29–53.
- Sircar, A., Yadav, K., Rayavarapu, K., Bist, N., & Oza, H. (2021). Application of machine learning and artificial intelligence in oil and gas industry. *Petroleum Research*, 6(4), 379–391.
- Sun, J., Zhang, R., Chen, M., Chen, B., Wang, X., Li, Q., & Ren, L. (2021). Identification of Porosity and Permeability While Drilling Based on Machine Learning. *Arabian Journal for Science and Engineering*, 46(7), 7031–7045.
- Wood, D. A. (2020). Predicting porosity, permeability and water saturation applying an optimized nearest-neighbour, machine-learning and data-mining network of well-log data. *Journal of Petroleum Science and Engineering*, 184, 106587.
- Yazmyradova, G., Hermana, M., & Soleimani, H. (2022). Estimation of porosity from well logs and seismic using artificial neural network. *IOP Conference Series: Earth and Environmental Science*, 1003(1), 012017.
- Zerrouki, A. A., Aïfa, T., & Baddari, K. (2014). Prediction of natural fracture porosity from well log data by means of fuzzy ranking and an artificial neural network in Hassi Messaoud oil field, Algeria. *Journal of Petroleum Science and Engineering*, 115, 78–89.