PETROVIETNAM Số 3 - 2025

DầuKhí

KHOA HỌC, CÔNG NGHỆ VÀ ĐỔI MỚI SÁNG TẠO

PETROVIETNAM

ĐẶC SAN CỦA TẬP ĐOÀN CÔNG NGHIỆP - NĂNG LƯỢNG QUỐC GIA VIỆT NAM ■ SỐ 3/2025

ISSN 3030-4075

TRƯỞNG BAN BIÊN TẬP
TS. Lê Xuân Huyên

PHÓ TRƯỞNG BAN BIÊN TẬP
TS. Lê Mạnh Hùng
ThS. Lê Ngọc Sơn

BAN BIÊN TẬP
TS. Trịnh Xuân Cường
TS. Nguyễn Anh Đức
ThS. Vũ Đào Minh
ThS. Trần Thái Ninh
ThS. Dương Mạnh Sơn
PGS.TS. Lê Văn Sỹ
ThS. Bùi Minh Tiến
ThS. Nguyễn Văn Tuấn

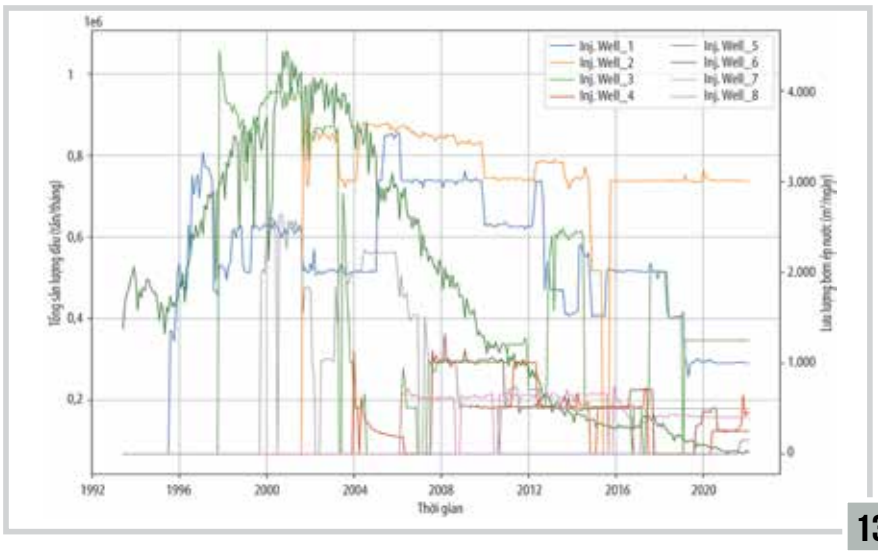
BAN TRỊ SỰ
ThS. Lê Văn Khoa
ThS. Nguyễn Thị Việt Hà
ThS. Nguyễn Trung Đạt

TỔ CHỨC THỰC HIỆN
Viện Dầu khí Việt Nam

BAN TRỊ SỰ
Tầng 16, Tòa nhà Viện Dầu khí Việt Nam - 167 Phố Trung Kính, Phường Yên Hòa, Thành phố Hà Nội
*Tel: 024-37727108 | 0982288671 * Fax: 024-37844156 * Email: tcdk@pvn.vn*

DầuKhí
KHOA HỌC, CÔNG NGHỆ VÀ ĐỔI MỚI SÁNG TẠO PETROVIETNAM

Giấy phép xuất bản số 153/GP - XBĐS cấp ngày 12/12/2024 của Cục Báo chí, Bộ Văn hóa, Thể thao và Du lịch



13

4. Ứng dụng AI dự báo khoảng mở vỉa trong điều kiện dữ liệu địa vật lý giếng khoan hạn chế tại mỏ Tê Giác Trắng

13. Sử dụng thuật toán mạng neural nhân tạo trong tối ưu bơm ép nước cho mỏ Bạch Hổ

22. Nghiên cứu xây dựng mô hình phân chia sản phẩm khai thác cho các mỏ kết nối tại Vietsovpetro trên nền tảng điện toán đám mây

32. Khai thác dữ liệu chuyên ngành với sự hỗ trợ của AI: Kinh nghiệm từ phần mềm quản lý cơ sở dữ liệu đường cong địa vật lý giếng khoan bể Cửu Long

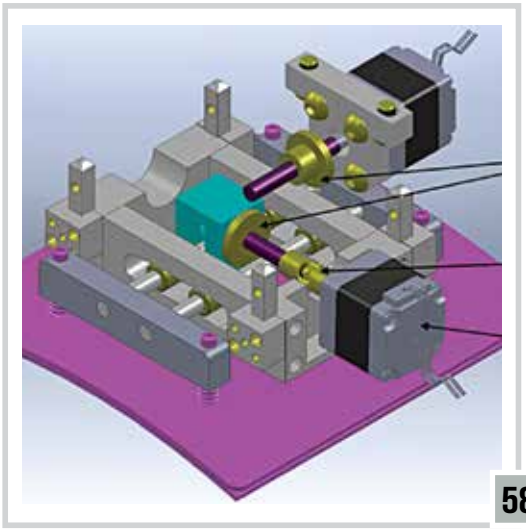
41. Ứng dụng Python và machine learning trong xử lý và phân loại dữ liệu xuất nhập khẩu

51. Ứng dụng AI trong tổng hợp thông tin thị trường và phân tích các chỉ số kinh tế vĩ mô

58. Nghiên cứu chế tạo thiết bị cầm tay phát hiện khuyết tật của hệ thống đường ống bằng phương pháp rò rỉ đường sức từ (MFL)

66. Nghiên cứu chế tạo robot phát hiện ăn mòn bồn nổi hình trụ đứng chứa xăng dầu

76. Kết hợp sử dụng ảnh viễn thám quang học và radar trong phân loại và giám sát vết dầu trên biển



LỜI GIỚI THIỆU

Chuyển đổi số đang tạo ra giá trị đáng kể cho ngành công nghiệp năng lượng. Nghiên cứu của McKinsey cho thấy khi ứng dụng công nghệ số thành công, các doanh nghiệp năng lượng có thể tăng sản lượng và hiệu suất từ 2 - 10%, đồng thời giảm chi phí vận hành từ 10 - 30%. Thực hiện Nghị quyết số 57-NQ/TW ngày 22/12/2024 của Bộ Chính trị về đột phá phát triển khoa học, công nghệ, đổi mới sáng tạo và chuyển đổi số quốc gia, Petrovietnam xác định ứng dụng trí tuệ nhân tạo, tích hợp với công nghệ tự động hóa vào hoạt động quản trị, sản xuất kinh doanh; với mục tiêu khoa học công nghệ, đổi mới sáng tạo và chuyển đổi số sẽ tạo ra 25% doanh thu và hiệu quả gấp 2 lần so với giá trị đầu tư.

Trong số này, Đặc san Dầu khí - Khoa học, công nghệ và đổi mới sáng tạo giới thiệu các nghiên cứu về ứng dụng trí tuệ nhân tạo (AI) và machine learning (ML) trong lĩnh vực thăm dò, khai thác dầu khí. Kết quả "ứng dụng AI dự báo khoảng mở vỉa trong điều kiện dữ liệu địa vật lý giếng khoan hạn chế tại mỏ Tê Giác Trắng" giúp xác định chính xác các tầng chứa dầu tiềm năng, góp phần gia tăng đáng kể sản lượng khai thác với chi phí tối ưu. Việc "sử dụng thuật toán mạng neural nhân tạo trong tối ưu bơm ép nước cho mỏ Bạch Hổ" đã giúp sản lượng dầu khai thác tăng trung bình 1,5% khi sử dụng các lược đồ bơm ép nước được tối ưu hóa. Đặc biệt, nghiên cứu "Xây dựng mô hình phân chia sản phẩm khai thác cho các mỏ kết nối tại Vietsovpetro trên nền tảng điện toán đám mây" đã giúp tối ưu hóa quá trình vận hành sản xuất, nâng cao độ chính xác và hiệu quả trong công tác phân chia sản phẩm cho các mỏ kết nối.

Cũng trong số này, Đặc san Dầu khí - Khoa học, công nghệ và đổi mới sáng tạo giới thiệu các nghiên cứu về "Khai thác dữ liệu chuyên ngành với sự hỗ trợ của AI: Kinh nghiệm từ phần mềm quản lý cơ sở dữ liệu đường cong địa vật lý giếng khoan bể Cửu Long" và "Ứng dụng Python và machine learning trong xử lý và phân loại dữ liệu xuất nhập khẩu", thể hiện xu hướng tích hợp AI vào quản lý dữ liệu lớn (big data). Bên cạnh đó, nghiên cứu "Ứng dụng AI trong tổng hợp thông tin và phân tích các chỉ số kinh tế vĩ mô" góp phần nâng cao năng lực dự báo, hoạch định chiến lược kinh doanh trong bối cảnh thị trường năng lượng thế giới biến động phức tạp.

Để góp phần hiện thực hóa mục tiêu xây dựng nhà máy/giàn khoan/công trình thông minh, Đặc san Dầu khí - Khoa học, công nghệ và đổi mới sáng tạo giới thiệu các nghiên cứu chế tạo robot phát hiện ăn mòn bồn nổi hình trụ đứng chứa xăng dầu; chế tạo thiết bị cầm tay phát hiện khuyết tật của hệ thống đường ống bằng phương pháp rò rỉ đường sức từ (MFL); sử dụng ảnh viễn thám quang học và radar trong phân loại và giám sát vết dầu trên biển.

Ban Biên tập hy vọng từ góc nhìn của các chuyên gia, nhà khoa học cùng các giải pháp về công nghệ sẽ góp phần hiện thực hóa mục tiêu xây dựng Petrovietnam có tiềm lực mạnh, quản trị hiện đại bằng công nghệ số theo chiến lược "AI First", bảo đảm an ninh năng lượng quốc gia đến năm 2030, tầm nhìn đến năm 2045.

TRƯỞNG BAN BIÊN TẬP
Phó Tổng giám đốc
Tập đoàn Công nghiệp - Năng lượng Quốc gia Việt Nam

TS. Lê Xuân Huyên

DầuKhí
KHOA HỌC, CÔNG NGHỆ VÀ ĐỔI MỚI SÁNG TẠO PETROVIETNAM

SỬ DỤNG THUẬT TOÁN MẠNG NEURAL NHÂN TẠO TRONG TỐI ƯU BƠM ÉP NƯỚC CHO MỎ BẠCH HỔ

Đoàn Huy Hiên[1], Lê Thế Hùng[1], Trần Xuân Quý[1], Phạm Trường Giang[1], Nguyễn Minh Quý[1], Nguyễn Thế Đức[2]
[1]Viện Dầu khí Việt Nam (VPI)
[2]Viện Cơ học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam (VAST)
Email: hiendh.epc@vpi.pvn.vn
https://doi.org/10.47800/PVSI.2025.03-02

Tóm tắt

Sản lượng dầu ngoài khơi Việt Nam được khai thác chủ yếu từ vỉa chứa móng Bạch Hổ, nơi có chế độ dòng chảy rất phức tạp do sự phân bố không gian bất đồng nhất của các thuộc tính vật lý thạch học như độ rỗng, độ thấm và độ bão hòa nước. Do đó, phương pháp dự báo sản lượng dầu truyền thống dựa trên mô phỏng vỉa chứa thường thiếu chính xác hoặc đòi hỏi nhiều công sức và thời gian để tối ưu hóa các thông số thủy động lực học. Sự phát triển gần đây của các thuật toán học máy (machine learning - ML) sẽ giúp dự đoán được sản lượng dầu từ lưu lượng bơm ép nước của từng giếng bơm nhanh hơn và đáng tin cậy hơn. Một khi có thể dự đoán sản lượng dầu bằng phương pháp học máy, việc tối ưu hóa quá trình bơm ép nước có thể được triển khai bằng nhiều thuật toán tối ưu hóa khác nhau.

Trong nghiên cứu này, sử dụng thuật toán mạng neural nhân tạo (ANN) đã cho kết quả tốt nhất: hệ số tương quan giữa các giá trị dự đoán và giá trị thực tế lần lượt là 0,98 và 0,95 đối với tập dữ liệu huấn luyện và kiểm tra. Sau đó, thuật toán tối ưu Gauss-Newton được áp dụng để tìm ra lưu lượng bơm ép nước tối ưu nhất cho từng giếng bơm nhằm tăng sản lượng dầu. Kết quả cho thấy sản lượng dầu khai thác tăng trung bình 1,5% khi sử dụng các lược đồ bơm ép mới được tối ưu hóa.

Từ khóa: Tối ưu hóa bơm ép nước, mô phỏng vỉa chứa, mạng neural nhân tạo, học máy, mỏ Bạch Hổ.

1. Giới thiệu

Kể từ khi được đưa vào khai thác tới nay, mỏ Bạch Hổ (Lô 09-1, bể Cửu Long, ngoài khơi Việt Nam) đã có đóng góp lớn nhất vào tổng sản lượng dầu của Việt Nam. Dòng dầu đầu tiên của mỏ Bạch Hổ được khai thác từ vỉa chứa Miocene vào năm 1986, sau đó lần lượt từ vỉa chứa Oligocene và móng vào năm 1987 và 1988. Sản lượng dầu tích lũy được khai thác từ đá móng bể Cửu Long chiếm hơn 95% tổng sản lượng dầu của Việt Nam, trong đó hơn 90% từ mỏ Bạch Hổ - con số này bắt đầu suy giảm từ năm 2005. Từ năm 1996, phương pháp bơm ép nước (waterflooding) đã được triển khai tại mỏ Bạch Hổ để duy trì áp suất vỉa chứa.

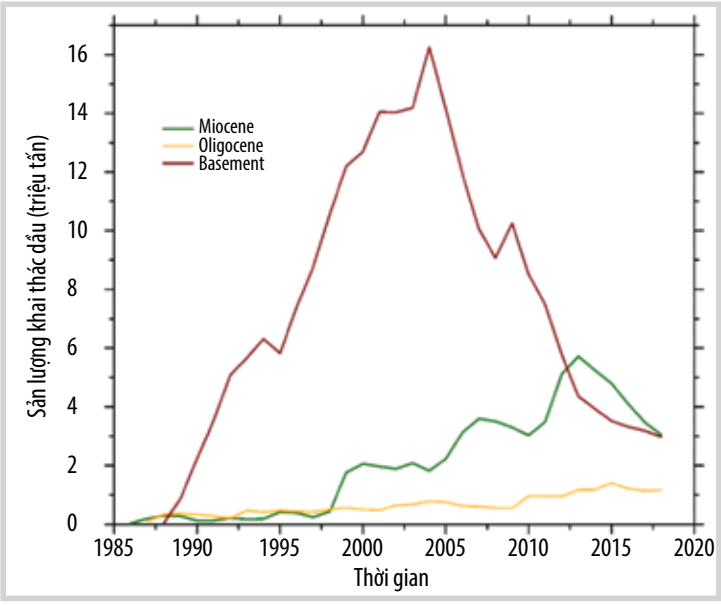
Mục tiêu chính của bơm ép nước là điều chỉnh các biến điều khiển (manipulated variables) sao cho tổng sản lượng dầu được tối đa hóa. Do đó, lượng dầu khai thác phải được dự đoán dựa trên tổng lượng nước bơm ép, việc này thường được thực hiện thông qua mô phỏng vỉa chứa động lực học. Quy trình bơm ép nước điển hình gồm 2 bước. Bước thứ nhất là dự đoán sản lượng tương lai dựa trên kết quả khớp lịch sử (history matching) từ mô phỏng vỉa chứa với lược đồ bơm ép nước hiện tại. Bước thứ hai là tối ưu hóa việc kiểm soát giếng (well control), gồm tối ưu hóa lưu lượng bơm và lắp đặt giếng bơm mới, bằng cách tối đa hóa lưu lượng dầu (Andrea Capolei và nnk, 2013 [1], Bjarne Foss và nnk, 2015 [2], Van den Hof và nnk, 2012 [3], Bjarne Foss và nnk, 2011 [4], J.D. Jansen và nnk, 2014 [5]). Phương pháp này xem xét tất cả ảnh hưởng của các quá trình vật lý và cho phép mô phỏng quá trình khai thác thực tế nhất có thể bằng cách giải các phương trình vi phân từng phần (partial differential equations). Độ chính xác của mô hình phụ thuộc vào mức độ phức tạp của các điều kiện địa chất, được thể hiện qua sự phân bố không gian của các thuộc tính vật lý thạch học như độ thấm, độ bão hòa nước và độ rỗng, cũng như năng lực của kỹ sư vỉa chứa trong việc thu thập dữ liệu để xây dựng và xác thực một mô hình vỉa chứa đủ chính xác. Đối với mô hình mô phỏng vỉa chứa quy mô lớn hoặc phức tạp, một lần chạy

Ngày nhận bài: 13/6/2025
Ngày đánh giá và sửa chữa: 13 - 22/6/2025
Ngày duyệt đăng: 22/6/2025

mô phỏng thuận (forward simulation run) có thể mất từ vài giờ đến vài ngày để hoàn thành (Tailai Wen và nnk, 2014 [6]; Zhenyu Guo và nnk, 2018 [7]).

Thay vì mô phỏng vỉa chứa kiểu vật lý, các kỹ thuật mô phỏng không dựa trên vỉa chứa, như mô phỏng đường dòng (streamline simulation) (Khalid Aziz và nnk, 1979 [8], Akhil Datta-Gupta và nnk, 2007 [9], M.R. Thiele và nnk, 2010 [10]), mô hình điện dung - điện trở (capacitance resistance model - CRM) (Nguyen Anh Phuong, 2012 [11], Daniel Brent Weber và nnk, 2009 [12], Ali A. Yousef và nnk, 2006 [13], L.W. Lake và nnk, 2007 [14]) và mô hình kết nối giữa các giếng (interwell interconnection model) (Hui Zhao và nnk, 2016 [15], Zhenyu Guo và nnk, 2018 [7]) có thể là các phương pháp thay thế nhằm tối ưu hóa bơm ép nước. Những phương pháp này sẽ ước tính tác động của cặp giếng bơm ép và giếng khai thác thông qua việc đảo ngược các hệ số truyền dẫn (transmissibility coefficients) sao cho khớp với sản lượng thực tế tại giếng khai thác. Việc triển khai các phương pháp trên sẽ đẩy nhanh quá trình mô phỏng bằng cách đơn giản hóa dòng chảy, nhưng việc tinh chỉnh các thông số mô hình, như tính thấm tương đối, độ bão hòa nước ban đầu và tổng thể tích lỗ rỗng, đòi hỏi phải có các chuyên gia giàu kinh nghiệm để có được kết quả tin cậy.

Việc áp dụng các phương pháp tiêu chuẩn trong giai đoạn tối ưu hóa bơm ép nước đòi hỏi rất nhiều nỗ lực. Nhiều trường hợp không có đủ dữ liệu để xây dựng các mô hình phức tạp dựa trên vật lý, đặc biệt đối với vỉa chứa móng. Các đặc trưng vật lý thạch học như độ rỗng, độ thấm và độ bão hòa nước cục bộ phụ thuộc đáng kể vào mạng lưới nứt nẻ (fracture network) - vốn rất khó xác định từ dữ liệu địa chấn truyền thống. Do đó, để có được một mô hình chấp nhận được cần phải đưa ra nhiều giả định.



Hình 1. *Sản lượng khai thác của bể Cửu Long theo thời gian.*

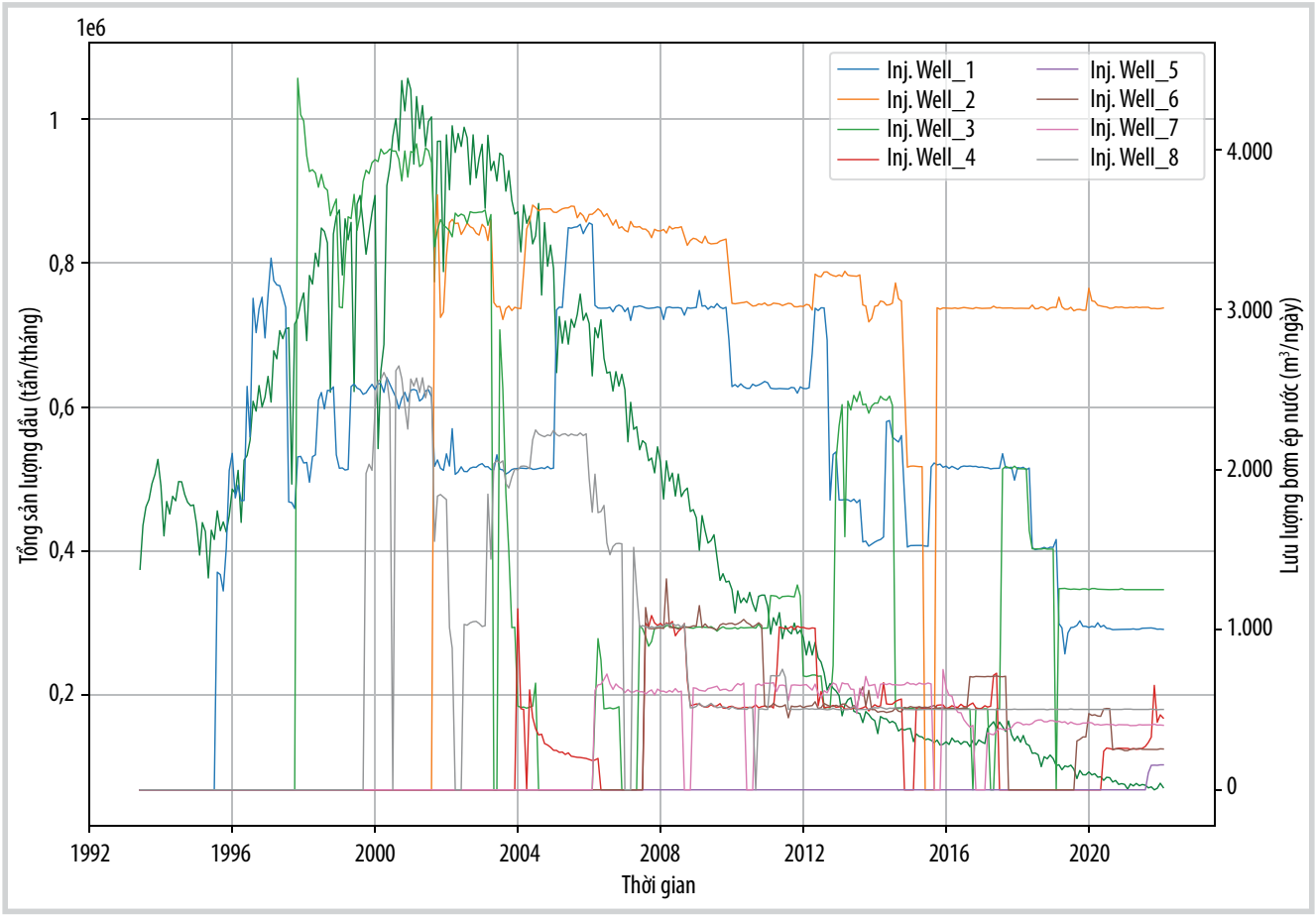
Gần đây, với sự phát triển của phương pháp học máy (ML), nhiều nghiên cứu đã áp dụng ML để dự đoán sản lượng dầu, tối ưu hóa lược đồ bơm ép nước nhằm nâng cao hiệu quả khai thác. Rui Zhang và nnk, 2021 [16] đã sử dụng chuỗi thời gian đa biến (multivariate time series) và phương pháp học máy tự hồi quy vector (vector autoregressive ML) để dự báo sản lượng dầu tại các vỉa chứa bơm ép nước. Trong nghiên cứu, các tác giả đã xác định các giếng bơm có ảnh hưởng đến giếng khai thác bằng cách tính hệ số tương quan đơn giản giữa lưu lượng bơm ép và lưu lượng khai thác dầu. Một số phương pháp ML khác như học tăng cường (reinforcement learning), mạng bộ nhớ dài hạn - ngắn hạn (long short-term memory network - LSTM network) đã được đề xuất để tối ưu hóa bơm ép nước (Farzad Hourfar và nnk, 2019 [17], Cuthbert Shang Wui Ng và nnk, 2022 [18]).

Trong nghiên cứu này, nhóm tác giả áp dụng quy trình làm việc (workflow) của Alfarizi và những người khác (Muhammad Gibran Alfarizi và nnk, 2022 [19]) kết hợp mạng neural nhân tạo (ANN) để dự đoán lưu lượng khai thác dầu và thuật toán tối ưu hóa Gauss-Newton nhằm tối ưu hóa quá trình bơm ép nước thông qua tối đa hóa giá trị hiện tại ròng (net present value - NPV).

Nhóm nghiên cứu đã thực hiện: i) thu thập số liệu về sản lượng dầu của các giếng khai thác cũng như lưu lượng bơm ép nước của các giếng bơm ép tại mỏ Bạch Hổ; ii) phân tích thống kê trên dữ liệu có sẵn để chọn lọc đặc trưng (feature selection); iii) tổng quan ngắn gọn về các phương pháp tiếp cận ANN cho bài toán hồi quy (regression problem) để dự đoán lưu lượng khai thác dầu của mỏ dầu đã cho từ lưu lượng nước bơm ép của các giếng bơm hiện tại; iv) đề xuất lược đồ bơm ép nước tối ưu nhằm nâng cao hiệu quả khai thác dầu bằng thuật toán tối ưu hóa Gauss-Newton.

2. Sản lượng dầu móng Bạch Hổ

Vỉa chứa móng nứt nẻ (fractured basement reservoir) được coi là một trường hợp đặc biệt trong ngành công nghiệp dầu khí. Do chế độ dòng chảy phức tạp, nhiều nghiên cứu đã được thực hiện để cải thiện việc khớp lịch sử sản lượng dầu (oil production history matching). Ví dụ, trong nghiên cứu của Lê Ngọc Sơn và nnk [20, 21], các thông số vật lý thạch học (petrophysical) và vỉa



Hình 2. *Lưu lượng bơm ép và lưu lượng khai thác dầu tại khu vực móng trung tâm mỏ Bạch Hổ.*

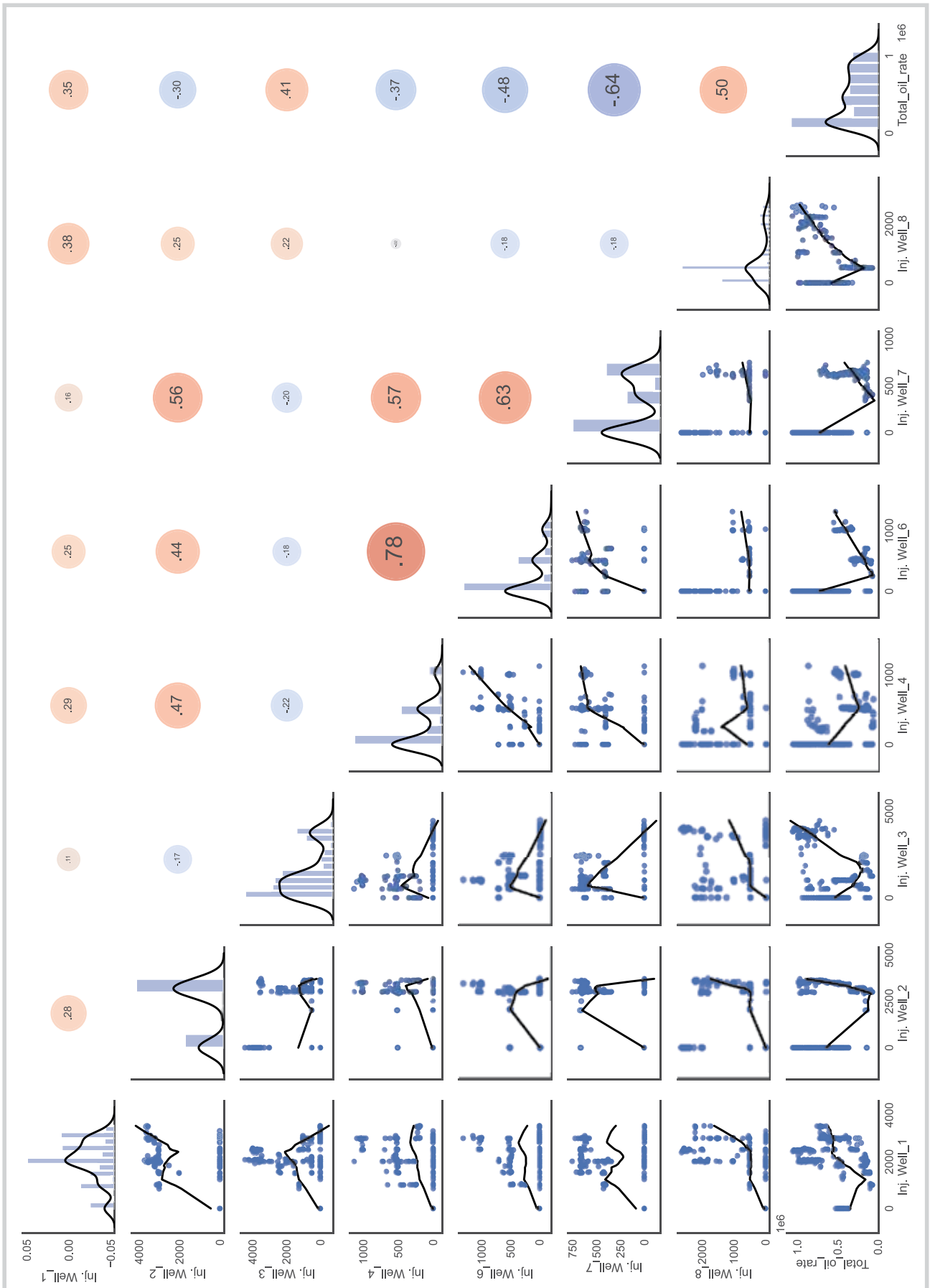
chứa như độ rỗng, độ thấm các thông số nén ép đá và kích thước tầng chứa nước (aquifer size) đã được điều chỉnh bằng cách sử dụng thuật toán tối ưu hóa trong quá trình mô phỏng nhằm đạt được kết quả khớp lịch sử sản lượng dầu tốt hơn. Một số phương trình thực nghiệm đã được đề xuất cho các mục đích này.

Tuy nhiên, các phương pháp học máy, vốn tiếp cận mục đích dự báo sản lượng dầu theo một cách khác, gần đây đã được sử dụng và ghi nhận nhiều thành công từ việc xem xét tài liệu. Chỉ dữ liệu thực tế về tổng sản lượng dầu từ các giếng khai thác và tốc độ bơm ép nước từ các giếng bơm mới được xem xét cho việc khớp và dự đoán tiếp theo.

Tại vỉa chứa móng Bạch Hổ, có gần 50 giếng bơm ép nước, tuy nhiên tại thời điểm nghiên cứu chỉ có 8 giếng đang hoạt động với lưu lượng bơm ép được thể hiện trong Hình 2 (trục dọc bên phải), cùng với tổng sản lượng dầu (trục dọc bên trái). Trong Hình 2, lưu lượng bơm ép tối đa khoảng 3.000 m^3/ngày tại giếng bơm ép số 2, trong khi giếng bơm ép số 5 mới được kích hoạt lại với lưu lượng bơm ép trung bình 154,3 m^3/ngày và sẽ không được xem xét trong tối ưu hóa bơm ép nước cũng như dự đoán tổng sản lượng dầu. Do đó, chỉ sử dụng 7 giếng bơm ép nước làm các đặc trưng đầu vào (input features) trong bài toán học máy.

Hình 3 cho thấy các biểu đồ cặp (pair plots) giữa các giếng bơm và tổng sản lượng dầu, sự phân bố và hệ số tương quan Pearson giữa các cặp. Hệ số tương quan Pearson chỉ ra rằng 4 giếng bơm có tương quan nghịch (negatively correlated) với tổng sản lượng dầu, trong khi 3 giếng còn lại có tương quan thuận (positively correlated). Điều này về mặt thống kê có nghĩa là có thể sơ bộ giảm lưu lượng bơm từ 4 giếng bơm có tương quan nghịch và tăng tốc độ bơm từ 3 giếng tương quan thuận để tăng tổng sản lượng dầu (total oil productivity).

Ngoài ra, các hệ số tương quan không đủ cao để cho phép sử dụng tất cả lưu lượng bơm ép nước từ các giếng bơm ép này làm đặc trưng đầu vào cho dự đoán tổng sản lượng của mỏ dầu. Sau khi được chọn, các đặc trưng dữ liệu sẽ được xáo trộn (shuffled) và chọn ngẫu nhiên 80% cho huấn luyện (training) và 20% cho kiểm tra (testing).



***Hình 3.** Hệ số tương quan giữa lưu lượng bơm ép và lưu lượng khai thác dầu tại khu vực trung tâm mỏ Bạch Hổ.*

3. Thuật toán ANN sử dụng cho dự báo khai thác

Trong nghiên cứu này, mô hình hồi quy (regression model) sẽ được sử dụng để ánh xạ lưu lượng bơm ép nước hiện tại từ các giếng bơm ép sang tổng lưu lượng dầu.

Theo thông lệ, $X = \{x_i\}_{i=1}^{d} \in R^{n \times d}$ biểu thị tập huấn luyện (training set), trong đó d là số lượng giếng bơm ép và n là số bước thời gian trong lịch sử bơm ép. Cột giá trị mục tiêu (target values) được biểu diễn dưới dạng vector $Y = \{y_i\}_{i=1}^{n} \in R^{n \times 1}$. Nói chung, bài toán đặt ra là cần tìm một hàm xấp xỉ $\hat{f}(x, \theta) : X \to Y$ bằng cách tối thiểu hóa hàm sai số (loss function) $\sum_{i=1}^{n} L\left(\hat{f}(x, \theta), y_i\right) \to min\, \theta$, trong đó $\hat{f}(x, \theta)$ là mô hình hồi quy với tham số mô hình θ.

- Mô hình hồi quy tuyến tính

Mô hình hồi quy tuyến tính được định nghĩa là $\hat{f}(x, w) = w^T x + w_0$. Để huấn luyện mô hình này, việc tối thiểu hóa hàm sai số đối với các hệ số $w^T \in R^{n \times 1}, w_0 \in R$ sẽ được thực hiện để tìm ra tham số mô hình tốt nhất, được biểu diễn như sau:

$$[w^T, w_0] = \underset{w^T, w_0}{argmin} \sum_{i=1}^{n} (w^T x_i + w_0 - y_i)^2 + R(w, \alpha) \quad (1)$$

Số hạng điều chuẩn (regularization term) R(w, α) sẽ loại bỏ các hệ số có giá trị cao nhằm tránh hiện tượng khớp quá mức (overfitting) và có thể giúp giảm thiểu hệ số của các đặc trưng có ảnh hưởng nhỏ đến biến mục tiêu. Có một số loại điều chuẩn:

$Lasso\ R(w, \alpha) = \alpha \sum_{j=1}^{d} |w_j|$ *(L₁ regularization);*

$Ridge\ R(w, \alpha) = \alpha \sum_{j=1}^{d} w_j^2$ *(L₂ regularization);*

Siêu tham số (hyperparameter) α liên quan đến điều chuẩn L_1/L_2 sẽ được tinh chỉnh bằng cách sử dụng các kỹ thuật điều chuẩn Lasso và Ridge của thư viện Scikit-learn cho thử nghiệm (Fabian Pedregosa và nnk, 2011) [22].

- Mạng neural nhân tạo (ANN)

Mô hình ANN có thể được biểu diễn dưới dạng:

$$\hat{f}(x) = \delta_k(w_k \delta_{k-1}(\dots w_2 \delta_1(w_1 x + b_1) + b_2) \dots) + b_k) \quad (2)$$

Trong đó:

δ_i: Hàm kích hoạt (activation function);

k: Số lớp;

$w_i \in R^{out_i \times in_i}$: Ma trận trọng số (weight matrix);

b_i: Độ lệch (bias) cho lớp thứ i.

Tương tự mô hình hồi quy tuyến tính, việc tối thiểu hóa hàm sai số bình phương (squared loss function) giữa giá trị tính toán và giá trị thực sẽ được thực hiện để khớp mô hình bằng cách sử dụng thuật toán giảm độ dốc ngẫu nhiên (stochastic gradient descent) hoặc phiên bản cải tiến có tên là Adam (Diederik P. Kingma và nnk, 2014 [23]) như sau:

$$[w_1, b_1, \dots, w_k, b_k] = \underset{w_1, b_1, \dots, w_k, b_k}{argmin} \frac{1}{n} \sum_{i=1}^{n} \left(\hat{f}(x_i) - y_i\right)^2 \quad (3)$$

- Đánh giá mô hình

Để ước tính mức độ chính xác mà các mô hình dự đoán hoạt động trên toàn bộ tập dữ liệu, nhóm tác giả đã sử dụng sai số tuyệt đối trung bình (MAE) (mean absolute error) làm phương pháp chấm điểm cho kiểm chứng chéo 5 lần (five-folds cross-validation). Chỉ số lỗi này được định nghĩa như sau:

$$MAE\left(\hat{f}(x), y\right) = \frac{1}{n} \sum_{i=1}^{n} \left|\hat{f}(x_i) - y_i\right| \quad (4)$$

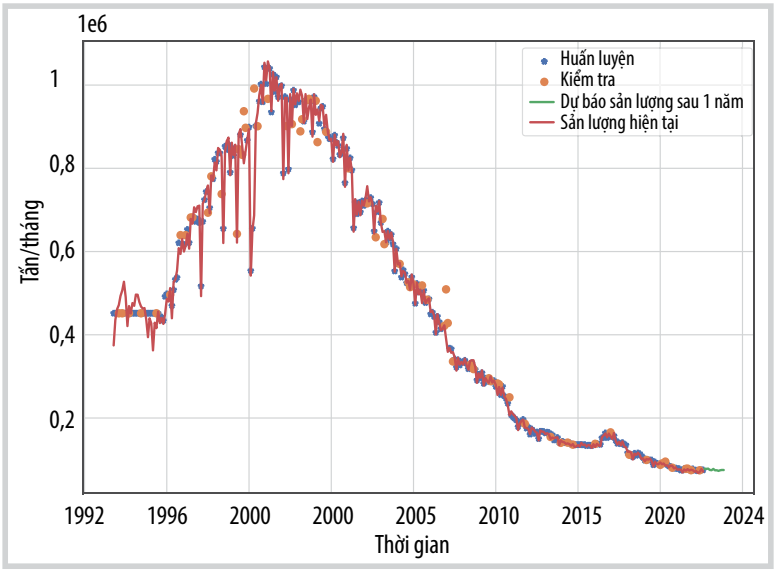
Để đánh giá khả năng áp dụng của mô hình trong thực tế và so sánh với đánh giá của nhà điều hành về hiệu quả của việc bơm ép, nhóm tác giả đã tiến hành thử nghiệm tính toán như sau:

- Chia mẫu dữ liệu thành tập huấn luyện (80%) và tập kiểm tra (20%) và huấn luyện một số mô hình đã đề cập ở trên với các tham số mặc định cho việc lựa chọn mô hình;

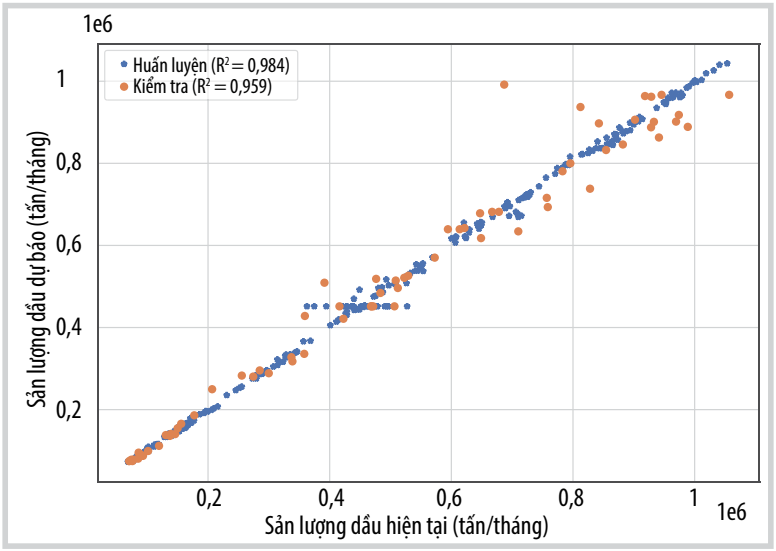
- Tinh chỉnh để tối ưu hóa các tham số mô hình trên tập dữ liệu huấn luyện. Sau đó, mô hình sẽ được áp dụng cho tập dữ liệu kiểm tra để xác minh việc khớp quá mức (verifying the overfit) trước khi được quyết định lựa chọn.

Kết quả dự đoán ANN trên tập huấn luyện và kiểm tra của dữ liệu sau khi xáo trộn và được chọn ngẫu nhiên được thể hiện trong Hình 4a. Tại đây, dự đoán trước 1 năm được thực hiện bằng cách giữ nguyên lược đồ bơm ép của năm hiện tại, sau đó nhóm tác giả áp dụng mô hình ANN để dự đoán xa hơn và hiển thị dưới dạng đường cong màu xanh lam trong Hình 4a. Dự đoán này sẽ được sử dụng làm trường hợp cơ sở (based case) cho việc tối ưu hóa lược đồ bơm ép nước tiếp theo.

Các chỉ số MAE (sai số tuyệt đối trung bình) là 37.644,24 và 61.160,15 tương ứng cho tập dữ liệu huấn luyện và tập dữ liệu kiểm tra, trong khi hệ số tương quan (coefficients) của mô hình ML giữa các giá trị đầu ra của tập huấn luyện và kiểm tra với sản lượng dầu thực tế lần lượt là 0,985 và 0,959, như thể hiện trên Hình 4b. Hệ số xác định cao của mô hình ANN và sự khác biệt không lớn giữa tập dữ liệu

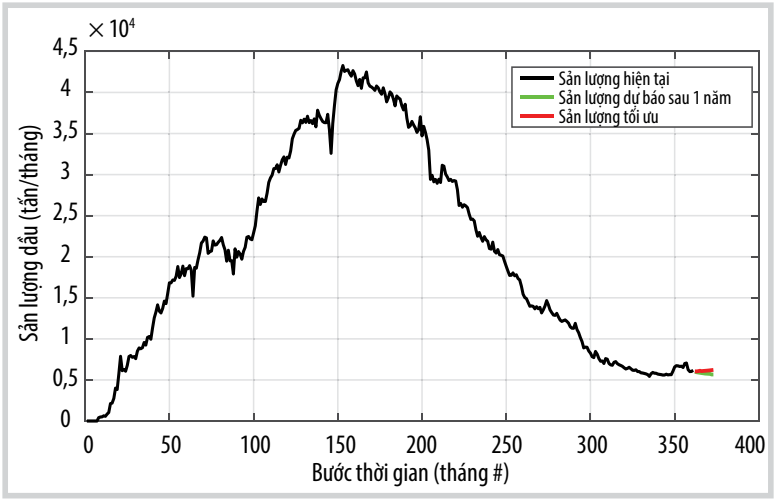


(a)



(b)

Hình 4. *Lưu lượng khai thác dầu được sử dụng để huấn luyện, kiểm tra và dự báo sau đó 1 năm với giả thiết lưu lượng bơm ép vẫn giữ nguyên so với hiện tại (a); hệ số tương quan giữa kết quả dự báo cho tập huấn luyện và kiểm tra (b).*



Hình 5. *Kết quả tối ưu bơm ép.*

huấn luyện và kiểm tra cho thấy mô hình này là chấp nhận được cho việc dự đoán và tối ưu hóa tiếp theo.

4. Tối ưu hóa bơm ép nước

Để tối ưu hóa lược đồ bơm ép nước (water injection scheme), hàm mục tiêu giá trị hiện tại ròng (NPV) được định nghĩa như sau:

$$NPV(u) = \sum_{i=1}^{n_t} \frac{\left(Q_o^i(u)P_o - Q_w^i(u)P_w - Q_{wi}^i(u)P_{wi}\right)\Delta t_i}{(1 + interest\ rate)^{\frac{t_i}{D}}} \quad (5)$$

Trong đó u là vector tham số tối ưu hóa; Q^i là lưu lượng bơm ép toàn mỏ tại bước thời gian i; P là cho giá hoặc chi phí; các chỉ số dưới o, w và wi lần lượt chỉ dầu, nước khai thác và nước bơm ép; n_t là tổng số bước thời gian. Trong nghiên cứu này, giá dầu (P_o) được đặt là 90 USD/thùng, trong khi cả chi phí nước khai thác (P_w) và chi phí nước bơm ép (P_{wi}) đều là 3 USD/thùng. Hơn nữa, tham số tối ưu hóa được sử dụng ở đây là tốc độ bơm ép toàn mỏ. Do đó, bài toán tối ưu hóa sẽ liên quan đến việc điều chỉnh tốc độ bơm ép nước toàn mỏ trong vòng 12 tháng của năm tiếp theo.

Nhiều thuật toán tối ưu hóa có thể được sử dụng để tối đa hóa hàm mục tiêu (7) trong phương pháp điều chỉnh của nhóm tác giả đã mô tả trong Mục 3. Bảy thuật toán tối ưu hóa truyền thống khác nhau đã được tích hợp vào chương trình máy tính của nhóm tác giả gồm: Thuật toán giảm độ dốc dốc nhất (Steepest descent algorithm) (Roger Fletcher, 1987 [24], William H. Press và nnk, 1992 [25]); thuật toán Gauss-Newton (Gauss-Newton algorithm) (RogerFletcher, 1987 [24], William H. Press và nnk, 1992 [25]); thuật toán xấp xỉ ngẫu nhiên xáo trộn đồng thời (simultaneous pertubation stochastic approximation - SPSA algorithm) (J.C. Spall 1992, 1998 [26, 27]); thuật toán SIMPLEX (SIMPLEX algorithm) (J.A. Nelder và nnk, 1965 [28]); thuật toán tập hợp hướng (Direction set algorithm) Roger Fletcher, 1987 [24]; William H. Press và nnk, 1992 [25]; thuật toán gradient liên hợp (Conjugate gradient algorithm) Roger Fletcher, 1987 [24]; William H. Press và nnk, 1992 [25]; thuật toán ma trận biến thiên (Variable metric algorithm) Roger

Bảng 1. *Lưu lượng tối ưu các giếng bơm ép*

Giếng bơm ép	Lưu lượng dòng hiện tại (m^3/ngày)	Lưu lượng tối ưu (m^3/ngày)	Thay đổi (m^3/ngày)
1	416,8	407,1	-9,7
2	523,2	620,5	97,3
3	703,4	752,6	49,2
4	2.009,0	2.135,3	126,2
5	2.005,1	2.214,0	208,9
6	3.007,2	2.321,0	-686,2
7	602,7	635,4	32,6

Fletcher, 1987 [24]; William H. Press và nnk, 1992 [25]. Mô tả chi tiết về các thuật toán tối ưu hóa liệt kê trên đây có thể được tìm thấy trong các tài liệu tham khảo.

Đối với bài toán cụ thể tại mỏ Bạch Hổ, thuật toán Gauss-Newton đã được chọn và kết quả tối ưu hóa được thể hiện trong Hình 5 với lưu lượng bơm ép tối ưu được chỉ ra trong Bảng 1.

Nghiên cứu tối ưu hóa, dựa trên các thay đổi đề xuất đối với lưu lượng bơm ép nước, cho thấy rằng trong khi dự đoán tương lai theo lưu lượng hiện tại (đường màu xanh lục) tiếp tục đi xuống, thì lưu lượng tối ưu hóa (đường màu đỏ) đã làm chững lại phần suy giảm dốc nhất và giúp gia tăng nhẹ trong dự báo sản lượng ngắn hạn. Tác động tích cực này, dù nhỏ, đạt được nhờ phân bổ lại nước bơm ép, đáng chú ý nhất là giảm 686,2 m^3/ngày tại giếng bơm ép số 7 (có khả năng là giếng có watercut cao) và chuyển hướng phần lớn lượng nước đó đến giếng bơm ép số 6 (+208,9 m^3/ngày), giếng bơm ép số 4 (+126,2 m^3/ngày), và giếng bơm ép số 2 (+97,3 m^3/ngày). Việc điều chỉnh lưu lượng bơm ép tối ưu này nhằm mục đích cải thiện hiệu quả quét theo thể tích (volumetric sweep efficiency) bằng cách hướng dòng nước đến các khu vực chưa được quét trước đây của vỉa chứa, qua đó tối đa hóa lượng dầu còn lại có thể thu hồi.

5. Kết luận và kiến nghị

Một quy trình mới để tối ưu hóa bơm ép nước tại vỉa chứa móng Bạch Hổ đã được phát triển thành công, tích hợp mạng neural nhân tạo để dự báo sản lượng dầu với thuật toán tối ưu Gauss-Newton nhằm tối đa hóa giá trị hiện tại ròng. Mô hình mạng neural nhân tạo đã chứng tỏ được độ chính xác dự đoán chấp nhận được, thể hiện sự tương quan cao và sai khác rất nhỏ giữa sản lượng thực tế và sản lượng dự báo trên cả tập dữ liệu huấn luyện và tập dữ liệu kiểm tra.

Lược đồ bơm ép nước tối ưu hóa thu được đã cho thấy một phát hiện quan trọng: việc giảm đáng kể lưu lượng bơm tại giếng bơm ép số 7 là yếu tố chủ đạo cho dự đoán gia tăng tổng sản lượng dầu và nâng cao NPV. Điều cốt yếu là, quy trình làm việc dựa trên dữ liệu này mang lại tốc độ tính toán nhanh hơn so với các phương pháp mô phỏng số truyền thống. Với ưu điểm như vậy, phương pháp này được khuyến nghị sử dụng cho việc sàng lọc và phân tích sơ bộ, nhưng kết quả thu được cần phải được xác thực bằng đánh giá chuyên môn và mô phỏng số đầy đủ trước khi triển khai thực tế ngoài mỏ.

Phương pháp tối ưu hóa hiệu quả và thực tiễn này được khuyến nghị để ứng dụng cho các vỉa chứa móng tương tự trong bể Cửu Long, bao gồm các mỏ Rồng, Sư Tử Nâu, và Sư Tử Trắng.

Tài liệu tham khảo

[1] Andrea Capolei, Eka Suwartadi, Bjarne Foss, and John Bagterp Jørgensen, "Waterflooding optimization in uncertain geological scenarios", *Computational Geosciences*, Volume 17, Issue 6, pp. 991 - 1013, 2013. DOI: 10.1007/s10596-013-9371-1.

[2] Bjarne Foss, Bjarne Grimstad, and Vidar Gunnerud, "Production optimization–facilitated by divide and conquer strategies", *IFAC-PapersOnLine*, Volume 48, Issue 6, pp. 1 - 8, 2015. DOI: 10.1016/j.ifacol.2015.08.001.

[3] Paul M.J. Van den Hof, Jan Dirk Jansen, and Arnold Heemink, "Recent developments in model-based optimization and control of subsurface flow in oil reservoirs", *IFAC Proceedings Volumes*, Volume 45, Issue 8, pp. 189 - 200, 2012. DOI: 10.3182/20120531-2-NO-4020.00047.

[4] Bjarne Foss and John Petter Jenson, "Performance analysis for closed-loop reservoir management", *SPE Journal*, Volume 16, Issue 1, pp. 183 - 190, 2011. DOI: 10.2118/138891-PA.

[5] J.D. Jansen, R.M. Fonseca, S. Kahrobaei, M.M. Siraj, G.M. Van Essen, and P.M.J. Van den Hof, "The egg model-a geological ensemble for reservoir simulation", *Geoscience Data Journal*, Volume 1, Issue 2, pp. 192 - 195, 2014. DOI: 10.1002/gdj3.21.

[6] Tailai Wen, Marco R. Thiele, David Echeverría Ciaurri, Khalid Aziz, and Yinyu Ye, "Waterflood management using two-stage optimization with streamline simulation", *Computational Geosciences*, Volume 18, Issue 3, pp. 483 - 504, 2014. DOI: 10.1007/s10596-014-9404-4.

[7] Zhenyu Guo, Albert C. Reynolds, and Hui Zhao, "A physics-based data-driven model for history matching, prediction, and characterization of waterflooding performance", *SPE Journal*, Volume 23, Issue 2, pp. 367 - 395, 2018. DOI: 10.2118/182660-PA.

[8] Khalid Aziz and Antonín Settari, *Petroleum reservoir simulation*. Society of Petroleum Engineers, 1979. DOI: 10.2118/9781613999646.

[9] Akhil Datta-Gupta and Michael King, *Streamline simulation: Theory and practice*. Society of Petroleum Engineers, 2007. DOI: 10.2118/9781555631116.

[10] M.R. Thiele, R.P. Batycky, and D.H. Fenwick, "Streamline simulation for modern reservoir-engineering workflows", *Journal of Petroleum Technology*, Volume 62, Issue 1, pp. 64 - 70, 2010. DOI: 10.2118/118608-JPT.

[11] Nguyen Anh Phuong, *"Capacitance resistance modeling for primary recovery, waterflood and water-CO_2 flood"*, University of Texas at Austin, 2012.

[12] Daniel Brent Weber, *"The use of capacitance-resistance models to optimize injection allocation and well location in waterflood"*, University of Texas at Austin, 2009.

[13] Ali A. Yousef, Pablo Gentil, Jerry L. Jensen, and Larry W. Lake, "A capacitance model to infer interwell connectivity from production-and injection-rate fluctuations", *SPE Reservoir Evaluation & Engineering*, Volume 9, Issue 6, pp. 630 - 646, 2006. DOI: 10.2118/95322-PA.

[14] L.W. Lake, X. Liang, T.F. Edgar, A. Al-Yousef, M. Sayarpour, and D. Weber, "Optimization of oil production based on a capacitance model of production and injection rates", *Hydrocarbon economics and evaluation symposium Dallas, Texas, U.S.A., 1 - 3 April 2007*. DOI: 10.2118/107713-MS.

[15] Hui Zhao, Zhijiang Kang, Xiansong Zhang, Haitao Sun, Lin Cao, and Albert C. Reynolds, "A physics-based data-driven numerical model for reservoir history matching and prediction with a field application", *SPE Journal*, Volume 21, Issue 6, pp. 2175 - 2194, 2016. DOI: 10.2118/173213-PA.

[16] Rui Zhang and Hu Jia, "Production performance forecasting method based on multivariate time series and vector autoregressive machine learning model for waterflooding reservoirs", *Petroleum Exploration and Development*, Volume 48, Issue 1, pp. 201 - 211, 2021. DOI: 10.1016/S1876-3804(21)60016-2.

[17] Farzad Hourfar, Hamed Jalaly Bidgoly, Behzad Moshiri, Karim Salahshoor, and Ali Elkamel, "A reinforcement learning approach for waterflooding optimization in petroleum reservoirs", *Engineering Applications of Artificial Intelligence*, Volume 77, pp. 98 - 116, 2019. DOI: 10.1016/j.engappai.2018.09.019.

[18] Cuthbert Shang Wui Ng, Ashkan Jahanbani Ghahfarokhi, and Menad Nait Amar, "Production optimization under waterflooding with Long Short-Term Memory and metaheuristic algorithm", *Petroleum*, Volume 9, Issue 1, pp. 53 - 60, 2022. DOI: 10.1016/j.petlm.2021.12.008.

[19] Muhammad Gibran Alfarizi, Milan Stanko, and Timur Bikmukhametov, "Well control optimization in waterflooding using genetic algorithm coupled with artificial neural networks", *Upstream Oil and Gas Technology*, Volume 9, 2022. DOI: 10.1016/j.upstre.2022.100071.

[20] Le Ngoc Son, Mahmound Jamiolahmady, Jean-Marie Questiaux, and Mehran Sohrabi, "An integrated geology and reservoir engineering approach for modelling and history matching of a Vietnamese fractured granite basement reservoir", *EUROPEC/EAGE Conference and Exhibition, London, UK, 11 - 14 June 2007*. DOI: 10.2118/107141-MS.

[21] Le Ngoc Son, Phan Ngoc Trung, Yoshihiro Masuda, Sumihiko Murata, Nguyen The Duc, Sunao Takagi, and Ahmad Ghassemi, "Development of a method for adjusting rock compaction parameters and aquifer size from production data and its application to Nam-Su fractured basement reservoir of Vietnam", *Journal of Petroleum Science and Engineering*, Volume 210, 2022. DOI: 10.1016/j.petrol.2021.109894.

[22] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and

Edouard Duchesnay, "Scikit-learn: Machine learning in Python", *Journal of Machine Learning Research*, Volume 12, pp. 2825 - 2830, 2011.

[23] Diederik P. Kingma and Jimmy Ba, "Adam: A method for stochastic optimization", *3rd International Conference for Learning Representations, San Diego, 22 December 2014*. DOI: 10.48550/arXiv.1412.6980

[24] Roger Fletcher, *Practical methods of optimization, second edition*. John Wiley & Sons, 1987.

[25] William H. Press, William T. Vetterling, *Numerical in Fortran: the Art of Scientific Computing*. Cambridge University Press, 1992.

[26] J.C. Spall, "Multivariate stochastic approximation using a simultaneous perturbation gradient approximation", *IEEE Transactions on Automatic Contron*, Volume 37, Issue 3, pp. 332 - 341, 1992. DOI: 10.1109/9.119632.

[27] J.C. Spall, "Implementation of the simultaneous perturbation algorithm for stochastic optimization", *IEEE Transactions on Aero space Electronic System*, Volume 34, Issue 3, pp. 817 - 823, 1998. DOI: 10.1109/7.705889.

[28] J.A. Nelder and R. Mead, "A simplex method for function minimization", *Computer Journal*, Volume 7, Issue 4, pp. 308 - 313, 1965. DOI: 10.1093/comjnl/7.4.308.

AN ARTIFICIAL NEURAL NETWORK APPROACH TO OPTIMIZE THE WATER FLOODING OF BACH HO OILFIELD, OFFSHORE VIETNAM

Doan Huy Hien[1], Le The Hung[1], Tran Xuan Quy[1], Pham Truong Giang[1], Nguyen Minh Quy[1], Nguyen The Duc[2]
[1]Vietnam Petroleum Institute (VPI)
[2]Institute of Mechanics, Vietnam Academy of Science and Technology (VAST)
Email: hiendh.epc@vpi.pvn.vn

Summary

The predominant oil production offshore Vietnam originates from the Bach Ho basement reservoir, where the flow regime is highly complicated due to the heterogeneous spatial distribution of petrophysical properties such as porosity, permeability, and water saturation. Therefore, the conventional reservoir-simulation-based methods for forecasting oil production often yield limited accuracy or require substantial time and effort to optimize the dynamic parameters. Recent advances in machine learning (ML) algorithms enable more rapid and accurate prediction of oil production rates from water injection rates at individual injection wells. Once the oil rate prediction is achieved using ML approach, the waterflooding optimization can then be achieved by any suitable optimization algorithm.

In this research, an artificial neural network (ANN) algorithm is used, yielding very good results: the correlation coefficients between the predicted and actual values are 0.98 and 0.95 for training and testing datasets, respectively. Subsequently, the Gauss-Newton optimization algorithm is applied to determine the optimal water injection rates for each injection well, aiming to enhance oil productivity. The results show that the newly optimized injection schemes yield an average oil production increase of 1.5%.

Key words: Waterflooding optimization, reservoir simulation, artificial neural network, machine learning, Bach Ho oilfield.

DầuKhí
KHOA HỌC, CÔNG NGHỆ VÀ ĐỔI MỚI SÁNG TẠO PETROVIETNAM

NGHIÊN CỨU XÂY DỰNG MÔ HÌNH PHÂN CHIA SẢN PHẨM KHAI THÁC CHO CÁC MỎ KẾT NỐI TẠI VIETSOVPETRO TRÊN NỀN TẢNG ĐIỆN TOÁN ĐÁM MÂY

Vũ Mai Khanh, Trần Quốc Thắng, Lê Việt Dũng, Trần Lê Phương, Chu Văn Lương, Phạm Thành Vinh*, Lê Thị Đoan Trang
Liên doanh Vietsovpetro
Email: vinhpt.rd@vietsov.com.vn
https://doi.org/10.47800/PVSI.2025.03-03

Tóm tắt

Trong quá trình kết nối thu gom và vận chuyển dầu khí giữa các mỏ lân cận, việc phân chia sản phẩm đóng vai trò quan trọng nhằm đảm bảo quyền lợi của các nhà đầu tư. Đây là lĩnh vực mới xuất hiện trong những năm gần đây đối với các mỏ kết nối trên thềm lục địa Việt Nam. Mô hình phân chia sản phẩm sử dụng nhiều nguồn dữ liệu đầu vào, bao gồm các thông số lưu lượng và đặc tính chất lưu, để xử lý kết quả. Các mô hình thực nghiệm được áp dụng để xác định sự thay đổi trạng thái pha của sản phẩm tách bậc theo nhiệt độ và áp suất trong hệ thống thu gom và vận chuyển, từ đầu giếng khai thác đến điểm tàng trữ sản phẩm cuối cùng.

Bài báo giới thiệu một cách hệ thống quy trình tính toán phân chia sản phẩm, chuyển đổi số quy trình và mô phỏng thuật toán nhằm xây dựng mô hình phân chia sản phẩm AIT trên nền tảng đám mây. Việc ứng dụng mô hình sẽ giúp tối ưu hóa quá trình vận hành sản xuất, nâng cao độ chính xác và hiệu quả trong công tác phân chia sản phẩm cho các mỏ kết nối.

Từ khóa: Kết nối mỏ, phân chia dầu khí, mô hình phân chia, điện toán đám mây.

1. Giới thiệu

Tại Vietsovpetro, từ đầu những năm 1981, cơ sở hạ tầng khai thác dầu khí ngoài khơi được xây dựng, cải hoán, nâng cấp hàng năm và đến thời điểm hiện tại đã xây dựng được cơ sở hạ tầng khá hoàn chỉnh tại mỏ Bạch Hổ và mỏ Rồng ở Lô 09-1. Nhằm tận dụng, sử dụng hiệu quả cơ sở hạ tầng hiện có với mục đích tiết giảm chi phí đầu tư và vận hành khai thác các mỏ lân cận, Vietsovpetro đã phối hợp với các công ty dầu khí khác để thực hiện kết nối các mỏ này vào hệ thống thu gom xử lý sẵn có tại Lô 09-1 ngoài khơi phía Nam Việt Nam.

Tuy nhiên, việc đưa các phát hiện dầu khí không cùng chủ sở hữu vào khai thác bằng cách kết nối với mỏ Bạch Hổ và mỏ Rồng ở Lô 09-1 để sử dụng cơ sở hạ tầng tại đây lại phát sinh một trở ngại rất lớn về vấn đề phân chia sản phẩm khai thác giữa các mỏ kết nối với nhau và được các bên liên quan đặc biệt quan tâm. Yêu cầu bắt buộc đối với việc phân chia sản phẩm khai thác là phải thực hiện trên cơ sở khoa học, khách quan, công bằng, minh bạch để đảm bảo quyền lợi của các chủ mỏ. Các vấn đề tranh chấp phát sinh trong quá trình phân chia sản phẩm sẽ rất khó để giải quyết khi thiếu những cơ sở và dữ liệu cần thiết. Công tác xây dựng quy trình và mô hình phân chia sản phẩm dầu - khí luôn được chú trọng, thường xuyên kiểm tra để đảm bảo sự phù hợp với quá trình đồng bộ hệ thống thiết bị, công nghệ hiện có trong quá trình vận hành khai thác mỏ.

Ngày nhận bài: 7/3/2024.
Ngày phản biện đánh giá và sửa chữa: 7/3/2024 - 27/9/2024.
Ngày bài báo được duyệt đăng: 27/9/2024.

2. Nội dung và kết quả nghiên cứu mô hình phân chia sản phẩm

Quá trình phân chia sản phẩm dựa trên các kết quả đo đếm các thông số vật lý về khối lượng, thể tích, năng lượng. Trong đó phân chia theo khối lượng, thể tích áp dụng cho những trường hợp phân chia hydrocarbon lỏng, phân chia theo năng lượng thường được sử dụng cho phân chia sản phẩm khai thác ở dạng khí [2].

Dựa trên những điều kiện cụ thể của hệ thống công nghệ thu gom, đo lường sản phẩm khai thác và trên cơ sở thỏa thuận giữa các bên liên quan mà có các nguyên lý phân chia khác nhau được áp dụng phổ biến là:

- Phân chia sản phẩm theo khối lượng theo nguyên tắc phân chia ngược;

- Phân chia sản phẩm theo đơn vị thể tích theo nguyên tắc phân chia ngược;

- Phân chia sản phẩm theo thành phần chất lưu;

- Phân chia sản phẩm theo các mô hình mô phỏng, tính toán.

Mỗi nguyên lý phân chia sản phẩm nói trên đều có những ưu nhược điểm, nhưng phổ biến và thường dùng nhất là phân chia ngược theo tỷ lệ. Lưu ý rằng, tất cả các lý thuyết và tài liệu được phổ biến về nguyên lý phân chia sản phẩm khai thác đều rất tổng quát và không thể áp dụng cho các trường hợp cụ thể nếu thiếu các nghiên cứu chuyên sâu, bao gồm cả phương pháp phân chia ngược theo tỷ lệ.

Nguyên lý phân chia sản phẩm theo nguyên tắc phân chia ngược được sử dụng và thừa nhận rộng rãi trong ngành công nghệ khai thác dầu và khí. Xét một mô hình phân chia sản phẩm theo nguyên tắc ngược có n nguồn hydrocarbon vận chuyển ra điểm B để xử lý và tàng trữ, với lưu lượng thể tích dầu đo được tại B (Q_B) quy về cùng một điều kiện. Theo đó lưu lượng hydrocarbon phân chia cho từng nguồn (i = 1...n) sẽ là [2]:

$$Q_{i=1...n} = k_{back\ allocation} \times Q_{imeasured}$$

$$k_{back\ allocation} = \frac{Q_B}{\sum Qi^{i=1...n}}$$

Trong đó:

$Q_{i=1...n}$: Lượng hydrocarbon được phân chia cho nguồn i;

$Q_{imeasured}$: lưu lượng thể tích dầu đo được ở cùng một điều kiện;

Hệ số $k_{back\ allocation}$: Hệ số bất cân bằng.

Lượng dầu đo được tại các công trình X quy về điều kiện chuẩn được xác định dựa trên các tham số sau:

+ Hàm lượng nước W_X;

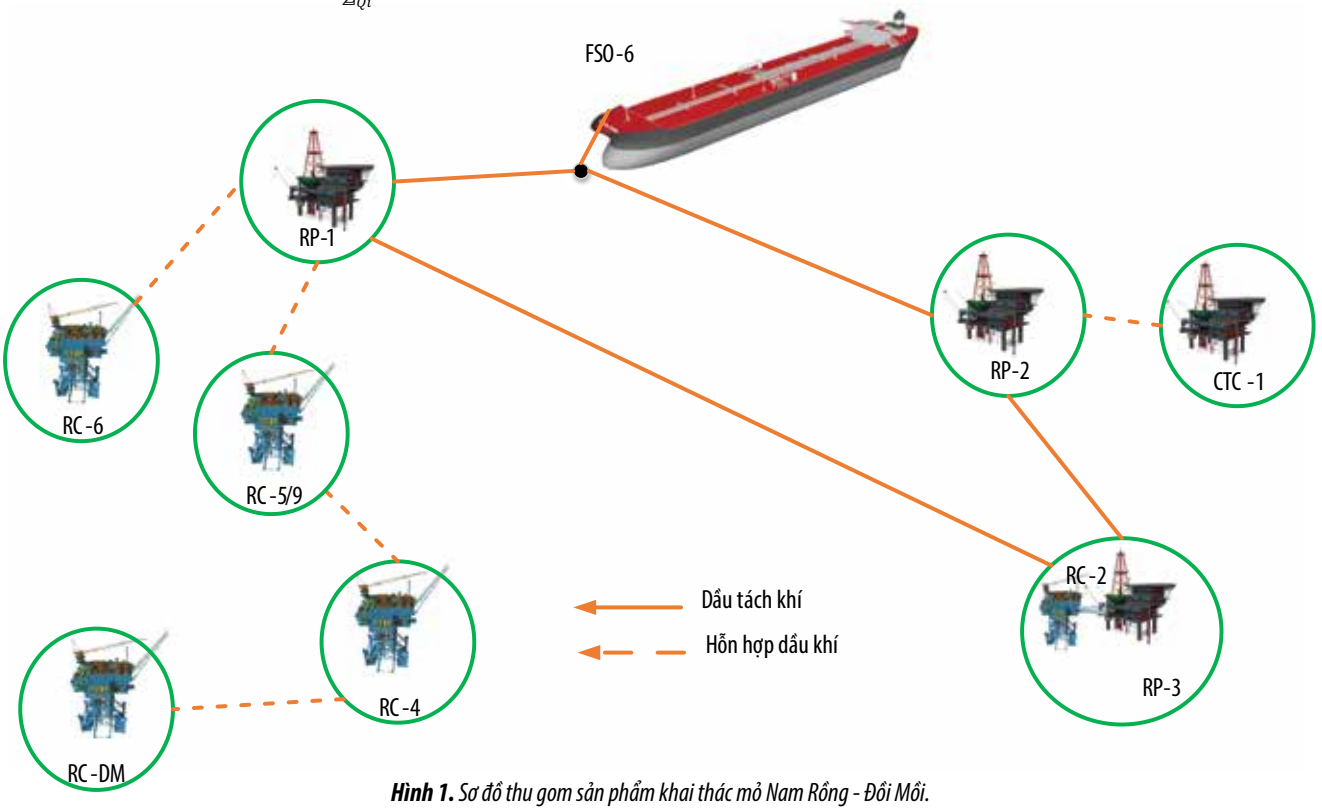
+ Lưu lượng theo thể tích chất lỏng ở điều kiện vận hành V_X;

+ Hệ số co ngót dầu của công trình X, S_X.

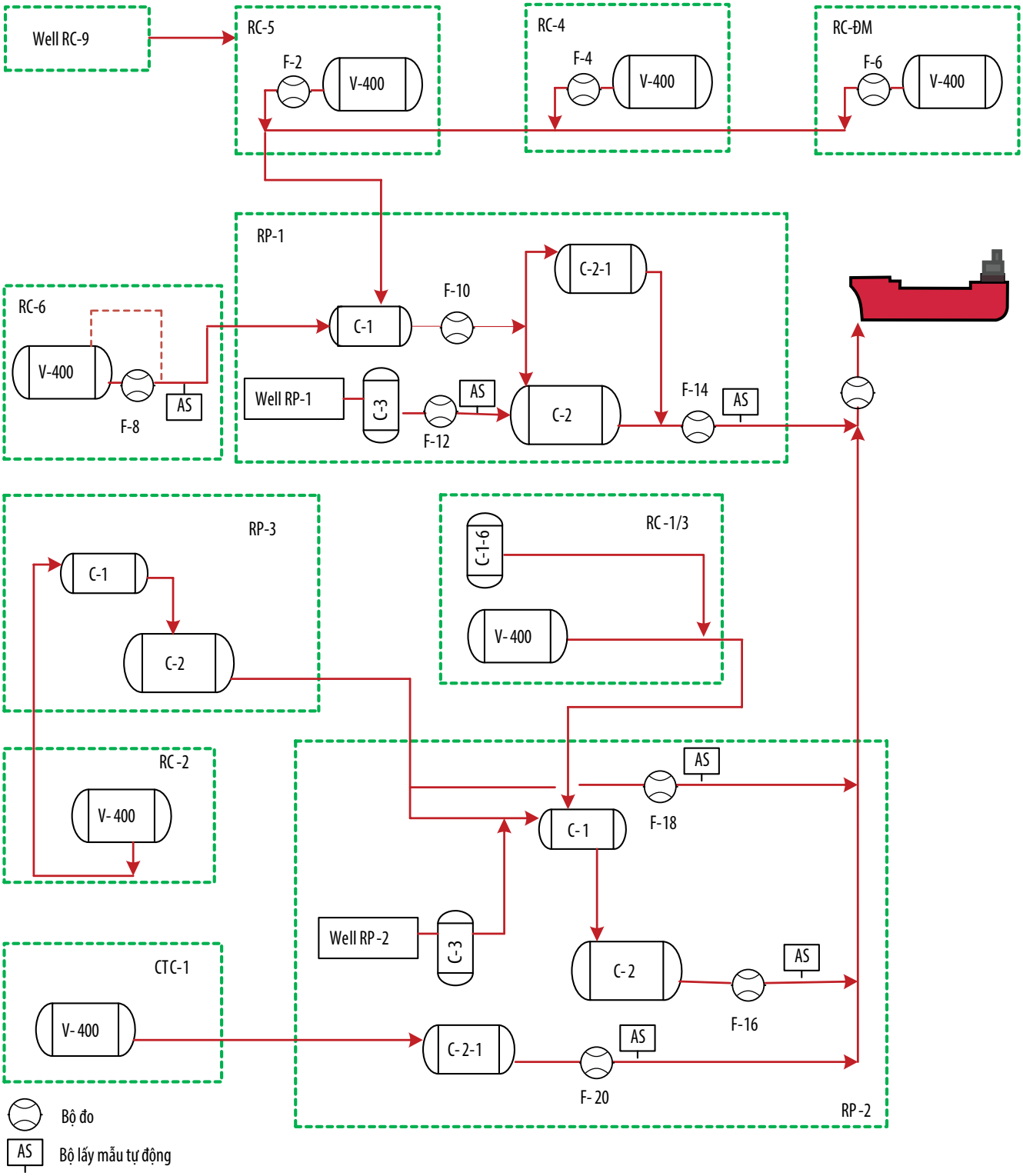
Như vậy, để giải quyết thành công bài toán phân chia sản phẩm dầu khí các mỏ kết nối chúng ta cần phải xác định chính xác số lượng sản phẩm tại điểm B (trên tàu chứa dầu FSO) và số lượng, tính chất sản phẩm của từng nguồn được vận chuyển tới điểm B. Để đưa về cùng một mặt bằng và thực hiện tính toán phân chia ngược, các thông số tính chất dầu khí ở các điều kiện vận hành khác nhau cần phải quy về cùng một điều kiện (thường là điều kiện chuẩn).

2.1. Phân chia sản phẩm mỏ Nam Rồng - Đồi Mồi

Mỏ Nam Rồng - Đồi Mồi (NR-ĐM) được kết nối vào hệ thống thu gom, vận chuyển sản phẩm khai thác sẵn có



Hình 1. *Sơ đồ thu gom sản phẩm khai thác mỏ Nam Rồng - Đồi Mồi.*



Hình 2. *Sơ đồ phân chia dòng dầu mỏ Nam Rồng - Đồi Mồi.*

của mỏ Rồng, Lô 09-1. Mỏ NR-ĐM có 2 giàn nhẹ được khai thác là RC-DM, RC-4.

Sản phẩm khai thác từ RC-DM, RC-4, RC-5/9 cùng với RC-6 được vận chuyển về RP-1 để tách khí và bơm về tàu chứa dầu FSO-6 để xử lý, tàng trữ và xuất bán. Tại FSO-6 đồng thời tiếp nhận các nguồn dầu bơm từ RP-2 bao gồm dầu mỏ Cá Tầm, RP-3. Sơ đồ thu gom vận chuyển và phân chia dòng dầu được thể hiện ở Hình 1 và 2.

Các công thức tính toán phân chia sản phẩm mỏ NR-ĐM đã được trình bày tại bài báo khoa học [2]. Ngoài ra, bằng thực nghiệm, phương trình tính hệ số co ngót cho các nguồn dầu được xác định như Bảng 1.

Trong đó: P: Áp suất tại điều kiện làm việc (psia);

T: Nhiệt độ tại điều kiện làm việc (°F);

P_{buf}: Áp suất bình tách (psia);

T_{buf}: Nhiệt độ bình tách (°F)

S: Hệ số co ngót (svol./vol.)

a, b, c, d, e, f, g, h, i, j: Các hệ sô thực nghiệm.

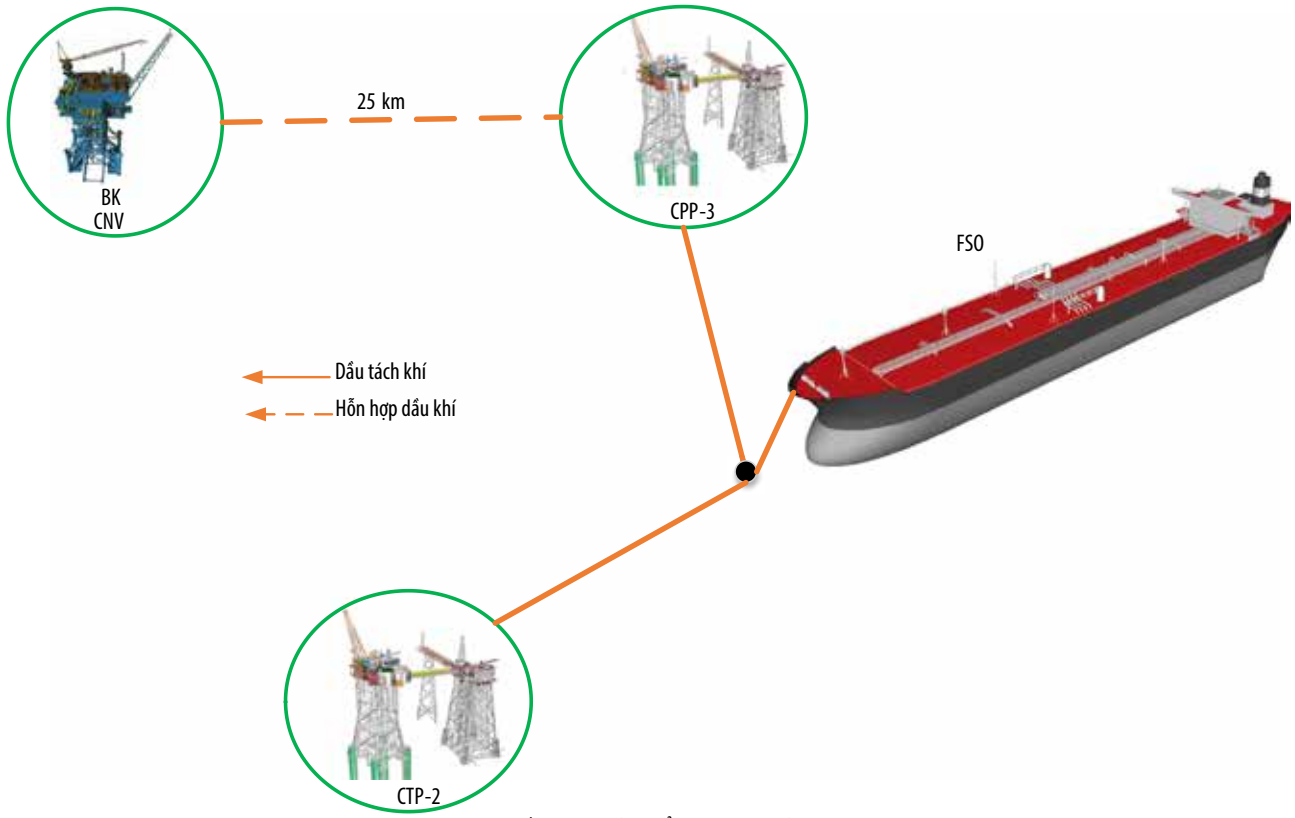
Các hệ số cho phương trình của Bảng 1 được xác định bằng kết quả phân tích hệ số co ngót mẫu dầu trong phòng thí nghiệm như Bảng 2.

Bảng 1. *Phương trình xác định hệ số co ngót cho các nguồn dầu*

Nguồn dầu	Công thức thực nghiệm
RP-1	$S = a \times Pbuf + b \times Tbuf + c \times P + d \times T + e$
RP-3	$S = a + b \times P + c/T + d \times P^2 + e/T^2 + f \times P/T + g \times P^3 + h/T^3 + i \times P/T^2 + j \times P^2/T$
RC-1/3 condensate	$S = a + b \times P + c/T + d \times P^2 + e/T^2 + f \times P/T + g \times P^3 + h/T^3 + i \times P/T^2 + j \times P^2/T$
RC-1/3 oil	$S = a + b \times P + c/T + d \times P^2 + e/T^2 + f \times P/T + g \times P^3 + h/T^3 + i \times P/T^2 + j \times P^2/T$
RP-2 oil	$S = a + b \times P + c/T + d \times P^2 + e/T^2 + f \times P/T + g \times P^3 + h/T^3 + i \times P/T^2 + j \times P^2/T$
CT C2-1	$S = a \times Pbuf + b \times Tbuf + c \times P + d \times T + e$

Bảng 2. *Kết quả phân tích hệ số co ngót cho các nguồn dầu*

Hệ số	RP-1	RP-3	RC-1/3 condensate	RC-1/3 oil
a	-1.39398E-04	0.852824	0.718812	0.772827
b	8.33320E-05	-4.18E-06	1.39E-05	1.34E-05
c	6.03621E-06	13.31185	46.47339	39.41745
d	-4.63042E-04	3.74E-10	-5.84E-11	-5.26E-10
e	1.01075E+00	214.9434	-3492.87	-3045.21
f		2.77E-03	-2.85E-03	-1.53E-03
g		8.84E-13	-5.55E-13	-4.07E-14
h		-42582.2	86548.16	86350.98
i		-0.15608	0.194113	0.085252
j		-1.81E-07	7.89E-08	5.45E-08



Hình 3. *Sơ đồ thu gom sản phẩm khai thác mỏ Cá Ngừ Vàng.*

2.2. Phân chia sản phẩm mỏ Cá Ngừ Vàng

Sản phẩm khai thác mỏ Cá Ngừ Vàng (CNV) được vận chuyển về giàn công nghệ trung tâm CTP-3 mỏ Bạch Hổ để xử lý. Dầu mỏ CNV xử lý tách nước và được bơm về tàu FSO để tàng trữ cùng với dầu mỏ Bạch Hổ từ giàn CTP-3 và giàn công nghệ trung tâm CTP-2.

Sơ đồ thu gom sản phẩm khai thác của mỏ CNV được thể hiện tại Hình 3. Sản phẩm của CNV đi vào 1 trong 3 đường công nghệ của CTP-3 và có thể được xử lý theo một đường công nghệ riêng hoặc trộn với dầu mỏ Bạch Hổ để tách khí nước và bơm về FSO.

Các công thức tính toán phân chia sản phẩm mỏ CNV được trình bày tại bài báo khoa học [2]. Ngoài ra, theo thực nghiệm, tỷ trọng dầu và hệ số co ngót cho các nguồn dầu được xác định như sau:

- Tỷ trọng và hệ số co ngót của dầu mỏ CNV bình tách:

$$D(P_{Sep'}, T_{Sep}) = a + b \times P_{Sep} + \frac{c}{T_{Sep}} + d \times P_{Sep^2} + \frac{e}{T_{Sep^2}} + f \times \frac{P_{Sep}}{T_{Sep}} + g \times P_{Sep^3} + \frac{h}{T_{Sep^3}} + i \times \frac{P_{Sep}}{T_{Sep^2}} + j \times \frac{P_{Sep^2}}{T_{Sep}}$$

Trong đó:

a = 6.92E-01 f = 0.00E+00

b = -1.02E-03 g = 0.00E+00

c = 5.23E+00 h = 0.00E+00

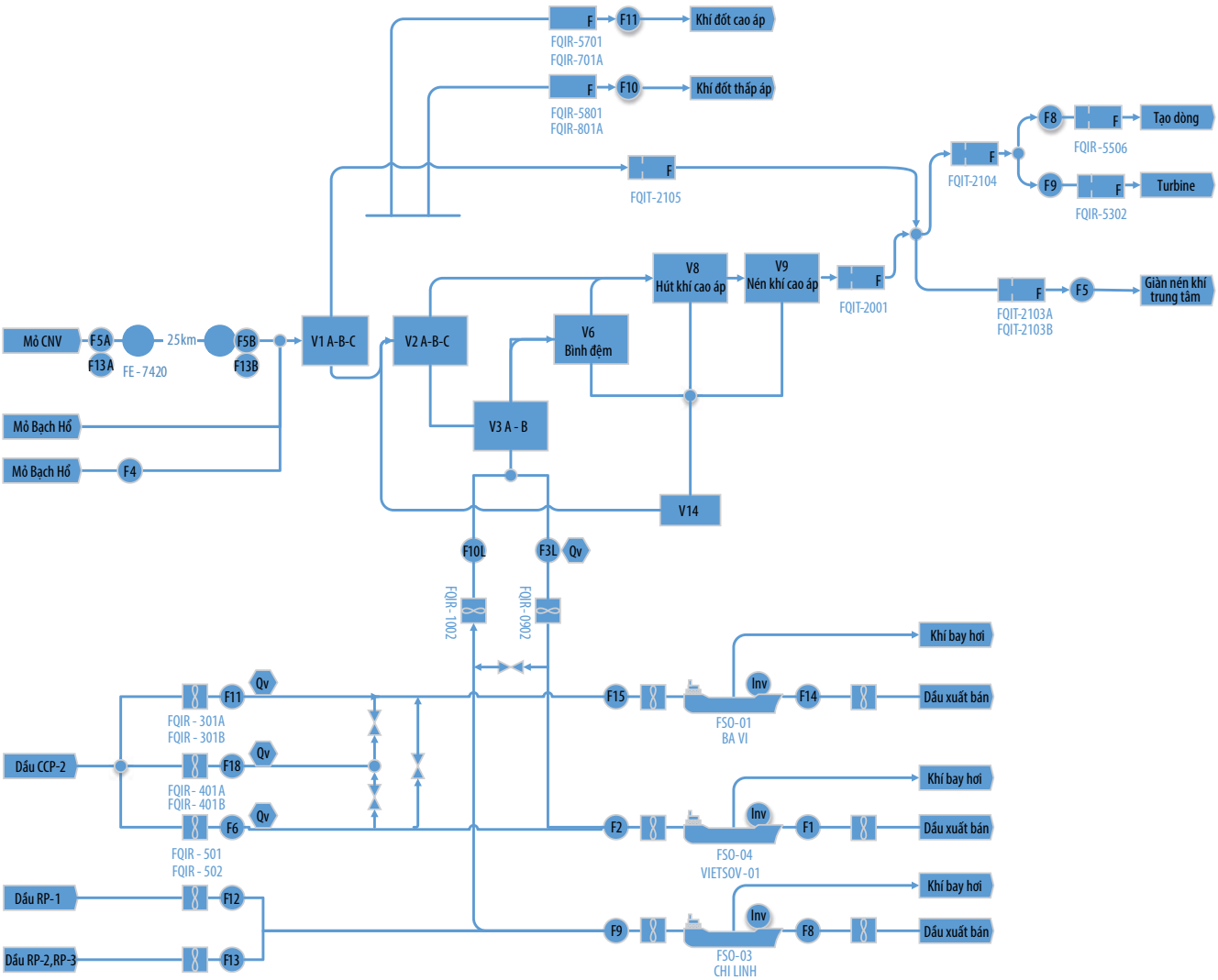
d = 0.00E+00 i = 0.00E+00

e = -7.70E+01 j = 0.00E+00

$$S(P_{Sep'}, T_{Sep}) = a + b \times P_{Sep} + \frac{c}{T_{Sep}} + d \times P_{Sep^2} + \frac{e}{T_{Sep^2}} + f \times \frac{P_{Sep}}{T_{Sep}} + g \times P_{Sep^3} + \frac{h}{T_{Sep^3}} + i \times \frac{P_{Sep}}{T_{Sep^2}} + j \times \frac{P_{Sep^2}}{T_{Sep}}$$

Trong đó:

P_{Sep}, T_{Sep}: Áp suất, nhiệt độ bình tách.



Hình 4. *Sơ đồ phân chia dòng dầu mỏ Cá Ngừ Vàng.*

a = 8.46E-01 f = 0.00E+00

b = -1.30E-03 g = 0.00E+00

c = 5.43E+00 h = 0.00E+00

d = 0.00E+00 i = 0.00E+00

e = -7.61E+01 j = 0.00E+00

- Tỷ trọng và hệ số co ngót của dầu sau bơm BH-CPP3:

$$D\&S(P_{Buffer}, T_{Buffer}, P_{Pump}, T_{Pump}) = a \times P_{Buffer} + b \times T_{Buffer} + c \times P_{Pump} + d \times T_{Pump} + d$$

Hệ số co ngót

a = -5.44E-03 f = 0.00E+00

b = 2.68E-04 g = 0.00E+00

c = 8.25E-05 h = 0.00E+00

d = -8.71E-04 i = 0.00E+00

e = 9.85E-01 j = 0.00E+00

Tỷ trọng

a = -2.23E-03 f = 0.00E+00

b = 9.95E-05 g = 0.00E+00

c = 6.96E-05 h = 0.00E+00

d = -7.34E-04 i = 0.00E+00

e = 8.35E-01 j = 0.00E+00

P_{Buffer}: Áp suất bình tách (barg);

T_{Buffer}: Nhiệt độ bình tách (°C);

P_{Pump}: Áp suất bơm (barg);

T_{Pump}: Nhiệt độ bơm (°C).

- Tỷ trọng và hệ số co ngót của dầu sau bơm BH-CPP2:

$$D\&S(P_{Buffer}, T_{Buffer}, P_{Pump}, T_{Pump}) = a \times P_{Buffer} + b \times T_{Buffer} + c \times P_{Pump} + d \times T_{Pump} + d$$

Hệ số co ngót

a = -4.70E-03 f = 0.00E+00

b = 1.93E-04 g = 0.00E+00

c = 8.88E-05 h = 0.00E+00

d = -8.80E-04 i = 0.00E+00

e = 9.93E-01 j = 0.00E+00

Tỷ trọng

a = -1.26E-03 f = 0.00E+00

b = 5.18E-05 g = 0.00E+00

c = 7.38E-05 h = 0.00E+00

d = -7.34E-04 i = 0.00E+00

e = 8.32E-01 j = 0.00E+00

P_{buffer}: Áp suất bình tách (barg);

T_{buffer}: Nhiệt độ bình tách (°C);

P_{Pump}: Áp suất bơm (barg);

T_{Pump}: Nhiệt độ bơm (°C).

- Tỷ trọng và hệ số co ngót của dầu sau bơm CNV-CPP3:

$$D\&S(P_{Buffer}, T_{Buffer}, P_{Pump}, T_{Pump}) = a \times P_{Buffer} + b \times T_{Buffer} + c \times P_{Pump} + d \times T_{Pump} + d$$

Hệ số co ngót

a = -7.80E-03 f = 0.00E+00

b = 3.31E-04 g = 0.00E+00

c = 1.04E-04 h = 0.00E+00

d = -9.23E-04 i = 0.00E+00

e = 9.82E-01 j = 0.00E+00

Tỷ trọng

a = -3.46E-03 f = 0.00E+00

b = 1.69E-04 g = 0.00E+00

c = 8.36E-05 h = 0.00E+00

d = -7.41E-04 i = 0.00E+00

e = 7.93E-01 j = 0.00E+00

P_{Buffer}: Áp suất bình tách (barg);

T_{Buffer}: Nhiệt độ bình tách (°C);

P_{Pump}: Áp suất bơm (barg);

T_{Pump}: Nhiệt độ bơm (°C).

- Tỷ trọng và hệ số co ngót của hỗn hợp dầu sau bơm BH-CPP3 và CNV-CPP3:

$$D(P_{Pump}, T_{Pump}, Mixratio) = a \times P_{Pump} + b \times T_{Pump} + c \times Mixratio + d$$

a = 7.60E-05

b = -7.37E-04

c = -3.90E-02

d = 8.40E-01

$$S(P_{Pump}, T_{Pump}, Mixratio) = a \times P_{Pump} + b \times T_{Pump} + c \times Mixratio + d$$

a = 9.17E-05

b = -8.90E-04

c = -4.06E-03

d = 9.98E-01

2.3. Phần mềm AIT phân chia dầu khí trên nền tảng điện toán đám mây

Trên cơ sở các quy trình phân chia sản phẩm đã được xác định, các tác giả tiến hành xây dựng phần mềm AIT phân chia dầu khí, mô phỏng động và tính toán trên nền tảng điện toán đám mây. Phần mềm AIT cho phép thiết lập mô hình phân chia sản phẩm tổng quát áp dụng vào phân chia dầu khí cho các mỏ kết nối, các nguồn sản phẩm khác nhau.

Để thiết lập mô hình phân chia tổng quát, nhóm phát triển đã áp dụng các công nghệ tiên tiến từ các hãng công nghệ hàng đầu như Google, Facebook, cũng như lưu trữ dữ liệu linh hoạt bằng MongoDB, một trong những hệ quản trị cơ sở dữ liệu NoSQL hiện được sử dụng phổ biến nhất thế giới.

NodeJS được phát hành vào năm 2009 và được duy trì bởi những người đóng góp từ khắp nơi trên thế giới. NodeJS, còn được biết với tên gọi chính thức là Node.js, là môi trường thời gian chạy (runtime environment) JavaScript đa nền tảng và mã nguồn mở. NodeJS cho phép các lập trình viên tạo cả ứng dụng front-end và back-end bằng JavaScript.

ReactJS là một mã nguồn mở được phát triển bởi Facebook, ra mắt vào năm 2013. Bản thân ReactJS là một thư viện JavaScript được dùng để xây dựng giao diện tương tác với các thành phần trên website. Một trong những điểm nổi bật nhất của ReactJS đó là việc trình bày dữ liệu không chỉ thực hiện được trên tầng máy chủ mà còn cả dưới trình duyệt.

Trong lập trình ứng dụng front-end, lập trình viên thường sẽ phải làm việc trên 2 thành phần chính là UI và xử lý tương tác của người dùng. UI là tập hợp những thành phần mà bạn nhìn thấy được trên bất kỳ một ứng dụng nào có thể kể đến như: menu, thanh tìm kiếm, những nút nhấn, card… Để tăng tốc quá trình phát triển và giảm những rủi ro có thể xảy ra trong khi lập trình, ReactJS còn cung cấp khả năng tái sử dụng mã nguồn (reusable code).

Chương trình phần mềm AIT cho phép giải quyết các bài toán phân chia sản phẩm khác nhau, ở mức độ tổng quát trên cơ sở các nguyên lý, nguyên tắc đã thống nhất, đồng thời cho phép thiết lập các cách thức phân chia, tùy biến, lập các cơ sở dữ liệu khác nhau làm đầu vào cho quá trình tính toán, xây dựng các cơ sở dữ liệu kết quả tính toán khác nhau để truy xuất các «project» theo các kết quả phân chia.

Chương trình sử dụng các nguyên tắc phân chia như sau:

- Nguyên tắc về phân chia ngược;
- Nguyên tắc quy về điều kiện chuẩn:

Theo nguyên tắc này, tổng lượng dầu đầu vào (in) và đầu ra (out) từ các nguồn trong mối tương quan lẫn nhau, được bảo toàn khi cùng đưa về điều kiện chuẩn. Hệ số chuyển đổi thể tích từ một điều kiện xác định về điều kiện chuẩn được hiểu là hệ số co ngót.

- Nguyên tắc về node:

Chương trình được xây dựng theo nguyên tắc về node. Mỗi node biểu thị cho bất kỳ một trong các đối tượng, loại hình công trình dầu khí: BK, MSP, CCP… Các node thể hiện các tính chất khác nhau của các công trình đó, bao gồm nguồn (source), hệ thống công nghệ xử lý thu gom dầu khí. Giữa các node liên hệ với nhau theo các nguyên lý tính toán phân chia đã đề cập như trên.

- Nguyên tắc về cấu hình hệ thống công nghệ:

Mỗi node, được hiểu như là một công trình biển, bao gồm và không giới hạn các loại hình: BK, MSP, mini BK, mini MSP, CCP… là tập hợp các hệ thống công nghệ khác nhau. Mỗi node cho phép hiển thị cấu hình hệ thống công nghệ, bao gồm:

+ Hệ thống bình tách;
+ Hệ thống bơm;
+ Hệ thống đo đếm;
+ Hệ thống lấy mẫu tự động (autosampler).

Việc sắp đặt các hệ thống này là tùy biến, cho phép thể hiện không giới hạn các cấu hình khác nhau của hệ thống, bao gồm tất cả các hệ thống tổng quan.

Giao diện và cách thức xây dựng mô hình cho các project được dựa trên nguyên tắc NODE:

1) NODE có 3 kiểu: OUTLET NODE, INLET NODE, END NODE

2) Mỗi NODE gồm 5 thành phần cơ bản:

- Bình tách (input: áp suất, nhiệt độ);
- Bơm (input: áp suất, nhiệt độ);
- Bộ đo lưu lượng (input: type, áp suất, nhiệt độ, lưu lượng theo khối lượng, lưu lượng theo thể tích, hàm lượng nước);
- Autosample (water-cut);
- Tính chất dầu thô.

3) Mỗi NODE được kết nối với nhau bằng LINE chỉ hướng của dòng chảy dầu, theo 2 cách thức:

- BYPASS: Dòng dầu từ NODE A không đi vào mà chỉ đi qua NODE B sau đó đi vào đường ống
- COMMINGLE: Dòng dầu từ NODE A đi vào
- SPLIT: Dòng dầu từ một NODE đi tới các NODE khác nhau

Kết quả thu được là kết quả tính toán cho từng NODE.

Chương trình có các phân vùng làm việc chính như sau:

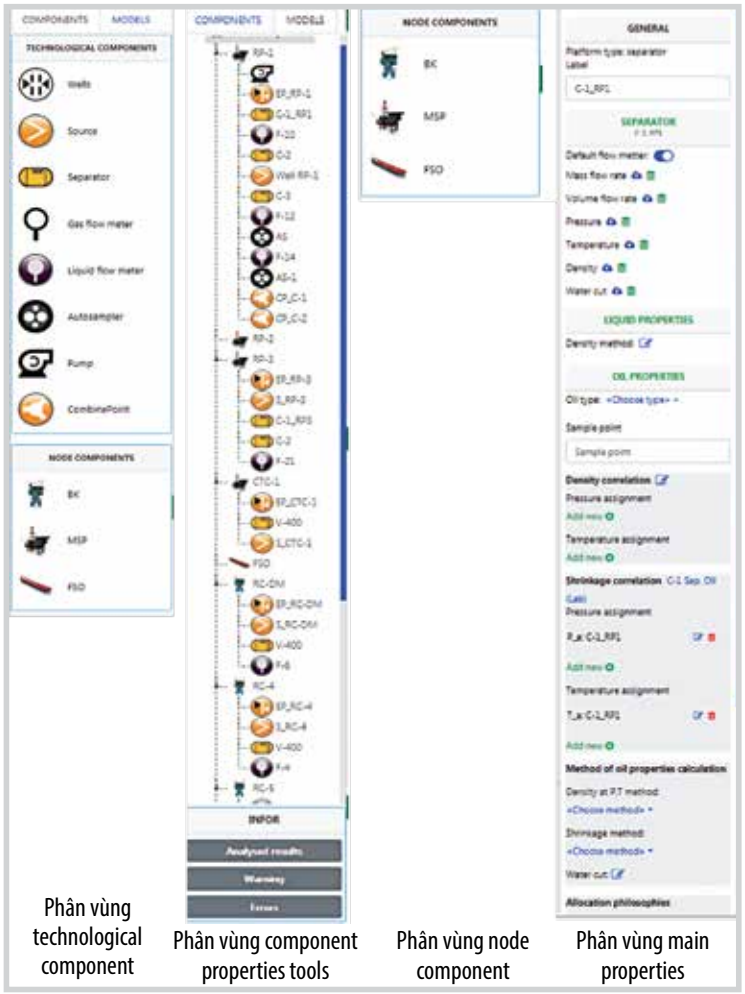
+ Phân vùng technological component;
+ Phân vùng node;
+ Phân vùng nodel working;
+ Phân vùng processing information;
+ Phân vùng mode view;
+ Phân vùng node view;
+ Phân vùng xử lý project;
+ Phân vùng tools;
+ Phân vùng report.

Giao diện của chương trình AIT được thể hiện tại Hình 5.

Chương trình AIT sử dụng để phân chia sản phẩm dầu khí các mỏ kết nối có các ưu điểm sau:

- Nghiên cứu áp dụng thành công các mô

Hình 5. *Giao diện của chương trình AIT.*



Hình 6. *Các phân vùng làm việc của chương trình AIT.*

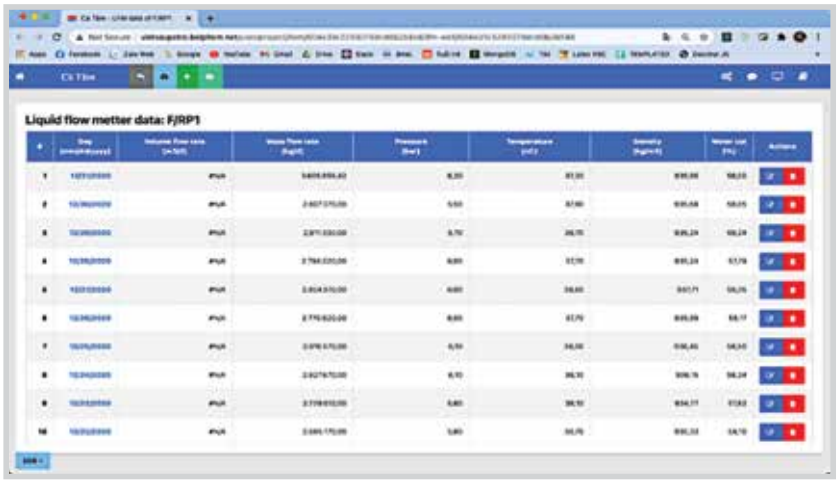
hình phân chia sản phẩm dầu khí đồng bộ với điều kiện đặc thù tại các mỏ kết nối;

- Xây dựng thành công hệ thống cơ sở dữ liệu tính chất chất lưu, thiết lập được mối tương quan với điều kiện khai thác thu gom, trên cơ sở đó phát triển các mô hình công thức thực nghiệm có độ chính xác cao;

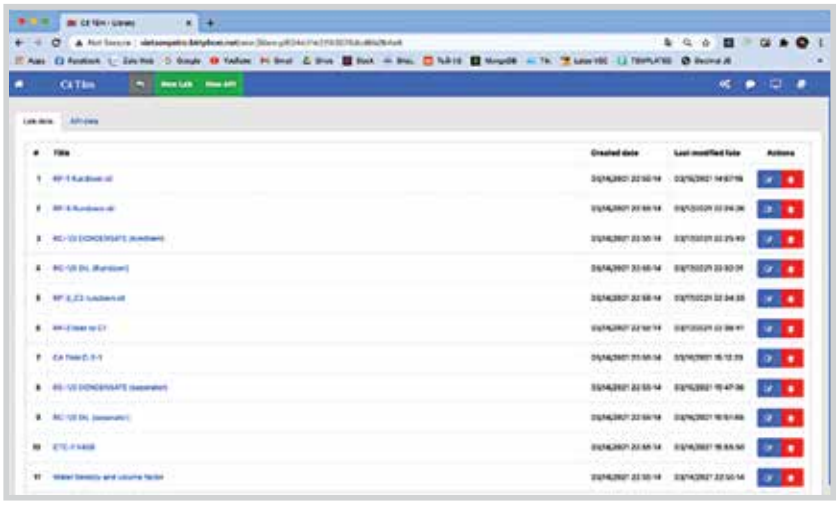
- Kết hợp dữ liệu thực nghiệm và bộ dữ liệu chuyên ngành cho phép đánh giá tính chất chất lưu chính xác và linh hoạt;

- Xây dựng mô hình phân chia có sự tham gia của nhiều nguồn chất lưu với tính chất khác nhau; xây dựng thành công mô hình đánh giá hỗn hợp chất lưu có khác biệt về tính chất;

- Lần đầu tiên xây dựng chương trình cho phép thiết lập mô hình phân chia cho hệ thống công nghệ khai thác dầu khí bất kỳ nhằm phân chia sản phẩm dầu khí cho các mỏ kết nối. Chương trình cũng cho phép đánh giá các thay đổi về tính chất chất lưu dầu khí trong quá trình xử lý công nghệ, thu gom, tàng trữ sản phẩm khai thác.



Dữ liệu input



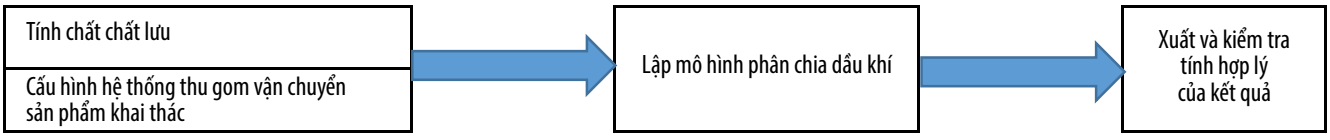
Các cấu trúc xử lý dữ liệu

Hình 7. *Cấu trúc xử lý dữ liệu của chương trình AIT.*

3. Kết luận

Bộ dữ liệu cơ sở các quy trình phân chia sản phẩm và chương trình AIT cho phép thiết lập mô hình phân chia cho bất kỳ một hệ thống công nghệ khai thác dầu khí nào nhằm phân chia sản phẩm một cách công bằng và minh bạch, làm tiền đề để các bên nhà thầu dầu khí chấp nhận kết quả phân chia. Bộ dữ liệu cơ sở các quy trình phân chia sản phẩm và chương trình AIT được áp dụng trong lĩnh vực khai thác dầu khí, công nghiệp năng lượng khi cần tiến hành phân chia dầu khí dưới các hình thức khác nhau, có mức tùy biến và độ chính xác cao, trên cơ sở các mô hình phân chia đã được phát triển và áp dụng.

Trong thời gian tới, xu thế kết nối mỏ sẽ trở thành một trong những định hướng phát triển chủ đạo trong hoạt động khai thác dầu khí tại Việt Nam nhằm tối ưu hóa việc sử dụng cơ sở hạ tầng hiện có và giảm thiểu chi phí thông qua chia sẻ nguồn lực. Trên cơ sở nghiên cứu các mô hình phân chia sản phẩm dầu khí, nhóm tác giả tại Vietsovpetro đã xây dựng các mô hình phân chia dựa trên các nguyên lý tách bậc sản phẩm kết hợp với mô hình biến đổi đặc tính chất lưu thiết lập từ thực nghiệm. Phần mềm AIT lần đầu tiên được nhóm tác giả phát triển và xây dựng, cho phép thiết lập mô hình phân chia sản phẩm phù hợp với bất kỳ một hệ thống công nghệ khai thác dầu và



Hình 8. *Dataflow của chương trình AIT.*

khí nào, phục vụ hiệu quả cho việc phân chia sản phẩm dầu khí giữa các mỏ kết nối.

Tài liệu tham khảo

[1] American Petroleum Institute, "Manual of petroleum measurement standards Chapter 20.3 measurement of multiphase flow", *The publication adopts the term mathematical process in their defining of allocation and use the term entity rather than contributing source, 2013*.

[2] Trần Lê Phương, Lê Đăng Tâm, Chu Văn Lương, Phạm Thành Vinh, Nguyễn Vi Hùng, Tống Cảnh Sơn, Nguyễn Viết Văn, Đỗ Dương Phương Thảo, A.G. Axmadev, Châu Nhật Bằng, Nguyễn Hữu Nhân, Đoàn Tiến Lữ, Trần Thị Thanh Huyền, Lê Thị Đoan Trang và Bùi Mai Thanh Tú, "Nghiên cứu, xây dựng phát triển các mô hình phân chia sản phẩm tại các mỏ kết nối của Vietsovpetro", *Tạp chí Dầu khí*, số 1, trang 41 - 48, 2021. Doi: 10.47800/pvj.2021.01-02.

DEVELOPMENT OF A PRODUCTION ALLOCATION MODEL FOR TIE-IN OIL FIELDS AT VIETSOVPETRO ON THE CLOUD COMPUTING PLATFORM

Vu Mai Khanh, Tran Quoc Thang, Le Viet Dung, Tran Le Phuong, Chu Van Luong, Pham Thanh Vinh*, Le Thi Doan Trang
Vietsovpetro
Email: vinhpt.rd@vietsov.com.vn

Summary

In the process of integrating oil and gas gathering and transportation between adjacent fields, production allocation plays a crucial role for tie-in oil and gas fields in ensuring the interests of investors. In recent years, this issue has emerged for tie-in fields on the continental shelf of Vietnam.

The production allocation model integrates multiple input data sources, including flow parameters and fluid properties, to process the outcomes. Experimental models are applied to determine phase state variations of stage-separated products under different temperature and pressure conditions within the gathering and transportation system, from the production wellhead to the final storage point.

This paper systematically presents the computational process for production allocation, the digital transformation of workflows, and algorithmic simulation to develop the AIT production allocation model on a cloud computing platform. The implementation of this model is expected to optimize production operations, enhance accuracy, and improve efficiency in production allocation for tie-in fields..

Key words: Tie-in fields, allocation, allocation model, cloud computing.

DầuKhí
KHOA HỌC, CÔNG NGHỆ VÀ ĐỔI MỚI SÁNG TẠO PETROVIETNAM

KHAI THÁC DỮ LIỆU CHUYÊN NGÀNH VỚI SỰ HỖ TRỢ CỦA AI: KINH NGHIỆM TỪ PHẦN MỀM QUẢN LÝ CƠ SỞ DỮ LIỆU ĐƯỜNG CONG ĐỊA VẬT LÝ GIẾNG KHOAN BỂ CỬU LONG

Vũ Tuyết Vy, Nguyễn Trung Sơn
Viện Dầu khí Việt Nam (VPI)
Email: vyvt@vpi.pvn.vn
https://doi.org/10.47800/PVSI.2025.03-04

Tóm tắt

Sự phát triển của trí tuệ nhân tạo (AI), đặc biệt là các mô hình ngôn ngữ lớn (LLMs) đã tạo ra sự thay đổi đáng kể trong quản lý và xử lý dữ liệu chuyên ngành, bao gồm dữ liệu có cấu trúc và phi cấu trúc. Bài viết giới thiệu kinh nghiệm ứng dụng AI trong khai thác dữ liệu tại Viện Dầu khí Việt Nam (VPI), đặc biệt trong hệ thống cơ sở dữ liệu đường cong địa vật lý giếng khoan bể Cửu Long. Đây là sản phẩm được phát triển dựa trên việc kết hợp AI để tối ưu quá trình quản lý, khai thác và phân tích dữ liệu địa vật lý giếng khoan, đảm bảo các yêu cầu về bảo mật.

Kết quả nghiên cứu cho thấy sản phẩm có khả năng tích hợp với các công cụ hỗ trợ khác, góp phần nâng cao trải nghiệm của người dùng; từ đó đề xuất các hướng phát triển trong tương lai nhằm tối ưu hóa hiệu quả và tính bảo mật của hệ thống.

Từ khóa: Trí tuệ nhân tạo, mô hình ngôn ngữ lớn, cơ sở dữ liệu, địa vật lý giếng khoan, bể Cửu Long.

1. Giới thiệu

Trong ngành công nghiệp - năng lượng, AI và học máy không chỉ giúp cải thiện quy trình thăm dò và khai thác mà còn tối ưu hóa việc bảo trì thiết bị, quản lý rủi ro và giảm thiểu chi phí sản xuất thông qua các công nghệ như phân tích dữ liệu thời gian thực và học sâu [1]. Các nghiên cứu cũng cho thấy AI hỗ trợ cải thiện an toàn lao động thông qua các hệ thống dự đoán và đánh giá rủi ro bằng phân tích dữ liệu quy mô lớn [2].

Việc khai thác hiệu quả dữ liệu là "chìa khóa" để đưa ra các quyết định chính xác và tối ưu hóa quy trình. Trong lĩnh vực khí tự nhiên, AI đã chứng minh khả năng phân tích dữ liệu từ các quy trình vận hành để hỗ trợ ra quyết định kịp thời và chính xác, giảm thiểu rủi ro và tối ưu hóa hoạt động [3]. Tuy nhiên, việc áp dụng AI không chỉ yêu cầu nguồn dữ liệu chất lượng cao mà còn đối mặt với các thách thức như an ninh mạng, quản trị dữ liệu và đảm bảo tính minh bạch của các thuật toán [4]. Các hệ thống AI tiên tiến trong ngành công nghiệp năng lượng đang tập trung vào bảo trì dự đoán và giảm thời gian ngừng hoạt động, giúp tiết kiệm chi phí và tăng hiệu quả sản xuất [2].

Trí tuệ nhân tạo mở ra tiềm năng to lớn cho việc phát triển các sản phẩm số thông minh. Việc tích hợp AI vào các sản phẩm này có thể mang lại nhiều lợi ích như cá nhân hóa trải nghiệm người dùng, tự động hóa các tác vụ và dự đoán chính xác. Đặc biệt, AI giúp giảm chi phí sản xuất, tăng năng suất và nâng cao chất lượng sản phẩm. AI không chỉ cải thiện hiệu quả hoạt động mà còn hỗ trợ quản lý năng lượng bền vững và tối ưu hóa sản xuất hydrocarbon thông qua các chiến lược dự đoán và tối ưu hóa dựa trên AI (Gowe AI [1]). Tuy nhiên, các thách thức lớn vẫn tồn tại, bao gồm đào tạo nhân sự chuyên môn cao, đầu tư ban đầu lớn và giải quyết các vấn đề đạo đức liên quan đến dữ liệu và tính minh bạch [3]. Ngoài ra, việc xây dựng các khung quản trị phù hợp cũng là yếu tố then chốt để triển khai AI hiệu quả và bền vững.

2. Ứng dụng AI trong sản phẩm số đặc thù cho lĩnh vực công nghiệp - năng lượng tại Viện Dầu khí Việt Nam (VPI)

2.1. Ứng dụng AI trong khai thác dữ liệu có cấu trúc

Dữ liệu có cấu trúc là dữ liệu được tổ chức thành các bảng với các trường thông tin rõ ràng, dễ dàng phân tích

Ngày nhận bài: 30/12/2024.
Ngày phản biện đánh giá và sửa chữa: 30/12/2024 - 23/1/2025.
Ngày bài báo được duyệt đăng: 23/1/2025.

và truy vấn bằng các công cụ cơ sở dữ liệu truyền thống. Việc ứng dụng trí tuệ nhân tạo và học máy truyền thống đối với dữ liệu có cấu trúc trong lĩnh vực dầu khí diễn ra rất phổ biến. Tại VPI, các ứng dụng này đã sớm được triển khai, đặc biệt phát triển trong vài năm trở lại đây với hàng loạt các sản phẩm số trong tất cả các khâu từ thượng nguồn đến hạ nguồn, chia thành các nhóm chính: phân tích xu hướng tương quan và dự báo, phát hiện bất thường, tối ưu hóa.

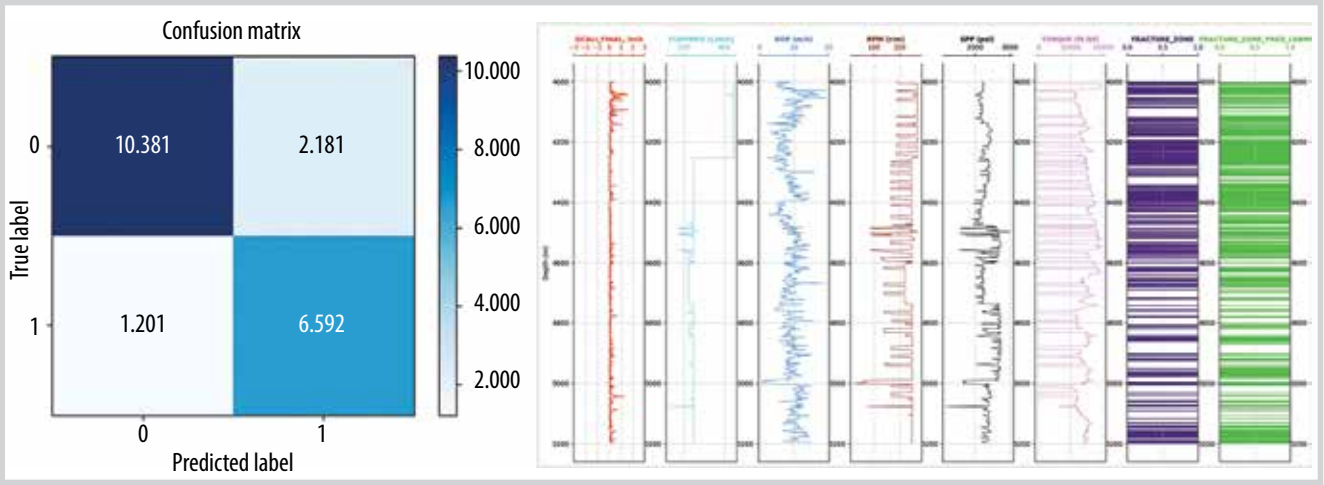
Đối với nghiệp vụ phân tích xu hướng tương quan và dự báo, AI hỗ trợ nhận diện mẫu (pattern recognition), phân tích đa chiều, khai phá, phân loại dữ liệu, dự báo… giúp các nhà nghiên cứu hiểu rõ hơn về các mối quan hệ phức tạp trong dữ liệu. Các sản phẩm tiêu biểu trong nhóm bao gồm: các mô hình kỹ thuật như mô hình dự báo độ thấm của vỉa chứa, mô hình dự báo đường cong địa vật lý giếng khoan, mô hình phân tích mối quan hệ giữa các thông số vỉa chứa…; các mô hình kinh tế như dự báo giá dầu thô Dated Brent, mô hình dự báo trích chi quỹ bình ổn xăng dầu…

Phát hiện bất thường trong dữ liệu có cấu trúc là một ứng dụng quan trọng của AI, giúp nhận diện các giá trị không bình thường hoặc các xu hướng bất thường trong quá trình khai thác dầu khí, được áp dụng trong ứng dụng quản lý khai thác của VPI. AI có thể phát hiện các sự thay đổi đột ngột trong các thông số đo đạc trong quá trình khai thác dầu khí, ví dụ thay đổi đột ngột trong nhiệt độ hoặc áp suất của miệng giếng hoặc đáy giếng, từ đó giúp các kỹ sư nhanh chóng điều chỉnh quá trình khai thác. Bên cạnh đó, các mô hình AI có thể nhận diện các dấu hiệu sớm của các vấn đề tiềm ẩn trong quá trình khai thác, như nguy cơ nổ hoặc rò rỉ dầu, giúp giảm thiểu rủi ro và đảm bảo an toàn cho quá trình vận hành.

Các bài toán tối ưu là một trong những bài toán điển hình của ngành (tối ưu vận hành nhà máy, tối ưu chi phí, tối ưu lựa chọn địa điểm…). Bên cạnh những thuật toán tối ưu truyền thống, việc ứng dụng đồng thời AI sẽ hỗ trợ tìm ra phương án tối ưu nhanh chóng, hiệu quả và tăng tính linh hoạt cho mô hình. Ngoài ra, khả năng học tăng cường, học theo thời gian cũng sẽ giúp mô hình thích nghi với các dữ liệu mới. Một số ứng dụng điển hình phải kể đến tối ưu hóa vị trí khoan, tối ưu hóa kỹ thuật khoan, tối ưu điều phối hệ thống vận chuyển nhiên liệu bằng xe bồn…

Khả năng về ngôn ngữ tự nhiên mạnh mẽ của các mô hình ngôn ngữ lớn đã mở ra hướng tiếp cận mới giúp khai thác dữ liệu có cấu trúc hiệu quả hơn. LLM rút ngắn khoảng cách giữa người dùng và cơ sở dữ liệu có cấu trúc với khả năng chuyển câu lệnh của người dùng thành các câu lệnh truy vấn hệ thống hiểu được (programming language) và diễn giải kết quả truy vấn bằng ngôn ngữ của người dùng. Kỹ thuật này được ứng dụng nhiều trong các sản phẩm số của VPI, chủ yếu trong lĩnh vực hạ nguồn, báo cáo kinh tế thị trường do các dữ liệu quen thuộc, phổ biến và yêu cầu bảo mật thấp. Việc ứng dụng kỹ thuật với dữ liệu tìm kiếm thăm dò khai thác gặp phải nhiều thách thức hơn do đặc thù của dữ liệu và yêu cầu bảo mật dữ liệu cao, cần phải thiết lập nền tảng và công nghệ chặt chẽ hơn, tiêu biểu là thử nghiệm với cơ sở dữ liệu địa vật lý giếng khoan bể Cửu Long (chi tiết trong phần tiếp theo).

Ngoài ra, sự kết hợp của LLM, các thuật toán AI truyền thống và hệ thống tự động hóa tác vụ (robotic process automation - RPA) giúp nhân rộng các ứng dụng đối với dữ liệu có cấu trúc thông qua các tác nhân (agents). Agents có khả năng tự lập kế hoạch, tự hành động và đưa ra quyết định dựa trên bối cảnh và yêu cầu mà người thiết



Hình 1. *Ứng dụng AI/ML trong việc xác định các đới nứt nẻ trong tìm kiếm thăm dò mỏ dầu khí. Nguồn: VPI, 2022.*

lập và người dùng cung cấp. Một dạng agent tiêu biểu đối với dữ liệu có cấu trúc đó là agent xây dựng báo cáo tự động. Sau khi nhận yêu cầu của người dùng (ví dụ như yêu cầu tạo báo cáo liên quan đến một chủ đề nhất định), agent tự động lập kế hoạch từng bước để xây dựng báo cáo, dựa vào nguồn dữ liệu đã được thiết lập sẵn hoặc có quyền truy cập để tự động trích xuất các dữ liệu liên quan, tự động phân tích, tính toán, vẽ biểu đồ và viết báo cáo theo chủ đề được yêu cầu.

2.2. Ứng dụng của AI trong khai thác dữ liệu phi cấu trúc

Sự phát triển của mô hình ngôn ngữ lớn cùng với khả năng kết nối với các thuật toán AI/ML mở ra tiềm năng khai thác dữ liệu phi cấu trúc (dữ liệu không có cấu trúc cố định, khó để sắp xếp vào các bảng cơ sở dữ liệu truyền thống nhưng lại chiếm tỷ trọng lớn). Các dạng dữ liệu phi cấu trúc phổ biến gồm: báo cáo thông tin địa chất, báo cáo mỏ, nhật ký khoan, hình ảnh mẫu lõi, mẫu vụn, mẫu chất lưu… và rất nhiều các loại báo cáo ở định dạng chữ, hình ảnh.

Một trong những ứng dụng phổ biến nhất ứng dụng AI đối với dữ liệu phi cấu trúc là xử lý ngôn ngữ tự nhiên, với đầu vào là tài liệu dạng văn bản. Xử lý ngôn ngữ tự nhiên (natural language processing - NLP) cho phép máy tính hiểu và xử lý ngôn ngữ tự nhiên của con người để thực hiện các tác vụ nâng cao khác như: tóm tắt, trích xuất thông tin, phân loại tài liệu.... Trong lĩnh vực thăm dò khai thác dầu khí, NLP có thể được ứng dụng để phân tích và trích xuất thông tin từ các báo cáo địa chất và nhật ký khoan. Trước khi có sự phát triển vượt bậc của các mô hình ngôn ngữ lớn, các ứng dụng NLP cũng đã xuất hiện tuy nhiên khá khiêm tốn, tập trung vào việc phân tích báo cáo địa chất, tóm tắt nhật ký khoan, dịch thuật tự động… Với sự phát triển của các mô hình NLP hiện nay, có thể tự động phân tích nội dung của các báo cáo địa chất, trích xuất các thông tin quan trọng như đặc điểm cấu trúc địa chất, thành phần khoáng vật và tự đánh giá tiềm năng chứa dầu khí theo những yêu cầu đã được thiết lập sẵn hoặc gợi ý của người sử dụng. Điều này giúp giảm thời gian xử lý dữ liệu và tăng độ chính xác trong việc nhận diện các yếu tố quan trọng. NLP có thể được sử dụng để tự động tóm tắt các ghi chép hàng ngày từ nhật ký khoan, giúp nhanh chóng nắm bắt được các thông tin quan trọng mà không cần đọc toàn bộ tài liệu.

Cùng với sự xuất hiện của ChatGPT và các mô hình ngôn ngữ lớn, việc xử lý ngôn ngữ tự nhiên nói chung và trong lĩnh vực dầu khí nói riêng cũng phổ biến và dễ dàng tiếp cận hơn, với độ chính xác cao hơn và khối lượng dữ liệu được xử lý cũng tăng mạnh. VPI đã tiếp cận hướng đi mới để phát triển các sản phẩm số thử nghiệm cho việc khai thác kho dữ liệu phi cấu trúc của VPI. Các ứng dụng chủ yếu tập trung vào sử dụng khả năng thấu hiểu ngôn ngữ tự nhiên của mô hình ngôn ngữ lớn kết hợp với kỹ thuật xử lý tài liệu theo hình thức vector hóa hoặc ứng dụng bản đồ tri thức (knowledge graph) và các kỹ thuật tìm kiếm nâng cao (retrieval-augmented generation - RAG) để xây dựng cơ sở dữ liệu tài liệu phi cấu trúc cùng với hệ thống chatbot để tìm kiếm, hỏi đáp các thông tin liên quan. Ngoài ra, LLM cũng hỗ trợ đắc lực trong việc tự động hóa tìm kiếm thông tin trên internet theo chủ đề xác định trước, tổng hợp và phân loại thông tin.

Sản phẩm thử nghiệm tiêu biểu phải kể đến việc ứng dụng LLM để phân tích các báo cáo địa chất và nhận diện các yếu tố quan trọng ảnh hưởng đến khả năng chứa dầu khí của bể Sông Hồng. Kết quả cho thấy, hệ thống có tích hợp LLM đã giảm được 20 - 30% thời gian xử lý dữ liệu và tăng độ chính xác trong việc phân loại các đặc điểm địa chất so với phương pháp thủ công, từ đó giúp việc tổng hợp và đánh giá thông tin trở nên dễ dàng và hiệu quả hơn.

Bên cạnh xử lý ngôn ngữ tự nhiên, việc ứng dụng thị giác máy tính (computer vision) cũng phổ biến trong lĩnh vực dầu khí nhằm phân tích hình ảnh mẫu lõi, tự động nhận dạng các đặc điểm địa chất như các tầng trầm tích khác nhau, vết nứt hoặc dấu hiệu của chất lưu trong mẫu lõi. Trong phân loại hình ảnh mẫu lõi, bằng việc sử dụng các neural network tích chập (CNN), hệ thống có thể tự động phân loại các loại đá và xác định các đặc điểm địa chất quan trọng từ hình ảnh mẫu lõi. Đồng thời có thể trợ giúp xác định những loại cổ sinh trong lát mỏng. Điều này không chỉ giúp tiết kiệm thời gian mà còn nâng cao độ chính xác trong việc phân tích mẫu lõi [3]. Đối với nhận diện đặc điểm các yếu tố địa chất, thị giác máy tính có thể phát hiện các dấu hiệu bất thường hoặc các đặc điểm địa chất quan trọng mà con người có thể bỏ sót như các hệ thống đứt gãy trên mặt cắt địa chấn, từ đó hỗ trợ quá trình thăm dò khai thác hiệu quả hơn.

3. Kinh nghiệm ứng dụng AI trong phần mềm quản lý cơ sở dữ liệu địa vật lý giếng khoan bể Cửu Long

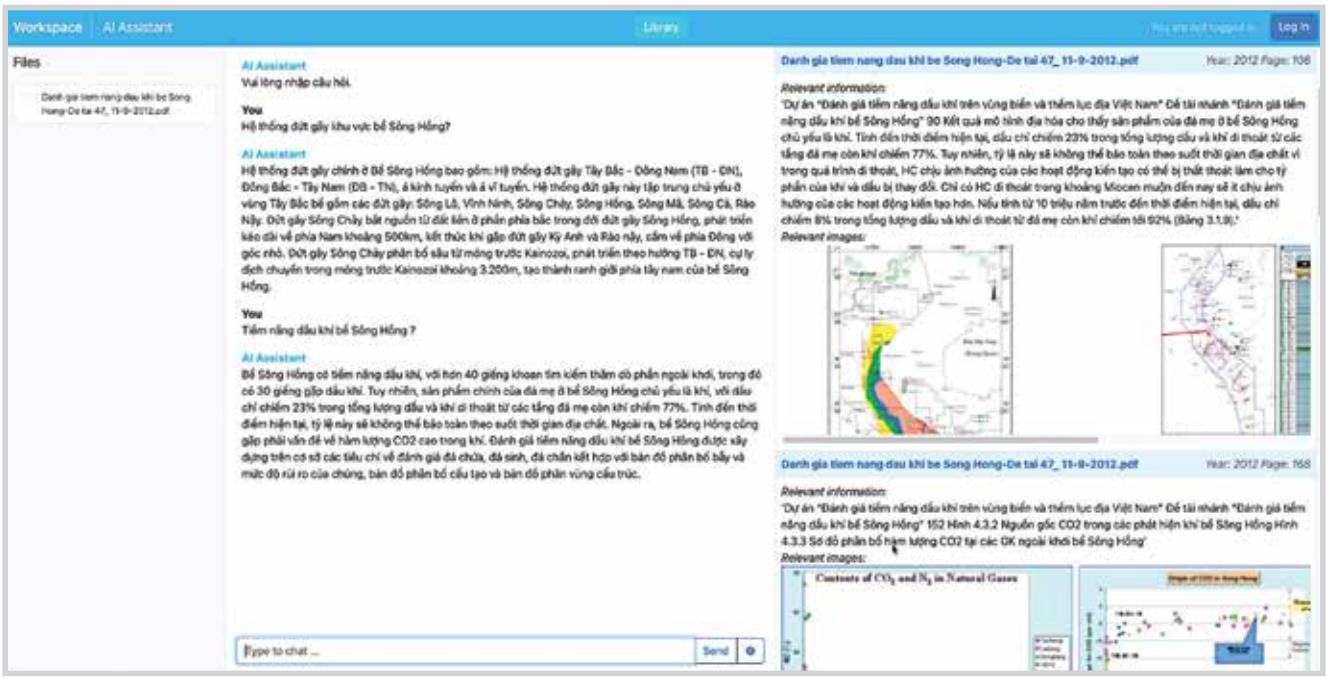
3.1. Phần mềm quản lý cơ sở dữ liệu địa vật lý giếng khoan bể Cửu Long

Cơ sở dữ liệu địa vật lý giếng khoan bể Cửu Long chứa lượng thông tin rất lớn về đặc điểm địa chất của các giếng khoan, bao gồm dữ liệu đo ghi địa vật lý giếng khoan, mẫu lõi và các kết quả phân tích (Hình 3). Việc ứng dụng AI vào quản lý và khai thác cơ sở dữ liệu này mang lại những lợi

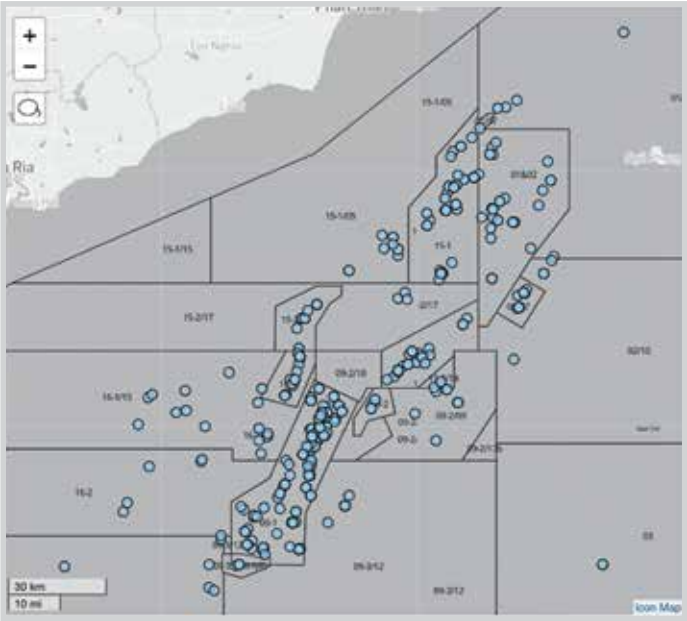
ích to lớn, như tăng cường độ chính xác trong phân tích và dự báo tiềm năng dầu khí [4]. Phần mềm quản lý cơ sở dữ liệu địa vật lý giếng khoan được xây dựng nhằm cung cấp một giải pháp toàn diện cho việc quản lý, khai thác và phân tích dữ liệu địa vật lý giếng khoan. Phần mềm được thiết kế với giao diện Power BI thân thiện, dễ sử dụng, đáp ứng các yêu cầu về an toàn bảo mật. Điểm nổi bật của sản phẩm này so với các sản phẩm truyền thống là khả năng tích hợp AI, cho phép người dùng tương tác với dữ liệu bằng ngôn ngữ tự nhiên và tự động hóa nhiều tác vụ, chạy trên máy tính nội bộ.

Giao diện người dùng đáp ứng nhu cầu trực quan và các yêu cầu kỹ thuật chuyên môn, giúp người dùng dễ truy cập, phân tích và khai thác dữ liệu hiệu quả. Việc xây dựng giao diện trên nền tảng Power BI với khả năng tương tác động mang lại nhiều lợi thế vượt trội. Power BI cho phép người dùng tương tác trực tiếp với dữ liệu thông qua các biểu đồ, bảng và bản đồ. Người dùng có thể dễ dàng sàng lọc dữ liệu, "đào sâu" vào chi tiết và thay đổi cách hiển thị dữ liệu phù hợp với nhu cầu phân tích. Power BI có thể kết nối với nhiều nguồn dữ liệu khác nhau, bao gồm cơ sở dữ liệu, tệp Excel và các dịch vụ đám mây. Điều này cho phép phần mềm dễ tích hợp với cơ sở dữ liệu địa vật lý giếng khoan và các hệ thống khác (Hình 4).

Địa vật lý giếng khoan đóng vai trò quan trọng trong việc thu thập, quản lý và phân tích dữ liệu liên quan đến giếng khoan (ví dụ: logs, khảo sát địa chấn xung quanh giếng...). Để đảm bảo dữ liệu nhất quán, dễ truy cập, dễ tích hợp và chia sẻ, VPI đã áp dụng mô hình dữ liệu PPDM (professional petroleum data management) để chuẩn hóa và quản lý dữ liệu địa vật lý giếng khoan. Mô hình dữ liệu (data model) được xây dựng chặt chẽ, liên kết vùng thông tin chung của giếng với các vùng thông tin khác: Tầng đánh dấu (marker), quỹ đạo giếng khoan (deviation), đường cong địa vật lý giếng khoan (logs), đồng thời duy trì tính nhất quán về thông tin giếng khoan xuyên suốt các vùng dữ liệu. Các khái niệm và đơn vị đo đã được chuẩn hóa và thống nhất trong từ điển dữ liệu chung (data dictionary) để tránh nhầm lẫn và dễ dàng chuyển giao, kế thừa trong các sản phẩm số khác. Bên cạnh đó, sản phẩm cơ sở dữ liệu địa vật lý giếng khoan bể Cửu Long được kế thừa dữ liệu đã qua kiểm soát chất lượng (QC) và chuẩn hóa bởi nhóm chuyên gia lĩnh vực địa vật lý giếng khoan của VPI. Vấn đề bảo mật và phân quyền dữ liệu cũng được đặt lên hàng đầu nhằm đảm bảo việc quản trị thông tin minh bạch, hiệu quả và đáp ứng tốt nhu cầu khác nhau của người dùng trong lĩnh vực dầu khí. Do đó cấu trúc cơ sở dữ liệu địa vật lý giếng khoan bể Cửu Long có khả năng mở rộng dữ liệu áp dụng cho các bể trầm tích khác, như bể Nam Côn Sơn, hay bể Sông Hồng, nơi cũng có nhiều giếng khoan với lượng dữ liệu đồ sộ, với thiết kế tổng quát và linh hoạt. Đồng thời, hệ thống đồng bộ sẽ hỗ trợ so sánh và phân tích dữ liệu hiệu quả giữa các trường thông tin khác nhau, ví dụ như thông tin khai thác và thông tin thăm dò trong cùng một khu vực hay các khu vực khác nhau, góp phần nâng cao khả năng quản lý và ra quyết



Hình 2. *Chatbot hỏi đáp báo cáo chuyên ngành. Nguồn: VPI, 2024.*



Hình 3. *Vị trí thông tin giếng khoan trong bể Cửu Long. Nguồn: VPI, 2024.*

Hình 4. *Giao diện ứng dụng quản lý cơ sở dữ liệu giếng khoan. Nguồn: VPI, 2024.*

định. Trên thực tế, nhóm tác giả đã và đang ứng dụng công nghệ và kỹ thuật xây dựng cơ sở dữ liệu của bể Cửu Long cho bể Nam Côn Sơn mà không cần thay đổi về cấu trúc dữ liệu.

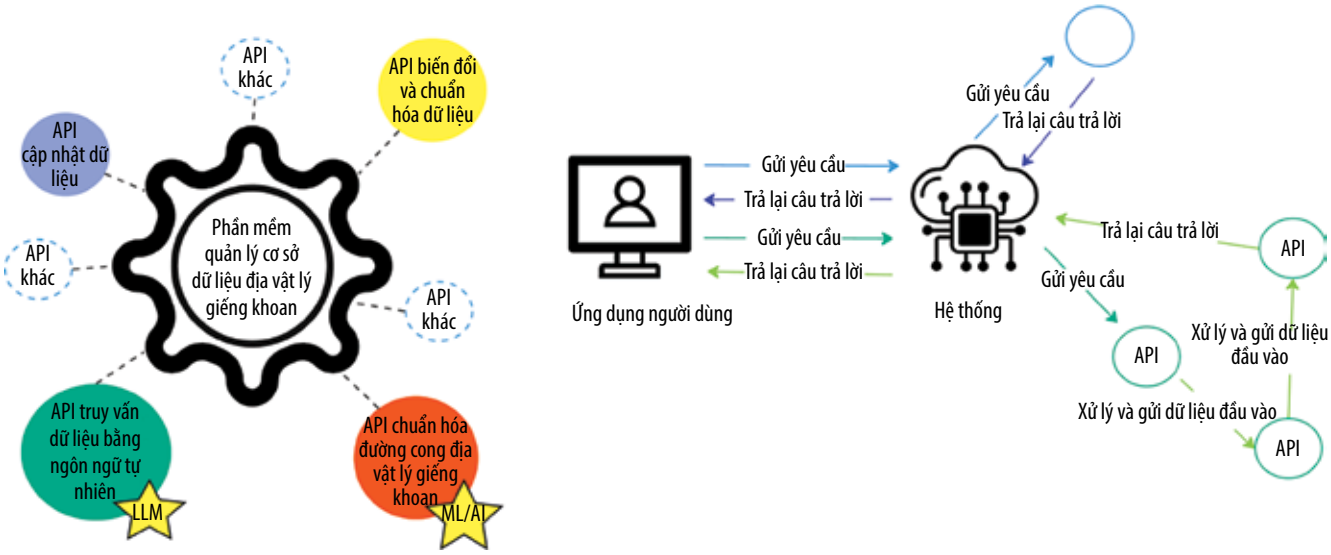
Với đặc thù của dữ liệu lĩnh vực tìm kiếm thăm dò, nền tảng yêu cầu truy cập an toàn, bảo mật và phân quyền chặt chẽ. Cơ sở dữ liệu tại chỗ (on-premise) kết hợp với hệ thống phân quyền chi tiết là giải pháp toàn diện mà nhóm tác giả đã triển khai để đảm bảo an toàn và kiểm soát dữ liệu. Dữ liệu được lưu trữ trên máy chủ nội bộ thay vì trên đám mây, từ đó giảm thiểu nguy cơ bị tấn công mạng hoặc mất mát dữ liệu do các lỗ hổng trên môi trường trực tuyến. Người dùng được giới hạn truy cập trực tiếp từ hạ tầng nội bộ theo phân quyền truy cập dựa trên các vai trò cụ thể, và có thể giới hạn phạm vi dữ liệu được truy cập (như theo giếng, mỏ, lô, bể…).

Một trong những điểm nổi bật trong thiết kế hệ thống của nền tảng quản lý cơ sở dữ liệu giếng khoan là khả năng tích hợp với các công cụ chuyên môn khác thông qua mảnh ghép số (API). Với định hướng của VPI trong việc phát triển hình thức mảnh ghép số và tích hợp, tái sử dụng trong các sản phẩm số khác nhau, nhóm tác giả lựa chọn nền tảng triển khai sản phẩm dễ dàng mở rộng trong tương lai, với định hướng nền tảng là nơi tập trung công cụ duy nhất hỗ trợ nhóm chuyên môn trong việc khai thác dữ liệu địa vật lý giếng khoan. Tính đến nay sản phẩm đã tích hợp thêm 2 mảnh ghép số khác từ các tác giả bộ phận chuyên môn và 1 mảnh ghép số hỗ trợ khai thác dữ liệu bằng ngôn ngữ tự nhiên phát triển bởi bộ phận Phân tích dữ liệu: API dự báo đường cong địa vật lý giếng khoan bị thiếu, API hiệu chỉnh đường cong địa vật lý giếng khoan, API chatbot hỏi đáp về dữ liệu.

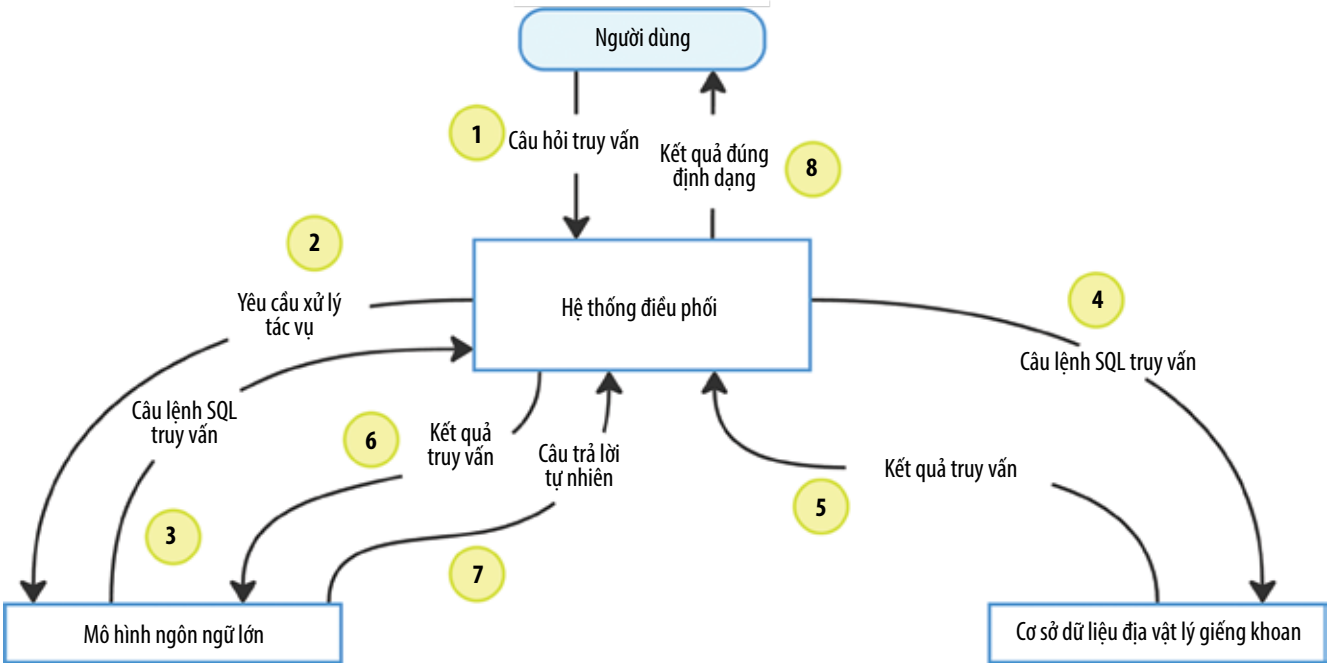
3.2. Kinh nghiệm ứng dụng AI trong phần mềm quản lý cơ sở dữ liệu địa vật lý giếng khoan bể Cửu Long

Khả năng hỏi đáp bằng ngôn ngữ tự nhiên là tính năng đột phá của ứng dụng, giúp người dùng tương tác với dữ liệu dễ dàng và tự nhiên hơn. Thay vì phải sử dụng các câu lệnh phức tạp hoặc truy vấn SQL, người dùng có thể đặt câu hỏi bằng ngôn ngữ hàng ngày và nhận được câu trả lời chính xác từ cơ sở dữ liệu. Nguyên lý hoạt động của hệ thống hỏi đáp như sau: Câu hỏi của người dùng sẽ được hệ thống gửi đến mô hình LLM để phân tích ngữ nghĩa và cú pháp của câu hỏi để hiểu rõ nội dung yêu cầu. Sau đó, mô hình sẽ chuyển đổi những yêu cầu được nhận biết, cùng với những hiểu biết về cơ sở dữ liệu, thành một câu lệnh truy vấn SQL mà hệ thống có thể thực hiện trên cơ sở dữ liệu. Câu lệnh truy vấn này sẽ được hệ thống gửi tới cơ sở dữ liệu để thực thi câu lệnh trên cơ sở dữ liệu và trả lại dữ liệu liên quan. Kết quả truy vấn (dạng dữ liệu, bảng biểu) một lần nữa được gửi qua LLM để chuyển thành ngôn ngữ tự nhiên, được trình bày dưới dạng văn bản, bảng biểu hoặc biểu đồ, giúp người dùng dễ dàng tiếp nhận thông tin (Hình 6).

Thách thức lớn nhất trong việc áp dụng mô hình ngôn ngữ lớn với cơ sở dữ liệu địa vật lý giếng khoan là yêu cầu bảo mật. Các mô hình có chất lượng tốt, có khả năng thấu hiểu người dùng và tạo ra câu lệnh truy vấn sát nhất đều sử dụng công nghệ đám mây, ngược lại, các mô hình có khả năng triển khai trên máy nội bộ, đảm bảo bảo mật lại cho chất lượng truy vấn thấp hơn, thậm chí trong một số trường hợp không hiểu được nhu cầu người dùng, hoặc diễn đạt không chính xác ý nghĩa của kết quả truy vấn. Để giải quyết vấn đề này, nhóm tác giả đã thiết kế và triển



Hình 5. *Ứng dụng cơ sở dữ liệu kết nối với các API khác. Nguồn: VPI, 2024.*

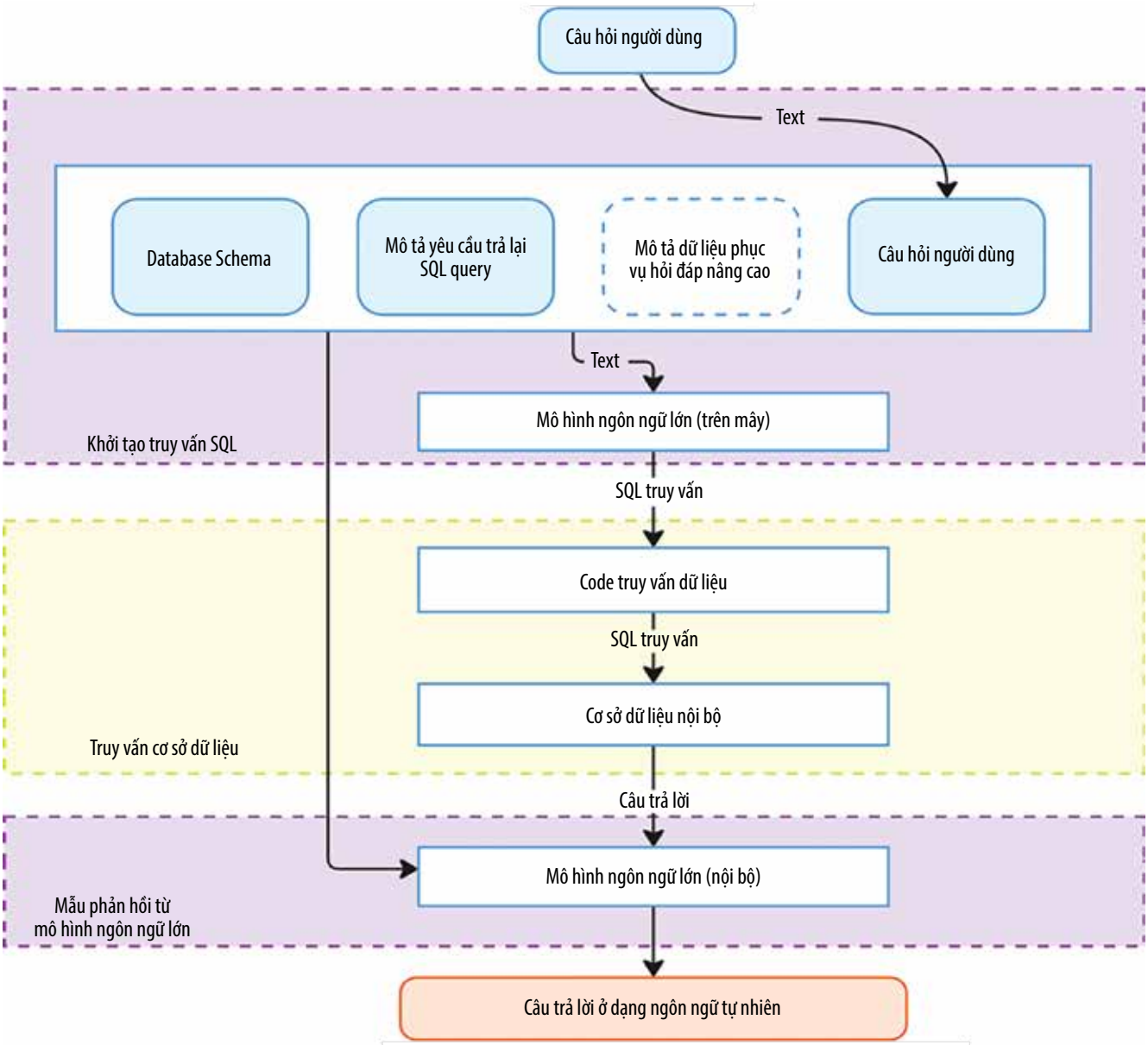


Hình 6. *Quy trình hoạt động sử dụng mô hình ngôn ngữ lớn hỏi đáp với cơ sở dữ liệu. Nguồn: VPI, 2024.*

khai phương án kết hợp giữa mô hình AI sử dụng cloud (mây) và mô hình trên máy nội bộ để đảm bảo chất lượng tốt và AI local để bảo mật thông tin.

Chiến lược kết hợp giữa mô hình AI sử dụng cloud và mô hình trên máy nội bộ được triển khai như sau: Đối với tác vụ tạo ra câu hỏi truy vấn, do đây là tác vụ cần chất lượng mô hình tốt, nhưng không yêu cầu đưa dữ liệu trực tiếp vào mô hình, nhóm tác giả sử dụng AI trên nền tảng đám mây của OpenAI, kết hợp với kỹ thuật xây dựng câu lệnh (prompt engineering). Kết quả cho thấy câu lệnh truy vấn rất tốt và sát với nhu cầu người dùng. Đối với tác vụ minh giải kết quả truy vấn, do tác vụ yêu cầu đưa dữ liệu trực tiếp vào mô hình để diễn giải, nhóm tác giả sử dụng mô hình LLM trên máy nội bộ với kích thước và số lượng tham số vừa đủ để đảm bảo hiệu suất hoạt động và bảo mật dữ liệu. Vấn đề về chất lượng diễn giải câu trả lời cũng được khắc phục thông qua kỹ thuật xử lý câu lệnh (Hình 7).

Câu lệnh (prompt) là đoạn văn bản hoặc thông tin đầu vào cung cấp cho mô hình. Vai trò của nó là "gợi ý" hoặc định hướng cho mô hình biết nội dung cần triển khai, các yêu cầu cụ thể, hoặc ngữ cảnh liên quan đến vấn đề. Do đó, việc đưa ra câu lệnh cụ thể và phản ánh rõ được nhu cầu của người dùng rất quan trọng và ảnh hưởng trực tiếp đến chất lượng câu trả lời của mô hình. Prompt engineering (kỹ thuật xây dựng câu lệnh) là quá trình thiết kế, tinh chỉnh và tối ưu prompt để có được



Hình 7. *Quy trình xử lý câu hỏi truy vấn dữ liệu kết hợp giữa AI cloud (trên mây) và AI local (nội bộ). Nguồn: VPI, 2024.*

câu trả lời mong muốn từ mô hình ngôn ngữ. Các tác vụ prompt engineering thường sẽ bao gồm lựa chọn ngôn từ, cú pháp, cung cấp ngữ cảnh chi tiết, quy ước về các tác vụ, quy ước về định dạng đầu ra, hoặc các tác vụ nâng cao khác. Việc lựa chọn câu lệnh phù hợp đòi hỏi sự tinh chỉnh, thử nghiệm lặp lại và sự "thấu hiểu" từng mô hình ngôn ngữ sử dụng của người xây dựng câu lệnh. Trong trường hợp này, do đặc thù dữ liệu chuyên môn nên việc để mô hình tự hiểu về cấu trúc cũng như ý nghĩa của dữ liệu trong cơ sở dữ liệu là không khả thi. Do đó, đối với tác vụ tạo câu lệnh truy vấn SQL, nhóm tác giả đã cung cấp cho mô hình chi tiết về cấu trúc của cơ sở dữ liệu (database schema không bao gồm dữ liệu) để đảm bảo mô hình biết được phạm vi của cơ sở dữ liệu và mối quan hệ giữa các bảng dữ liệu trong cơ sở dữ liệu. Bên cạnh đó, nhóm tác giả đã tận dụng từ điển dữ liệu đã được khai báo trước đó trong cơ sở dữ liệu (theo quy định về dữ liệu số của VPI) để mô hình ngôn ngữ hiểu về ý nghĩa của các bảng và các trường dữ liệu, đồng thời có thông tin về loại dữ liệu, tần suất, nguồn dữ liệu… từ đó đưa ra câu lệnh truy vấn chính xác hơn. Đối với tác vụ diễn giải câu trả lời, bên cạnh dữ liệu kết quả, trong câu lệnh gửi tới mô hình trên máy nội bộ, nhóm tác giả đã đưa kèm câu hỏi của người dùng và toàn bộ thông tin của cơ sở dữ liệu, cùng với chỉ dẫn cụ thể cho các trường hợp thường xuyên được hỏi liên quan đến chuyên môn để mô hình đưa ra câu trả lời sát nhất với nhu cầu và tuân thủ định dạng yêu cầu.

Ví dụ, khi thống kê giếng khoan theo lô (Hình 8), câu hỏi của người dùng được đưa vào hệ thống và gửi cùng với câu lệnh để đưa vào LLM trên đám mây để phân tích

câu hỏi và tạo ra câu lệnh truy vấn kèm theo định dạng hiển thị mong muốn. Trong trường hợp này, nội dung của LLM tạo ra như Hình 8.

Bên cạnh câu lệnh SQL phù hợp với cấu trúc dữ liệu trong cơ sở dữ liệu, LLM tự định dạng loại đồ thị phù hợp với phương thức thống kê này là đồ thị hình cột, đồng thời tự động đặt tên và mô tả cho hiển thị. Các thông tin này sau đó được bóc tách câu lệnh để gửi vào hệ thống cơ sở dữ liệu và nhận lại dữ liệu tương ứng, sau đó dữ liệu này dùng phần định nghĩa về hình vẽ và các mô tả khác sẽ được gửi tiếp vào hệ LLM trên máy nội bộ để đưa ra kết quả hiển thị cuối cùng ở dạng câu trả lời và hình vẽ như Hình 9.

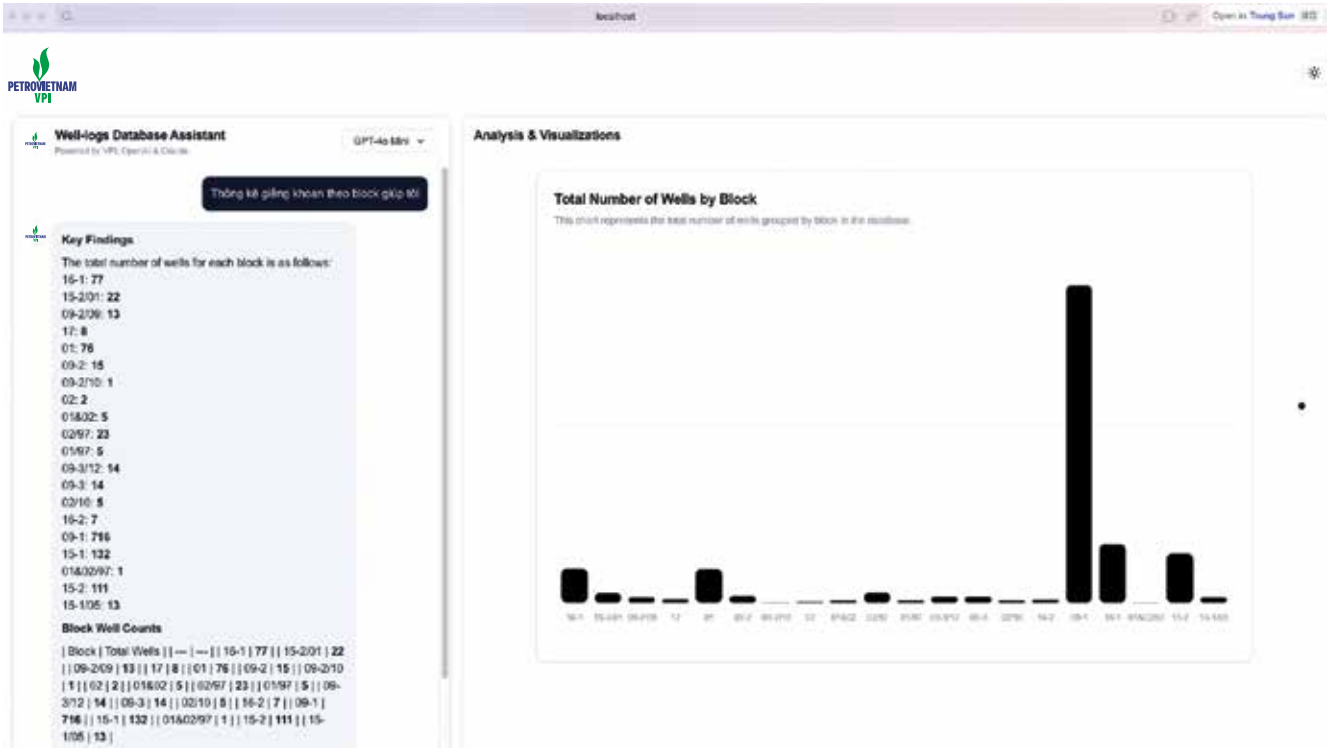
Kinh nghiệm ứng dụng kết hợp giữa AI trên máy nội bộ và trên mây có thể ứng dụng được trong rất nhiều trường hợp. Việc phân tách dựa trên yêu cầu bảo mật dữ liệu là nguyên tắc có thể áp dụng cho các dữ liệu chuyên ngành cần yêu cầu bảo mật cao nhưng vẫn đòi hỏi chất lượng truy vấn chính xác. Trong một số trường hợp, thay vì sử dụng mô hình trên máy nội bộ, người dùng có thể cân nhắc sử dụng AI trên mây để tạo ra mẫu câu trả lời và ghép dữ liệu một cách cơ học vào câu trả lời đưa tới người dùng. Việc này cũng đảm bảo dữ liệu không bị rò rỉ trên mây thông qua mô hình. Ngoài ra, việc phân tách cũng có thể dựa trên nguyên tắc yêu cầu về chất lượng và chi phí. Ví dụ những câu hỏi đơn giản hoặc yêu cầu độ chính xác không quá cao có thể xử lý trên máy nội bộ để tiết kiệm chi phí, những câu hỏi phức tạp hơn, yêu cầu chất lượng cao hơn sẽ được xử lý với các mô hình trên đám mây.

```
{

Chart Type: bar

Table Name: wellinfo

Title: Total Number of Wells by Block

Description: This chart represents the total number of wells grouped by block in
the database.

SQL Query:

SELECT COUNT("WELL_UWI") AS "count", "BLOCK"

FROM "wellinfo"

GROUP BY "BLOCK";

}
```

Hình 8. *Mẫu câu lệnh truy vấn cơ sở dữ liệu tạo ra từ mô hình ngôn ngữ lớn.*

4. Kết luận

Phần mềm quản lý cơ sở dữ liệu địa vật lý giếng khoan bể Cửu Long có tính ứng dụng cao và phù hợp với nhiều sản phẩm số tương tự, đặc biệt là các sản phẩm yêu cầu bảo mật dữ liệu. Với tốc độ phát triển công nghệ, chất lượng phản hồi của chatbot đối với dữ liệu chuyên môn sẽ ngày càng được cải thiện, ngay cả khi sử dụng các mô hình trên máy nội bộ. Điều này có thể đạt được thông qua việc đào tạo và tinh chỉnh các mô hình chuyên biệt dựa trên dữ liệu đầu vào từ các lĩnh vực chuyên ngành như: mô hình embedding



Hình 9. *Giao diện hỏi đáp với cơ sở dữ liệu địa vật lý giếng khoan bằng ngôn ngữ tự nhiên. Nguồn: VPI, 2024.*

(xử lý tài liệu đầu vào và xử lý câu hỏi người dùng), hay mô hình dịch thuật tài liệu chuyên ngành…

Đối với các mô hình AI/ML dựa trên dữ liệu có cấu trúc, việc thiết lập hệ thống chỉ số đo lường tương đối đơn giản và có tính nhất quán cao. Tuy nhiên, đối với các ứng dụng sử dụng ngôn ngữ tự nhiên như chatbot, việc đo lường chính xác câu trả lời trở nên khó khăn hơn, đặc biệt là do mô hình có thể thiếu hiểu biết về vấn đề, hoặc gặp phải hiện tượng ảo giác (hallucination). Để cải thiện, cần có sự can thiệp chuyên môn và hệ thống bộ câu hỏi thử nghiệm tiêu chuẩn để đánh giá hiệu quả. Ngoài ra, việc cập nhật tri thức và nâng cao chất lượng câu trả lời cần phải được cải tiến liên tục theo thời gian. Ưu điểm rõ rệt nhất là việc giảm thiểu thời gian người dùng tiếp cận và khai thác dữ liệu, nâng cao trải nghiệm người dùng, xóa bỏ rào cản công nghệ, giúp người dùng dễ dàng hơn trong việc sử dụng và tương tác với hệ thống.

Tài liệu tham khảo

[1] Ganesh Shank Gowekar, "How oil and gas industry are transforming with AI and ML", *World Journal of Advanced Research and Reviews*, Volume 23, Issue 3, pp. 1234 - 1238, 2024. DOI: 10.30574/wjarr.2024.23.3.2722.

[2] Chuka Anthony Arinze, Vincent Onuegbu Izionworu, Daniel Isong, Dominic Daudu, and Adedayo Adefemi "Integrating artificial intelligence into engineering processes for improved efficiency and safety in oil and gas operations", *Open Access Research Journal of Engineering and Technology*, Volume 6, Issue 1, pp. 39 - 51, 2024. DOI: 10.53022/oarjet.2024.6.1.0012.

[3] Siddhartha Nuthakki, Chinmay Shripad Kulkarni, Suraj Kumar, Satish Kathiriya, and Yudhisthir Nuthakki, "Artificial intelligence applications in natural gas industry: A literature review", *International Journal of Engineering and Advanced Technology*, Volume 13, Issue 3, 2024. DOI: 10.35940/ijeat.C4383.13030224.

[4] Muoi Duy Nguyen, Hoa Minh Nguyen, and Ngan Thi Bui, "Application of artificial neural network and seismic attributes to predict the distribution of Late Oligocene sandstones in the Cuu Long basin", *Journal of Mining and Earth Sciences*, Volume 64, Issue 3, pp: 24 - 31, 2023. DOI: 10.46326/jmes.2023.64(3).03.

[5] I.K. Turgazinov and T.M. Nassigazin, "Application of artificial intelligence for construction of well IPR", *Neft'i Gaz*, Volume 6, Issue 13, pp. 136 - 150, 2023. DOI: 10.37878/2708-0080/2023-6.13.

[6] Vivian Ofure Eghaghe, Olajide Soji Osundare, Chikezie Paul-Mikki Ewim, and Ifeanyi Chukwunonso Okeke, "Navigating the ethical and governance challenges of AI deployment in AML practices within the financial industry", *International Journal of Scholarly Research and Reviews*, Volume 5, Issue 2, pp. 30 - 51, 2024. DOI: 10.56781/ijsrr.2024.5.2.0047.

[7] PPDM Association, *"PPDM data model version 3.2. PPDM Standard, 3.2.1"*, 2020.

AI-INTEGRATED DOMAIN-SPECIFIC DATA MANAGEMENT: EXPERIENCE FROM DEVELOPING WELL-LOG DATABASE MANAGEMENT SOFTWARE IN THE CUU LONG BASIN

Vu Tuyet Vy, Nguyen Trung Son
Vietnam Petroleum Institute (VPI)
Email: vyvt@vpi.pvn.vn

Summary

The rapid advancement of artificial intelligence (AI), particularly Large Language Models, has significantly transformed data management and processing practices for both structured and unstructured data. This paper presents the experience of designing, implementing, and operating an AI-integrated geophysical well log database software for the Cuu Long basin at the Vietnam Petroleum Institute (VPI). The software was developed by leveraging AI to optimize the management, retrieval, and analysis of well logging data while ensuring strict security requirements.

The research results show that the product can be integrated with other supporting tools, contributing to an enhanced user experience. Based on these findings, future development directions are proposed to optimize the system's efficiency and security.

Key words: Artificial intelligence, large language models, database, well log geophysics, Cuu Long basin.

DầuKhí
KHOA HỌC, CÔNG NGHỆ VÀ ĐỔI MỚI SÁNG TẠO PETROVIETNAM

ỨNG DỤNG PYTHON VÀ MACHINE LEARNING TRONG XỬ LÝ VÀ PHÂN LOẠI DỮ LIỆU XUẤT NHẬP KHẨU

Đỗ Hồng Hạnh[1], Đoàn Tiến Quyết[1], Đoàn Trọng Sinh[2], Nguyễn Bằng Linh[1]
[1]Viện Dầu khí Việt Nam (VPI)
[2]Tập đoàn Công nghiệp - Năng lượng Quốc gia Việt Nam (PVN)
Email: hanhdh@vpi.pvn.vn
https://doi.org/10.47800/PVSI.2025.03-05

Tóm tắt

Dữ liệu xuất nhập khẩu tại Việt Nam đang gia tăng cả về quy mô và độ phức tạp, tạo ra thách thức lớn trong việc chuẩn hóa và phân loại thông tin khai báo hải quan. Bài viết đề xuất quy trình xử lý dữ liệu tự động hóa sử dụng ngôn ngữ lập trình Python, nhằm nâng cao hiệu quả và tính nhất quán trong phân tích thông tin khai báo hải quan. Tập dữ liệu đầu vào gồm hơn 10.000 bản ghi xuất nhập khẩu thực tế, được xử lý qua các bước kỹ thuật như: chuẩn hóa tên hàng, quy đổi đơn vị tính, tính toán chỉ tiêu định lượng và gán nhãn nhóm sản phẩm dựa trên từ khóa.

Kết quả cho thấy quy trình này vận hành hiệu quả trên tập dữ liệu có quy mô vừa và tính phức tạp cao, đồng thời đảm bảo độ chính xác và tính đồng nhất trong phân loại nhóm sản phẩm. Trên cơ sở đó, nhóm tác giả đề xuất tích hợp mô hình học máy như công cụ hỗ trợ để nâng cao khả năng khái quát hóa và thích ứng với các trường hợp ngoại lệ, khi tên hàng không được chuẩn hóa và thay đổi liên tục theo thực tiễn thương mại quốc tế.

Từ khóa: Xuất nhập khẩu, học máy, chuẩn hóa dữ liệu, TF-IDF, random forest, Python.

1. Giới thiệu

Dữ liệu xuất nhập khẩu đóng vai trò thiết yếu trong việc phân tích chuỗi cung ứng, theo dõi thương mại và quản lý nhà nước. Tuy nhiên, tính chất phi cấu trúc và không đồng nhất của dữ liệu đầu vào, đặc biệt ở trường "Tên hàng" hoặc "Mã HS khai báo", gây cản trở lớn cho việc khai thác thông tin. Việc xử lý thủ công hoặc bán thủ công bằng Excel thường tiềm ẩn sai sót, thiếu tính mở rộng và khó kiểm soát chất lượng. Do đó, yêu cầu đặt ra là cần xây dựng quy trình xử lý dữ liệu hiệu quả, linh hoạt và có khả năng tự động hóa cao. Trong nghiên cứu này, nhóm tác giả áp dụng các thư viện Python để xây dựng một chuỗi quy trình (pipeline) xử lý dữ liệu nhập khẩu nhựa từ file Excel được cung cấp bởi bên thứ ba. Các bước gồm làm sạch dữ liệu, ánh xạ, chuẩn hóa đơn vị, phân loại sản phẩm và tính toán các chỉ tiêu định lượng. Kết quả được so sánh với phương pháp truyền thống nhằm đánh giá hiệu quả xử lý và đề xuất khả năng mở rộng ứng dụng học máy (machine learning).

Ngày nhận bài: 12/6/2025
Ngày nhận bài sửa: 12 - 29/6/2025
Ngày chấp nhận đăng: 29/6/2025

Trong thực tế, dữ liệu khai báo xuất nhập khẩu tại Việt Nam chủ yếu được xử lý thủ công hoặc bán thủ công bằng phần mềm Excel. Mặc dù phổ biến nhờ tính trực quan và khả năng tiếp cận rộng rãi, nhưng công cụ này đang bộc lộ nhiều bất cập trong việc xử lý tập dữ liệu có quy mô vừa và tính phức tạp cao, không chuẩn hóa. Các nghiên cứu thực tiễn đã cho thấy các thao tác như chuẩn hóa tên hàng, phân loại từ khóa, tính toán định lượng và kiểm tra logic dữ liệu nếu thực hiện hoàn toàn trong Excel sẽ tiêu tốn nhiều thời gian, dễ xảy ra lỗi và thiếu tính mở rộng. Ngược lại, quy trình xử lý dữ liệu được thiết kế và thực thi bằng Python cho thấy ưu thế vượt trội trên nhiều khía cạnh: hiệu suất xử lý, độ chính xác, khả năng tái sử dụng.

Về hiệu suất xử lý, nghiên cứu của Kelvin và cộng sự [1] đã chỉ ra rằng Python có mức sử dụng CPU trung bình là 10,01%, thấp hơn so với 30,8% của Excel khi thực hiện các tác vụ cơ bản như đọc dữ liệu, áp dụng hàm logic và tạo báo cáo tổng hợp. Thử nghiệm này sử dụng kiểm định thống kê Wilcoxon để xác nhận sự khác biệt có ý nghĩa giữa 2 nền tảng. Điều này không chỉ phản ánh hiệu quả sử dụng tài nguyên hệ thống mà còn nhấn mạnh rằng Python vận hành ổn định hơn trong các tác vụ xử lý tập dữ liệu có quy

mô vừa và tính phức tạp cao, một đặc điểm thường xuyên xuất hiện trong phân tích dữ liệu xuất nhập khẩu.

Nghiên cứu của Mansour và cộng sự [2] thực hiện trên các mô hình dòng tiền đã minh chứng khả năng rút ngắn thời gian xử lý lên đến 99% khi chuyển từ Excel sang Python. Trong trường hợp cụ thể, mô hình dòng tiền được tính toán trong Excel mất 179 giây, trong khi phiên bản tương đương trong Python chỉ mất 0,8 giây, mà vẫn duy trì sai số tuyệt đối dưới 0,01%. Ngoài ra, Python còn cho phép triển khai các kỹ thuật tính toán hiện đại như vector hóa (trong Pandas) và lập trình mảng (NumPy), giúp tăng tốc xử lý đồng thời giữ được khả năng mở rộng với các dữ liệu phức tạp hơn.

Về độ chính xác và hiệu quả, nghiên cứu của Panko và Halverson [3] cho thấy trong các kiểm toán thực tế, tỷ lệ ô sai sót trong bảng tính thường dao động từ 1% đến 3%, có trường hợp thấp hơn (0,38%). Kết quả thí nghiệm của tác giả với Excel cũng ghi nhận rằng sinh viên làm 1 mình tạo ra bảng tính có tỷ lệ lỗi trung bình 4,6%, trong khi làm theo nhóm 3 người con số này giảm xuống còn 1,0%. Trong khi đó, các nền tảng lập trình như Python có thể tích hợp logic kiểm tra điều kiện, kiểm soát lỗi và tự động hóa xử lý, giúp giảm rủi ro sai sót phát sinh từ yếu tố con người. Quy trình Excel phụ thuộc nhiều vào thao tác thủ công nên dễ mắc lỗi định dạng, gán sai đơn vị, hoặc bỏ sót bước kiểm tra logic. Nghiên cứu của Panko [4] đã nhấn mạnh rằng gần như tất cả bảng tính đều chứa lỗi. Với tỷ lệ lỗi công thức ở mức 5%, khi 1 bảng tính có khoảng 100 công thức gốc thì xác suất tồn tại ít nhất 1 lỗi gần như đạt 100%. Bên cạnh đó, khả năng phát hiện lỗi của người dùng còn hạn chế: các nghiên cứu cho thấy cá nhân khi kiểm tra bảng tính chỉ phát hiện được khoảng 50 - 60% lỗi. Trong khi đó, môi trường sử dụng ngôn ngữ Python có thể tích hợp các bước chuẩn hóa, lọc ngoại lệ và kiểm tra tính nhất quán của dữ liệu ngay trong quy trình xử lý, giúp tỷ lệ dòng dữ liệu hợp lệ sau xử lý đạt trên 98%, cao hơn đáng kể so với khoảng 90 - 92% của phương pháp thủ công.

Về khả năng tái sử dụng và mở rộng, Excel thường yêu cầu người dùng chỉnh sửa lại toàn bộ công thức hoặc logic nếu danh mục sản phẩm thay đổi hoặc mở rộng. Panko [4] cảnh báo rằng khi công thức phải tham chiếu đến các ô nằm ngoài màn hình hoặc trên các trang tính khác, tỷ lệ lỗi phát sinh tăng đáng kể, đồng thời làm giảm khả năng duy trì và kiểm tra mô hình. Ngược lại, ngôn ngữ Python cho phép tổ chức từ khóa và logic xử lý thành các module riêng biệt (function, script, config), nhờ đó dễ dàng cập nhật và mở rộng hệ thống mà không cần can thiệp thủ công vào từng bước xử lý. Nghiên cứu của Ziogas và cộng sự [5] đã chứng minh rằng Python có thể đạt được hiệu suất cao và khả năng mở rộng tốt khi xử lý tập dữ liệu có quy mô vừa và tính phức tạp cao, đặc biệt khi sử dụng các công cụ như NumPy và các thư viện hỗ trợ tính toán song song. Điều này cho thấy Python không chỉ đơn giản và dễ học mà còn mạnh mẽ trong việc xử lý tập dữ liệu có quy mô vừa và tính phức tạp cao.

Nhìn chung, các bằng chứng định lượng và thực nghiệm đã chỉ ra rằng quy trình xử lý dữ liệu bằng Python không chỉ cải thiện đáng kể hiệu suất, mà còn nâng cao độ chính xác, khả năng kiểm soát chất lượng và khả năng thích ứng với các yêu cầu thay đổi trong thực tế. Các doanh nghiệp và cơ quan quản lý ngày càng phải xử lý lượng dữ liệu phức tạp và đa dạng, việc chuyển đổi từ Excel sang nền tảng xử lý tự động hóa như Python giúp nâng cao năng lực xử lý và phân tích chuyên sâu trên tập dữ liệu.

2. Quy trình xử lý dữ liệu bằng ngôn ngữ lập trình Python

2.1. Mô tả dữ liệu

Dữ liệu sử dụng trong nghiên cứu được trích xuất từ các bộ số liệu khai báo xuất nhập khẩu hàng tháng do cơ quan hải quan công bố hoặc được cung cấp bởi bên thứ 3. Đây là dữ liệu thực tế phản ánh hoạt động giao thương thường xuyên, được chuẩn hóa theo từng kỳ báo cáo, với quy mô mỗi tháng dao động trong khoảng từ 2.000 - 3.000 dòng dữ liệu, tập trung vào các mặt hàng nhựa polypropylene (PP). Mỗi dòng tương ứng với một tờ khai nhập khẩu gồm các trường thông tin như mã tờ khai, mã chi cục, loại hình nghiệp vụ, mã HS khai báo, tên hàng, số lượng, đơn vị tính và trị giá giao dịch. Quy trình xử lý dữ liệu được thiết kế với mục tiêu tự động hóa toàn bộ quá trình xử lý thông tin nhập khẩu nhựa PP từ file Excel đầu vào. Các bước gồm làm sạch dữ liệu, chuẩn hóa chuỗi văn bản, ánh xạ thông tin chi cục và loại hình, phân loại sản phẩm theo từ khóa kỹ thuật, quy đổi đơn vị và tính toán chỉ tiêu định lượng như sản lượng tính theo tấn, đơn giá và trị giá theo USD. Kết quả đầu ra có thể được sử dụng để thống kê, phân tích, trực quan hóa hoặc làm đầu vào cho các mô hình học máy nhằm hỗ trợ ra quyết định quản lý.

Điểm đáng chú ý là quy mô dữ liệu theo từng tháng không lớn, có tính chất đơn lẻ, chi phí triển khai và vận hành hệ thống hiện tại rất thấp, ngay cả khi áp dụng cho nhiều tháng hoặc nhiều sản phẩm khác nhau, hệ thống vẫn duy trì khả năng xử lý ổn định mà không đòi hỏi nâng cấp tài nguyên tính toán đáng kể. Do đó, quy trình xử lý này đặc

biệt phù hợp để áp dụng riêng cho từng nhóm mặt hàng xuất nhập khẩu theo mã HS hoặc theo từng sản phẩm, như nhóm nhựa PP trong nghiên cứu. Điều này mở ra tiềm năng triển khai linh hoạt theo từng ngành hàng, giúp tăng cường khả năng phân tích theo chiều sâu và hỗ trợ hoạt động giám sát chuyên đề trong lĩnh vực xuất nhập khẩu.

2.2. Chuẩn hóa toàn bộ dữ liệu văn bản

Quy trình xử lý dữ liệu xuất nhập khẩu được thiết kế và thực hiện trong môi trường Google Colab (Python 3.10.2), tối ưu hóa cho tập dữ liệu có quy mô vừa và tính phức tạp cao. Các trường văn bản như tên doanh nghiệp xuất nhập khẩu, tên hàng thường tồn tại dưới nhiều định dạng không đồng nhất, chứa ký tự đặc biệt. Để đảm bảo khả năng xử lý và phân tích tự động, tất cả các cột dạng văn bản được chuẩn hóa bằng hàm biến đổi tùy chỉnh chuyển về chữ in hoa, loại ký tự đặc biệt và chuẩn hóa khoảng trắng.

2.3. Kiểm tra dữ liệu

Dữ liệu được rà soát theo 3 tiêu chí: (i) Tính chính xác: Kiểm tra đơn vị và giá trị của các trường dữ liệu để loại bỏ các dòng bất hợp lệ; (ii) Tính nhất quán: Kiểm tra định dạng dữ liệu; (iii) Tính đầy đủ: Kiểm tra dữ liệu thiếu.

2.4. Trích xuất và ánh xạ mã chi cục, loại hình

Trong cấu trúc chuỗi "Tờ khai", thông tin quan trọng như mã chi cục hải quan và mã loại hình nghiệp vụ thường được nhúng dưới dạng ký hiệu kết hợp, ví dụ: 106977513900/C11/47I1. Để phục vụ mục đích phân tích theo vùng địa lý và mục đích nghiệp vụ, hệ thống thực hiện trích xuất có điều kiện từ chuỗi dạng này như sau:

- Mã chi cục được xác định là 4 ký tự cuối cùng trong chuỗi "Tờ khai" sau khi loại bỏ khoảng trắng và chuẩn hóa viết hoa.

- Mã loại hình được trích xuất là ký hiệu ở vị trí thứ hai sau dấu gạch chéo (/), thường theo định dạng chuẩn của Tổng cục Hải quan (ví dụ: C11, A12, B13...).

Quy trình trích xuất được chuẩn hóa bằng phép biến đổi chuỗi như sau:

Trích xuất mã chi cục: 4 ký tự cuối

df["Mã chi cục"] = df["Tờ khai"].str[-4:]

Trích xuất mã loại hình: phần tử thứ hai trong chuỗi tách bằng "/"

df["Mã loại hình"] = df["Tờ khai"].str.split("/").str[1]

Sau đó, mã chi cục được ánh xạ với bảng mã địa phương để gán nhãn miền (Bắc, Trung, Nam), còn mã loại hình được đối chiếu với bảng danh mục mã nghiệp vụ để phục vụ phân tích mục đích khai báo như kinh doanh, gia công, sản xuất xuất khẩu.

2.5. Phân loại nhóm sản phẩm

Việc phân loại tên hàng vào các nhóm sản phẩm như PP_HOMO, PP_COPOLYMER, PP_COMPOUND được thực hiện bằng phương pháp khớp từ khóa. Danh sách từ khóa đã chuẩn hóa được ánh xạ với từng nhãn phân loại. Hệ thống thực hiện quy trình này tự động dựa trên Regex hoặc tìm kiếm chuỗi, có khả năng xử lý toàn bộ dữ liệu nhanh chóng. Với những dòng dữ liệu chưa được khớp theo phương pháp ánh xạ do thiếu hụt danh sách từ khóa, tác giả tiếp tục phân loại bằng phương pháp học máy được giới thiệu trong nghiên cứu này. Trong thực tế triển khai, dù hệ thống phân loại bằng từ khóa (rule-based) và mô hình học máy đã cho thấy hiệu quả nhất định trong xử lý tên hàng nhập khẩu, vẫn tồn tại một tỷ lệ nhỏ các trường hợp không thể phân loại do không khớp với từ khóa hiện có hoặc chứa nội dung không rõ ràng. Để đảm bảo tính bao phủ và độ tin cậy của toàn bộ quy trình, nghiên cứu đề xuất tích hợp một cơ chế phản hồi có sự tham gia của chuyên gia (Human-in-the-loop - HITL) nhằm xử lý linh hoạt các ngoại lệ và cải thiện hệ thống theo thời gian. Cơ chế HITL được triển khai như một bước hậu kiểm, cụ thể như sau:

(i) Sau khi hệ thống thực hiện phân loại tên hàng bằng cả rule-based và mô hình học máy, các dòng không phân loại được (unclassified) hoặc có độ tin cậy thấp sẽ được tự động ghi log vào một bảng riêng (ví dụ: unclassified_items.xlsx) kèm theo các thông tin hỗ trợ như mô tả tên hàng, ngày khai báo, đơn vị nhập khẩu và kết quả phân loại gần đúng (nếu có);

(ii) Các dòng này sau đó được chuyển tới bộ phận chuyên môn để rà soát và gán nhãn thủ công dựa trên hiểu biết kỹ thuật và kinh nghiệm nghiệp vụ;

(iii) Kết quả rà soát được lưu lại và sử dụng để cập nhật vào các tập luật (keyword_map.json) hoặc bổ sung vào tập huấn luyện của mô hình học máy ở các chu kỳ sau, từ đó nâng cao độ chính xác và khả năng thích ứng của hệ thống.

Cơ chế đảm bảo rằng dữ liệu không rõ ràng hoặc chưa từng xuất hiện sẽ không bị loại bỏ, mà thay vào đó trở thành đầu vào cho quá trình cải tiến liên tục của hệ thống. Hơn nữa, việc lưu lại log đầy đủ giúp quá trình rà soát minh bạch, dễ kiểm tra và có thể đánh giá hiệu quả phân loại

theo thời gian. Đây là một bước quan trọng trong việc xây dựng quy trình xử lý dữ liệu linh hoạt và có khả năng học hỏi, một yêu cầu thiết yếu trong môi trường nghiệp vụ thực tế, nơi dữ liệu phát sinh liên tục với mức độ biến động cao

2.6. Quy đổi đơn vị và tính toán các chỉ tiêu định lượng

Tất cả dữ liệu về khối lượng được quy đổi về đơn vị chuẩn là tấn thông qua bảng chuyển đổi đơn vị. Bước này đảm bảo tính đồng nhất trong báo cáo sản lượng. Trên cơ sở đơn giá và sản lượng đã chuẩn hóa, hệ thống tự động tính ra trị giá USD và các chỉ tiêu thống kê cần thiết.

2.7. Kiểm tra chất lượng dữ liệu đầu ra

Dữ liệu được rà soát theo 3 tiêu chí:

- Loại bỏ các dòng trùng lặp hoàn toàn;

- Loại bỏ giá trị ngoại lai/không phù hợp (outliers) dựa theo thống kê và chuyên môn nghiệp vụ;

- Kiểm tra logic đơn vị và giá trị để loại bỏ các dòng bất hợp lệ.

3. Đánh giá hiệu quả của quy trình xử lý dữ liệu

Để kiểm chứng hiệu quả thực tiễn của quy trình xử lý dữ liệu đề xuất, nghiên cứu tiến hành thực nghiệm trên tập dữ liệu gồm 10.000 dòng dữ liệu xuất nhập khẩu phát sinh trong 1 tháng. Đây là khối lượng dữ liệu hải quan thực tế, phản ánh mật độ nghiệp vụ cao và tính chất biến động liên tục của hoạt động xuất nhập khẩu trong ngắn hạn. Mục tiêu của đánh giá là xác định mức độ cải thiện về hiệu suất xử lý, độ chính xác dữ liệu đầu ra và khả năng mở rộng hệ thống khi sử dụng Google Colab (Python 3.10.2).

3.1. Thời gian xử lý và hiệu suất vận hành

Kết quả đo lường thực tế (Bảng 1) cho thấy tổng thời gian xử lý toàn bộ quy trình đối với 10.000 dòng dữ liệu trong môi trường Google Colab (Python 3.10.2) là khoảng 303 giây (tương đương 5 phút 3 giây). Trong đó, các công đoạn như chuẩn hóa dòng và cột, quy đổi đơn vị và tính toán định lượng, kiểm tra lỗi logic và lọc dòng ngoại lệ đều được thực hiện gần như tức thời, chỉ mất khoảng 1 giây mỗi bước. Điều này đạt được nhờ khả năng xử lý song song và vector hóa mạnh mẽ của các thư viện Python như Pandas, Regex và NumPy, giúp thao tác nhanh chóng trên toàn bộ DataFrame. Ngược lại, khi thực hiện các công đoạn tương tự bằng Excel, thời gian xử lý tổng thể ước tính dao động trong khoảng 3 - 6 phút, tùy thuộc vào cấu hình máy tính và phiên bản Excel sử dụng. Đặc biệt, công đoạn phân loại tên hàng theo từ khóa, vốn đòi hỏi quét chuỗi và so khớp điều kiện, khi thực hiện bằng các hàm như SEARCH, IF, INDEX-MATCH hoặc TEXTJOIN trong Excel có thể gây chậm đáng kể khi áp dụng trên tập dữ liệu có quy mô vừa và tính phức tạp cao. Trong một số trường hợp, Excel còn xuất hiện tình trạng "Not responding" khi xử lý bảng tính có nhiều công thức mảng động. Bên cạnh yếu tố tốc độ, phương pháp xử lý bằng Python còn cho thấy tính ổn định và linh hoạt cao hơn. Cụ thể, môi trường Google Colab không gặp tình trạng treo ứng dụng như Excel và cho phép mở rộng thuật toán kiểm tra lỗi, lọc ngoại lệ hoặc thiết lập quy tắc phân loại phức tạp mà không bị giới hạn bởi giao diện. Ngoài ra, môi trường mã nguồn mở như Python còn hỗ trợ khả năng tự động hóa, kiểm thử và ghi log chi tiết, điều mà Excel chỉ có thể thực hiện khi tích hợp thêm VBA hoặc Power Query với độ phức tạp cao hơn. Kết quả cho thấy việc sử dụng Python trên Google Colab trong xử lý dữ liệu khai báo sản phẩm mang lại hiệu quả vượt trội về thời gian, tính linh hoạt và khả năng mở rộng, đặc biệt khi áp dụng cho các quy trình lặp lại hoặc tập dữ liệu có quy mô vừa và tính phức tạp cao.

3.2. Độ chính xác và tính nhất quán của dữ liệu đầu ra

Bên cạnh hiệu suất thời gian, nghiên cứu cũng đánh giá chất lượng đầu ra thông qua các chỉ số về tỷ lệ lỗi và dữ liệu không hợp lệ sau xử lý. Cụ thể, 3 loại lỗi được thống kê gồm: (i) không phân loại được tên hàng, (ii) lỗi đơn vị và quy đổi sai, và (iii) không tính toán được giá trị.

Kết quả cho thấy chất lượng đầu ra của quy trình xử lý dữ liệu bằng Google Colab (Python 3.10.2) đạt độ chính xác cao, với tổng tỷ lệ lỗi ghi nhận xấp xỉ bằng 0% trên tập dữ liệu thử nghiệm đã được chuẩn hóa gồm 10.000 dòng (Bảng 2). Cụ thể, 3 loại lỗi được quan sát đều không xảy ra

Bảng 1. *Thời gian xử lý trung bình bằng Google Colab (Python 3.10.2)*

Công đoạn xử lý	Google Colab (Python 3.10.2)
Chuẩn hóa dòng và cột	1 giây
Phân loại	300 giây
Quy đổi đơn vị và tính toán định lượng	1 giây
Kiểm tra lỗi logic và lọc dòng ngoại lệ	1 giây
Tổng thời gian xử lý	303 giây

trong quá trình xử lý gồm: không gán nhãn được các dòng tên hàng dù đã chứa từ khóa có trong bộ quy tắc, lỗi trong quá trình quy đổi đơn vị và sai số lượng, cũng như lỗi không thể tính toán giá trị do đầu vào không hợp lệ. Điều này cho thấy quy trình tiền xử lý như chuẩn hóa chuỗi văn bản, lọc ký tự đặc biệt, đồng nhất hóa đơn vị và kiểm tra tính hợp lệ của số liệu đã được thiết kế đầy đủ và hoạt động hiệu quả.

Việc sử dụng ngôn ngữ lập trình Python giúp đảm bảo tính minh bạch trong logic, khả năng kiểm soát lỗi rõ ràng và dễ dàng thiết lập các rào chắn kỹ thuật để ngăn chặn dữ liệu sai sót ngay từ bước đầu vào. Tuy nhiên, cần nhấn mạnh rằng kết quả này có được trong điều kiện dữ liệu đã được kiểm tra sơ bộ và chuẩn hóa tương đối tốt trước khi đưa vào hệ thống xử lý tự động. Trong môi trường thực tế, khi dữ liệu đầu vào đến từ nhiều nguồn khác nhau với mức độ "nhiễu" cao, việc duy trì tỷ lệ lỗi bằng 0% có thể khó đạt được nếu không có cơ chế kiểm tra và log lỗi đầy đủ. Do đó, nghiên cứu cũng đề xuất rằng trong các lần triển khai thực tế, cần tích hợp thêm các bước đánh dấu dòng nghi ngờ, phân loại lỗi và xác nhận thủ công đối với các trường hợp không rõ ràng, nhằm tăng cường độ tin cậy và tính linh hoạt của hệ thống. Kết quả về tỷ lệ lỗi cho thấy hệ thống xử lý bằng môi trường Google Colab (Python 3.10.2) không chỉ đạt hiệu quả về thời gian mà còn đảm bảo được tính chính xác và độ sạch của dữ liệu sau xử lý, là tiền đề quan trọng để thực hiện các bước phân tích tiếp theo như thống kê, mô hình hóa hoặc trực quan hóa thông tin.

Để kiểm chứng vai trò của bước chuẩn hóa trong quy trình xử lý, nghiên cứu đã tiến hành thực nghiệm đối chứng trên tập dữ liệu đầu vào ở trạng thái ban đầu, tức dữ liệu chưa trải qua bất kỳ bước xử lý sơ bộ nào như chuyển chữ in hoa, loại bỏ ký tự đặc biệt hay chuẩn hóa khoảng trắng. Trong trạng thái này, các chuỗi văn bản mô tả tên hàng và tên doanh nghiệp xuất nhập khẩu tồn tại nhiều định dạng khác nhau, bao gồm chữ hoa, chữ thường lẫn lộn, ký hiệu kỹ thuật không thống nhất, dấu câu xen lẫn và khoảng trắng thừa. Đây là đặc điểm

Bảng 2. *Tỷ lệ lỗi trung bình trong dữ liệu sau xử lý*

Loại lỗi	Google Colab (Python 3.10.2) (%)
Không gán nhãn được các dòng tên hàng dù đã chứa từ khóa có trong bộ quy tắc	0
Gán sai đơn vị hoặc quy đổi không chính xác	0
Không thể tính toán giá trị (lỗi đầu vào)	0
Tổng tỷ lệ dòng lỗi	0

Bảng 3. *Tỷ lệ lỗi của dữ liệu khi chưa được chuẩn hóa*

Tiêu chí	Tập dữ liệu chuẩn hóa	Tập dữ liệu chưa chuẩn hóa
Số dòng phân loại sai	0	610
Số dòng không phát hiện trùng lặp	0	280
Tỷ lệ lỗi tổng hợp (%)	0	8,9

phổ biến trong dữ liệu nhập khẩu thực tế, khi doanh nghiệp khai báo không theo chuẩn định dạng thống nhất. Kết quả thử nghiệm trên 10.000 dòng dữ liệu cho thấy hệ thống vẫn có thể xử lý được dữ liệu thô (raw), tuy nhiên độ chính xác giảm mạnh. Cụ thể, có 610 dòng không được phân loại chính xác theo nhóm sản phẩm tương ứng và 280 dòng bị bỏ sót do không nhận diện được trùng lặp, tổng cộng 890 dòng lỗi, tương đương với tỷ lệ lỗi 8,9%. Nguyên nhân là do việc mô tả tên hàng trong trạng thái raw không đảm bảo tính nhất quán cú pháp, khiến hệ thống rule-based không thể khớp đúng từ khóa kỹ thuật. Đồng thời, sự thiếu thống nhất trong cách viết tên doanh nghiệp (viết tắt, chữ thường, khoảng trắng thừa) khiến các dòng dữ liệu hoàn toàn trùng khớp về thông tin giao dịch nhưng khác định dạng bị hệ thống hiểu thành các bản ghi riêng biệt, không lọc trùng được. Điều này dẫn đến sai lệch thống kê về số lượng giao dịch, sản lượng và giá trị nhập khẩu.

Thử nghiệm này khẳng định rằng bước chuẩn hóa không chỉ là thao tác kỹ thuật phụ trợ mà là bước bắt buộc trong quy trình xử lý dữ liệu, ảnh hưởng trực tiếp đến khả năng phân loại chính xác, nhận diện trùng lặp và tính nhất quán của kết quả phân tích. Việc bỏ qua bước chuẩn hóa dẫn đến sai sót tích lũy trong toàn bộ quy trình, đặc biệt ở các khâu thống kê định lượng và lọc dữ liệu đầu vào. Như vậy, thử nghiệm trên tập raw không những phản ánh giới hạn của hệ thống khi thiếu chuẩn hóa, mà còn làm nổi bật vai trò thiết yếu của quy trình tiền xử lý trong việc đảm bảo hiệu quả, độ tin cậy và tính tự động hóa cao của quy trình xử lý dữ liệu xuất nhập khẩu.

3.3. Khả năng mở rộng và bảo trì hệ thống

Dưới góc độ kỹ thuật và thực tiễn triển khai, việc sử dụng ngôn ngữ lập trình Python trong xây dựng hệ thống xử lý dữ liệu xuất nhập khẩu mang lại lợi thế vượt trội về khả năng mở rộng và bảo trì so với các phương pháp truyền thống như Microsoft Excel. Trong khi Excel phụ thuộc chủ yếu vào công

thức cố định và thao tác thủ công, dễ gây lỗi lan truyền khi mở rộng quy mô hoặc cập nhật danh mục sản phẩm mới, thì hệ thống Python cho phép tổ chức kiến trúc mở, tách biệt rõ ràng giữa mã xử lý và các tệp cấu hình (ví dụ: keyword_map.json, unit_map.csv). Điều này cho phép người dùng chỉ cần cập nhật các file ánh xạ mà không phải sửa đổi mã nguồn chính, từ đó nâng cao tính linh hoạt, khả năng kiểm soát và khả năng thích ứng với các yêu cầu nghiệp vụ thay đổi liên tục. Không chỉ dừng lại ở tính tiện lợi, nghiên cứu [6] đã chỉ ra rằng quản lý cấu hình tách biệt là một trong những thực hành tốt nhất trong các dự án xử lý dữ liệu bằng Python, giúp giảm rủi ro vận hành và tăng khả năng mở rộng hệ thống khi xử lý các luồng dữ liệu phức tạp trong môi trường doanh nghiệp. Bên cạnh đó, Python cũng được ghi nhận là ngôn ngữ có khả năng tích hợp mạnh mẽ với các nền tảng đám mây như Google Colab, AWS, Azure - điều mà các nền tảng như Excel không thể đáp ứng ở quy mô lớn. Theo Ayyagiri và cộng sự [7], các công cụ xử lý song song như Dask và PySpark trong hệ sinh thái Python có thể tăng tốc độ xử lý lên đáng kể và giảm tải cho máy cá nhân, đặc biệt hữu ích khi xử lý dữ liệu nhập khẩu ngày càng mở rộng về quy mô và độ phức tạp.

Kết quả trên cho thấy quy trình xử lý dữ liệu xuất nhập khẩu sử dụng ngôn ngữ lập trình Python triển khai trên nền tảng Google Colab đã mang lại kết quả tích cực đặc biệt về hiệu suất thời gian, độ chính xác và khả năng mở rộng, bảo trì hệ thống, cho thấy tính khả thi của ứng dụng trong thực tiễn. Về hiệu suất, toàn bộ quy trình xử lý 10.000 dòng dữ liệu chỉ mất khoảng 303 giây (tương đương hơn 5 phút). Kết quả này đạt được nhờ vào việc khai thác khả năng xử lý theo vector và song song của các thư viện mã nguồn mở như Pandas, NumPy và Regex. Các bước như chuẩn hóa chuỗi văn bản, tính toán định lượng, và kiểm tra điều kiện logic đều được thực hiện gần như tức thì. Trong khi đó, Excel tuy phổ biến và trực quan thường gặp khó khăn khi áp dụng vào các tác vụ xử lý chuỗi phức tạp hoặc tập dữ liệu có quy mô vừa và tính phức tạp cao. Việc sử dụng công thức lồng nhau hoặc công thức mảng có thể gây chậm trễ rõ rệt, thậm chí dẫn đến tình trạng "treo" ứng dụng khi dữ liệu vượt ngưỡng xử lý hiệu quả (thường từ 5.000 - 10.000 dòng trở lên, tùy cấu hình máy tính và phiên bản Excel). Về độ chính xác, hệ thống xử lý bằng Python cho thấy kết quả đầu ra có độ tin cậy rất cao, với tỷ lệ lỗi sau xử lý bằng 0% trong thử nghiệm. Thành công này phản ánh hiệu quả của kiến trúc pipeline tách biệt rõ từng khâu: từ chuẩn hóa dữ liệu, kiểm tra logic đến phân loại bằng từ khóa. Nhờ khả năng kiểm soát logic lập trình chặt chẽ, việc truy vết và xử lý các tình huống ngoại lệ trở nên rõ ràng, minh bạch hơn so với Excel, nơi các thao tác chủ yếu dựa vào thủ công hoặc công thức cố định, ít khả năng kiểm tra lỗi tự động và khó truy xuất nguyên nhân khi có sai sót xảy ra. Đặc biệt, phương pháp xử lý bằng Python thể hiện ưu thế vượt trội về khả năng mở rộng và bảo trì hệ thống. Các tệp cấu hình riêng biệt (như keyword_map.json cho từ khóa hoặc unit_map.csv cho đơn vị đo) giúp hệ thống dễ dàng cập nhật theo thay đổi thực tiễn mà không cần can thiệp vào mã nguồn cốt lõi. Khi xuất hiện sản phẩm mới, biến thể tên thương mại, hay yêu cầu phân tích theo khu vực địa lý, người dùng chỉ cần cập nhật file cấu hình thay vì sửa hàng loạt công thức trong bảng tính như với Excel. Điều này giúp giảm thiểu rủi ro lỗi hệ thống, đồng thời tăng tính linh hoạt và khả năng thích ứng với dữ liệu mới trong môi trường nghiệp vụ biến động nhanh.

Kết quả đánh giá cho thấy việc triển khai quy trình xử lý dữ liệu bằng Python trên Google Colab không chỉ mang lại lợi thế rõ ràng về hiệu suất và độ chính xác, mà còn đặt nền móng cho một hệ thống bền vững, có khả năng tự động hóa, bảo trì dễ dàng và thích ứng cao với thay đổi thực tế.

4. Ứng dụng học máy trong phân loại dữ liệu xuất nhập khẩu

Khi xử lý dữ liệu khai báo hải quan, nhóm nghiên cứu đang sử dụng phương pháp phân loại tên hàng dựa trên quy tắc từ khóa (rule-based keyword matching) nhờ tính đơn giản và khả năng kiểm soát trực tiếp. Theo Sebastiani [8], các hệ thống phân loại dựa trên quy tắc (rule-based) thường phát huy hiệu quả đối với các trường hợp có quy chuẩn từ ngữ rõ ràng và ổn định, cho phép ánh xạ trực tiếp văn bản đầu vào sang các nhóm sản phẩm nhờ các quy tắc xác định sẵn. Trên thực tế, phương pháp này hoạt động hiệu quả đối với các mô tả sản phẩm phổ biến, có thể tra cứu được thông tin kỹ thuật trên các nguồn dữ liệu công khai và đã được chuẩn hóa trước đó. Tuy nhiên, theo thời gian, dữ liệu thực tế phát sinh liên tục khiến phương pháp rule-based dần bộc lộ những hạn chế cố hữu. Trung bình mỗi tháng, nhóm nghiên cứu ghi nhận khoảng 400 dòng tên hàng mới không thể phân loại theo cách truyền thống. Nguyên nhân chủ yếu đến từ việc các mô tả này không khớp hoàn toàn với tập từ khóa hiện có, sử dụng cách diễn đạt khác biệt, viết tắt theo nội bộ doanh nghiệp, hoặc không mang đủ thông tin kỹ thuật để xác định chính xác. Việc mở rộng hệ thống phân loại rule-based trong trường hợp này không chỉ mất thời gian, mà còn đòi hỏi sự can thiệp thủ công của chuyên gia, làm giảm tính linh hoạt và khả năng thích nghi của hệ thống [8].

Để khắc phục tình trạng trên, nhóm nghiên cứu đề xuất tích hợp module học máy như một công cụ bổ trợ, nhằm tự động gợi ý nhóm phân loại đối với các tên hàng chưa từng xuất hiện. Theo hướng tiếp cận này, mô hình học máy được huấn luyện trên tập dữ liệu có gán nhãn từ trước, sử dụng biểu diễn văn bản bằng kỹ thuật TF-IDF, một phương pháp vẫn được xem là hiệu quả và dễ triển khai trong các tác vụ phân loại văn bản ngắn, văn bản không có cấu trúc ngữ pháp rõ ràng [9]. Biểu diễn này cho phép mô hình học các tín hiệu ngôn ngữ tiềm ẩn từ văn bản gốc mà không cần xây dựng đặc trưng thủ công. Để thực hiện bài toán phân loại, thuật toán random forest được lựa chọn do tính ổn định và hiệu suất cao trong môi trường dữ liệu nhiều chiều và không tuyến tính - vốn là đặc trưng của tên hàng hóa nhập khẩu. Nghiên cứu của Fernández-Delgado và cộng sự [10] cho thấy random forest nằm trong nhóm mô hình hiệu quả nhất trong hơn 100 thuật toán phân loại được đánh giá trên các tập dữ liệu thực tế, đạt độ chính xác trung bình 94,1%. Khi dữ liệu khai báo hải quan thường có đặc trưng phi cấu trúc, không chuẩn hóa và thay đổi liên tục, các phương pháp phân loại thuần rule-based gặp nhiều hạn chế về khả năng bao phủ và tính thích ứng. Các nghiên cứu gần đây như Aubaid và cộng sự [11] đã chỉ ra rằng các hệ thống phân loại lai (hybrid) kết hợp giữa rule-based và học máy, chẳng hạn như random forest, decision tree hoặc các mô hình embedding, có thể cải thiện đáng kể hiệu suất phân loại, đặc biệt trong trường hợp thiếu dữ liệu nhãn hoặc không thể xác lập quy tắc thủ công rõ ràng. Bằng cách sử dụng kiến thức chuyên gia và khả năng học đặc trưng từ dữ liệu (trong AI-based), mô hình hybrid cho thấy ưu thế vượt trội về độ chính xác và khả năng mở rộng trong phân loại văn bản ngắn như mô tả hàng hóa trong khai báo hải quan. Các ứng dụng trong lĩnh vực hải quan như Singh và cộng sự [12] chỉ ra tiềm năng của học máy trong việc xử lý dữ liệu khai báo bất cấu trúc và phát hiện bất thường.

Khi hệ thống phân loại dựa trên từ khóa truyền thống dần bộc lộ những giới hạn về khả năng bao phủ và tính mở rộng, việc tích hợp mô hình học máy như 1 module bổ trợ mang lại triển vọng cải thiện đáng kể hiệu quả vận hành. Phương pháp tiếp cận dựa trên học máy không nhằm thay thế hoàn toàn quy trình hiện tại, mà đóng vai trò hỗ trợ việc gợi ý phân loại đối với các dòng tên hàng chưa từng xuất hiện hoặc khó quy chuẩn hóa bằng luật cố định. Các phần tiếp theo sẽ trình bày chi tiết quy trình xây dựng, huấn luyện và đánh giá mô hình phân loại này, với trọng tâm là tính khả thi và khả năng ứng dụng thực tế trong môi trường xử lý dữ liệu.

4.1. Dữ liệu và tiền xử lý

Nghiên cứu sử dụng tập dữ liệu gồm 8.236 dòng tên hàng nhập khẩu, trong đó mỗi dòng chứa 2 trường thông tin chính: (i) TEN_HANG là mô tả sản phẩm được ghi dưới dạng chuỗi văn bản tự do trong tờ khai hải quan và (ii) PHAN_LOAI là nhãn phân loại tương ứng được gán thủ công theo các nhóm tiêu chuẩn như PP, PE, PS... Đây là những dữ liệu phản ánh thực tế trong lĩnh vực nhập khẩu nhựa tại Việt Nam, với tính đa dạng cao về cách viết, tên thương mại, từ viết tắt và ký hiệu kỹ thuật. Do mục tiêu là đánh giá khả năng phân loại tự động từ văn bản gốc, nghiên cứu không áp dụng bất kỳ đặc trưng kỹ thuật nào được trích xuất thủ công. Toàn bộ thông tin đầu vào được xử lý trực tiếp từ chuỗi văn bản TEN_HANG. Quy trình tiền xử lý được thiết kế nhằm loại bỏ nhiễu và chuẩn hóa định dạng văn bản gồm: chuyển toàn bộ ký tự về dạng chữ thường, loại bỏ ký tự đặc biệt, chuẩn hóa khoảng trắng và chuyển đổi văn bản sang dạng biểu diễn số học bằng TF-IDF (term frequency-inverse document frequency). TF-IDF được áp dụng với ngữ cảnh unigram và bigram, giới hạn tối đa 500 đặc trưng xuất hiện phổ biến nhất nhằm duy trì tính tổng quát, tránh hiện tượng quá khớp và giảm chi phí tính toán. Cách biểu diễn này giúp mã hóa được các tín hiệu định danh quan trọng trong mô tả sản phẩm mà không cần hiểu ngữ nghĩa một cách đầy đủ, qua đó tạo điều kiện thuận lợi cho mô hình học máy xử lý và khai thác giá trị từ dữ liệu văn bản.

4.2. Mô hình hóa và huấn luyện

Trên cơ sở dữ liệu đã được chuẩn hóa, nghiên cứu triển khai nhiệm vụ phân loại tên hàng bằng mô hình random forest, một thuật toán thuộc nhóm học máy tổ hợp (ensemble learning), nổi bật với khả năng xử lý dữ liệu phi tuyến, kháng nhiễu tốt và đặc biệt phù hợp với các bài toán phân loại đa lớp. Đây là lựa chọn phù hợp khi tên hàng nhập khẩu có biểu hiện đa dạng, nhiều biến thể ngôn ngữ và không tuân theo quy tắc ngữ pháp cụ thể. Quy trình mô hình hóa được tổ chức thành 1 pipeline 2 giai đoạn: giai đoạn đầu thực hiện trích xuất đặc trưng văn bản bằng TF-IDF (sử dụng lại kết quả tiền xử lý từ phần trước) và giai đoạn tiếp theo là huấn luyện mô hình phân loại bằng random forest với 200 cây quyết định. Đặc biệt, để xử lý vấn đề mất cân bằng dữ liệu khi một số nhóm sản phẩm như PP_TERT hay PP_RANDOM có số lượng mẫu huấn luyện rất nhỏ, tham số class_weight='balanced' được tích hợp trong mô hình. Tham số này giúp điều chỉnh trọng số của từng lớp theo tỷ lệ xuất hiện trong tập huấn luyện, từ đó tăng cường khả năng nhận diện đối với các lớp hiếm

mà không làm ảnh hưởng đáng kể đến hiệu suất chung của mô hình. Việc tích hợp toàn bộ quy trình vào pipeline không chỉ đảm bảo tính nhất quán và khả năng tái sử dụng trong triển khai thực tế mà còn cho phép linh hoạt điều chỉnh các tham số kỹ thuật trong quá trình tối ưu hóa. Tập dữ liệu được chia theo phương pháp phân tầng (stratified sampling) nhằm đảm bảo phân bố đồng đều các nhãn trong cả hai tập huấn luyện và kiểm tra. Cụ thể, 5.765 dòng dữ liệu được sử dụng để huấn luyện mô hình, trong khi 2.471 dòng còn lại được giữ lại để đánh giá hiệu quả phân loại. Cách chia này giúp giảm thiểu sai lệch thống kê và cải thiện độ tin cậy của kết quả, đặc biệt quan trọng khi một số nhóm sản phẩm có số lượng quan sát rất thấp.

4.3. Đánh giá hiệu quả mô hình

Hiệu quả của mô hình random forest được kiểm định trên tập kiểm tra gồm 2.471 dòng dữ liệu tên hàng hóa nhập khẩu đã được gán nhãn, sử dụng các chỉ số phổ biến trong bài toán phân loại đa lớp như độ chính xác (precision), độ bao phủ (recall), điểm F1 (F1-score) và số lượng mẫu (support) trên từng nhóm nhãn cụ thể (Bảng 4). Kết quả cho thấy mô hình đạt độ chính xác tổng thể (accuracy) ở mức 81%, với các chỉ số trung bình gia quyền của precision (0,84), recall (0,76) và F1-score (0,79) (Bảng 5). Đây là kết quả khả quan khi dữ liệu thực tế có phân bố không đều giữa các lớp, nhiều nhóm nhãn có số lượng mẫu rất nhỏ và nội dung mô tả tên hàng không theo cấu trúc ngôn ngữ chuẩn hóa.

Bảng 4. *Hiệu suất phân loại theo từng nhóm nhãn trên tập kiểm tra*

Nhãn (Label)	Độ chính xác (Precision)	Độ bao phủ (Recall)	Điểm F1 (F1-Score)	Hỗ trợ (Support)
PE	1	0,5	0,67	10
PP	0,86	0,88	0,87	1100
PP_Block_Copolymer	1	0,97	0,98	59
PP_Compound	1	0,75	0,86	4
PP_Copolymer	0,77	0,85	0,81	327
PP_Homo	0,84	0,78	0,81	441
PP_Impact	0,82	0,68	0,74	301
PP_Other	0,54	0,64	0,58	140
PP_Random	0,58	0,57	0,57	86
PP_Tert	1	1	1	3

Bảng 5. *Chỉ số trung bình đánh giá tổng thể mô hình phân loại*

Chỉ số (Metrics)	Giá trị (Score)
Độ chính xác (Accuracy)	0,81
Trung bình cộng độ chính xác (Macro avg precision)	0,84
Trung bình cộng độ bao phủ (Macro avg recall)	0,76
Trung bình cộng điểm F1 (Macro avg F1-score)	0,79
Trung bình gia quyền độ chính xác (Weighted avg precision)	0,82
Trung bình gia quyền độ bao phủ (Weighted avg recall)	0,81
Trung bình gia quyền điểm F1 (Weighted avg F1-score)	0,81

Xét theo từng nhóm nhãn, mô hình thể hiện năng lực phân loại tốt đối với các nhóm phổ biến có quy mô dữ liệu lớn. Nhóm PP với 1.100 mẫu đạt F1-score 0,87, phản ánh khả năng học hiệu quả các đặc trưng từ vựng đại diện cho dòng nhựa polypropylene nói chung. Tương tự, các nhóm PP_Homo (441 mẫu) và PP_Copolymer (327 mẫu) cũng đạt F1-score lần lượt là 0,81, cho thấy mô hình có thể nhận diện tốt các biến thể kỹ thuật khác nhau trong cùng một họ sản phẩm. Đối với nhóm PP_Block_Copolymer, F1-score đạt tới 0,98 với recall lên đến 0,97, thể hiện độ chính xác cao khi phân biệt lớp có từ khóa kỹ thuật rõ ràng. Tuy nhiên, đối với các nhóm hiếm như PE (10 mẫu), PP_Compound (4 mẫu) và PP_TERT (3 mẫu), mô hình vẫn gặp khó khăn nhất định. Mặc dù precision đạt tuyệt đối (1), recall lại thấp hoặc không ổn định (lần lượt là 0,5, 0,75 và 1), dẫn đến F1-score chỉ dao động từ 0,67 - 1. Điều này phản ánh thực trạng thường thấy trong các bài toán mất cân bằng lớp, khi mô hình thiếu dữ liệu huấn luyện để học được các đặc trưng chuyên biệt của nhóm, dẫn đến hiện tượng bỏ sót (false negative).

Ngoài ra, các nhóm có ranh giới mô tả không rõ ràng như PP_Other (140 mẫu) và PP_Random (86 mẫu) chỉ đạt F1-score lần lượt là 0,58 và 0,57. Điều này cho thấy mô hình gặp khó khăn trong việc phân biệt các tên hàng chứa nội dung không đặc trưng hoặc dễ gây nhầm lẫn với các nhóm khác có từ vựng tương đồng. Đặc biệt, chỉ số macro average recall đạt mức 0,7669, thấp hơn rõ rệt so với macro precision (0,849) và macro F1-score (0,795), phản ánh sự chênh lệch trong hiệu suất mô hình giữa các lớp phổ biến và các lớp hiếm. Tuy nhiên, so với mô hình không áp dụng trọng số cho từng lớp, việc sử dụng tham số class_weight='balanced' trong huấn luyện đã giúp cải thiện đáng kể khả năng nhận diện các nhóm nhãn có số lượng mẫu hạn chế. Các chỉ số macro recall và F1-score đều tăng lên, cho thấy mô hình đã giảm được độ thiên lệch trong phân loại và trở nên công bằng hơn giữa các lớp.

Khi dữ liệu nhập khẩu thực tế liên tục phát sinh các mô tả mới, việc mô hình hoạt động

kém trên các nhóm ít phổ biến là một vấn đề cần được khắc phục. Các giải pháp kỹ thuật phù hợp có thể gồm: (i) áp dụng kỹ thuật tăng cường dữ liệu như smote hoặc data augmentation for text để mở rộng tập huấn luyện cho các lớp nhỏ; (ii) kết hợp các mô hình phân loại tiên tiến hơn như: XGBoost, LightGBM, hoặc Transformer-based models (BERT, PhoBERT) nếu bài toán yêu cầu nhận diện ngữ cảnh sâu hơn trong các mô tả sản phẩm dài. Mô hình được xây dựng bằng biểu diễn TF-IDF, phương pháp đơn giản, hiệu quả và dễ triển khai, kết quả đạt được vẫn có tính khả thi và ổn định. Điều này cho thấy trong bài toán có cấu trúc dữ liệu ngắn, không đầy đủ ngữ pháp như tên hàng trong tờ khai hải quan, TF-IDF vẫn là lựa chọn phù hợp. Bên cạnh đó, random forest được lựa chọn vì khả năng xử lý tốt dữ liệu nhiều chiều, tính ổn định cao, không quá nhạy với nhiễu và phù hợp với bài toán không yêu cầu tốc độ suy luận quá nhanh như trong thời gian thực. Kết quả này phù hợp với nghiên cứu của Fernández-Delgado và cộng sự [10], nghiên cứu chỉ ra rằng random forest là thuật toán phân loại có độ chính xác cao trong số hơn 100 thuật toán được đánh giá trên nhiều tập dữ liệu khác nhau.

Mô hình học máy không nhằm thay thế hoàn toàn hệ thống phân loại dựa trên từ khóa truyền thống, mà đóng vai trò như module hỗ trợ, cung cấp gợi ý phân loại trong các trường hợp mà phương pháp rule-based không thể áp dụng, ví dụ như với các mô tả mới, viết tắt, hoặc thiếu thông tin định danh rõ ràng... Điều này giúp nâng cao hiệu quả xử lý tự động, giảm bớt công việc thủ công trong quá trình rà soát và gán nhãn cho sản phẩm mới.

5. Kết luận

Nghiên cứu trình bày quy trình ứng dụng ngôn ngữ lập trình Python và học máy vào việc xử lý và phân loại dữ liệu xuất nhập khẩu, với mục tiêu cải thiện hiệu suất, độ chính xác và khả năng thích ứng của hệ thống xử lý dữ liệu khai báo hải quan. Thông qua việc triển khai chuỗi thao tác kỹ thuật bao gồm chuẩn hóa tên hàng, ánh xạ địa lý, quy đổi đơn vị, tính toán định lượng và phân loại sản phẩm, quy trình tự động hóa được xây dựng bằng Python đã chứng minh ưu thế rõ rệt so với phương pháp truyền thống sử dụng Excel trong các nghiên cứu quốc tế.

Kết quả thực nghiệm cho thấy việc sử dụng công cụ mới đã rút ngắn đáng kể thời gian xử lý, giảm thiểu sai sót, tăng tính nhất quán và dễ dàng mở rộng khi có dữ liệu phát sinh theo thời gian. Khi số lượng và biến thể tên hàng gia tăng nhanh, việc tích hợp mô hình học máy, cụ thể là thuật toán random forest kết hợp với biểu diễn văn bản TF-IDF, đã mở ra hướng tiếp cận khả thi nhằm tự động hóa việc gợi ý phân loại cho các dòng tên hàng mới. Mô hình đạt độ chính xác tổng thể 83% trên tập kiểm tra, với hiệu suất đặc biệt tốt ở các nhóm sản phẩm phổ biến. Tuy nhiên, các nhóm có số lượng mẫu hạn chế vẫn gặp khó khăn trong việc nhận diện, cần cải tiến thông qua việc mở rộng tập dữ liệu huấn luyện, áp dụng kỹ thuật cân bằng lớp hoặc thử nghiệm các mô hình tiên tiến hơn như LightGBM hoặc BERT.

Kết quả của nghiên cứu này khẳng định tính khả thi và hiệu quả của việc ứng dụng ngôn ngữ lập trình Python và học máy vào xử lý dữ liệu xuất nhập khẩu, cung cấp khung kỹ thuật có thể tái sử dụng và mở rộng cho các lĩnh vực tương tự như thương mại, logistics, và phân tích dữ liệu phi cấu trúc trong quản lý doanh nghiệp.

Tài liệu tham khảo

[1] Kelvin Kelvin, Wahidin Wahab, and Meirista Wulandari, "Computer resource utilization analysis for microsoft excel and python in data processing", *Engineering, Mathematics and Computer Science Journal* (EMACS), Volume 6, Issue 2, pp. 137 - 142, 2024. DOI: 10.21512/emacsjournal.v6i2.11736.

[2] Mohamed Fakhry Mansour, Tarek Aly, and Mervat Gheith, "Python based end user computing framework to empowering excel efficiency", *International Journal for Research in Applied Science and Engineering Technology*, Volume 12, Issue 4, pp. 2719 - 2729, 2024. DOI: 10.22214/ijraset.2024.60097.

[3] Raymond R. Panko and Richard P. Halverson Jr., "An experiment in collaborative spreadsheet development", *Journal of the Association for Information Systems*, Volume 2, No. 1, pp. 1 - 31, 2001. DOI: 10.17705/1jais.00016.

[4] Raymond R. Panko, "Thinking is bad: Implications of human error research for spreadsheet research and practice", *European Spreadsheet Risk Interest Group*, 2007. DOI: 10.48550/arXiv.0801.3114.

[5] Alexandros Nikolaos Ziogas, Timo Schneider, Tal Ben-Nun, Alexandru Calotoiu, Tiziano De Matteis, Johannes de Fine Licht, Luca Lavarini, and Torsten Hoefler, "Productivity, portability, performance", *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, 2021*. DOI: 10.1145/3458817.3476176.

[6] Diyyala Sravani, Jonnala Rohith Reddy, Pilla Sri Viswas, N.M. Jyothi, and Potru Chandukiran, "Python security in devOps: Best practices for secure

coding, configuration management, and continuous testing and monitoring", *4th International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 6 - 8 July 2023*. DOI: 10.1109/icesc57686.2023.10193128.

[7] Aravind Ayyagiri, Arpit Jain, and Om Goel, "Utilizing Python for scalable data processing in cloud environments", *Darpan International Research Analysis*, Volume 12, Issue 2, pp. 183 - 198, 2024. DOI: 10.36676/dira.v12.i2.78.

[8] Fabrizio Sebastiani, "Machine learning in automated text categorization", *ACM Computing Surveys*, Volume 34, Issue 1, pp. 1 - 47, 2002. DOI: 10.1145/505282.505283.

[9] ChengXiang Zhai and Sean Massung, *Text data management and analysis: A practical introduction to information retrieval and text mining*. Association for Computing Machinery and Morgan & Claypool, 2016. DOI: 10.1145/2915031.

[10] Manuel Fernández-Delgado, Eva Cernadas, Senén Barro, and Dinani Amorim, "Do we need hundreds of classifiers to solve real world classification problems?", *Journal of Machine Learning Research*, Volume 15, pp. 3133 - 3181, 2014.

[11] Asmaa M. Aubaid, Alok Mishra, and Atul Mishra, "Machine learning and rule-based embedding techniques for classifying text documents", *International Journal of Systems Assurance Engineering and Management*, Volume 15, Issue 12, pp. 5637 -5652, 2024. DOI: 10.1007/s13198-024-02555-w.

[12] Karandeep Singh, Yu-Che Tsai, Cheng-Te Li, Meeyoung Cha, and Shou-De Lin, "GraphFC: Customs fraud detection with label scarcity", *32nd ACM International Conference on Information and Knowledge Management, 2023*. DOI: 10.1145/3583780.3614690.

APPLICATION OF PYTHON AND MACHINE LEARNING IN PROCESSING AND CLASSIFYING IMPORT-EXPORT DATA

Do Hong Hanh[1], Doan Tien Quyet[1], Doan Trong Sinh[2], Nguyen Bang Linh[1]
[1]Vietnam Petroleum Institute (VPI)
[2]Vietnam National Industry - Energy Group (PVN)
Email: hanhdh@vpi.pvn.vn

Summary

Vietnam's import-export data is increasing substantially in both scale and complexity, creating significant challenges in standardizing and classifying customs declaration information. This study proposes an automated data-processing pipeline implemented in the Python programming language, with the objective of enhancing efficiency and ensuring greater consistency in the analysis of customs information. The input dataset comprises more than 10,000 real-world import-export records, which are processed through a structured sequence of technical steps, including product name normalization, unit conversion, computation of quantitative indicators, and keyword-based product group labeling.

The experimental results demonstrate that this processing pipeline operates effectively on medium-scale, high-complexity datasets, while considerably improving classification accuracy and ensuring uniformity across product categories. Based on these findings, the authors propose integrating machine learning models as a supplementary tool to enhance generalization capabilities and adaptability to exceptional cases - particularly relevant in a trade environment where product names are increasingly diverse, unstandardized, and continuously evolving.

Key words: Import, export, Python, machine learning, data normalization, product classification, TF-IDF, random forest, automation.

DầuKhí
KHOA HỌC, CÔNG NGHỆ VÀ ĐỔI MỚI SÁNG TẠO PETROVIETNAM

ỨNG DỤNG AI TRONG TỔNG HỢP THÔNG TIN THỊ TRƯỜNG VÀ PHÂN TÍCH CÁC CHỈ SỐ KINH TẾ VĨ MÔ

Đào Hương Giang, Đỗ Hồng Hạnh
Viện Dầu khí Việt Nam (VPI)
Email: giang.dh@vpi.pvn.vn
https://doi.org/10.47800/PVSI.2025.03-06

Tóm tắt

Phân tích dữ liệu trong lĩnh vực kinh tế đang chuyển từ các phương pháp truyền thống sang hiện đại dựa trên trí tuệ nhân tạo (AI) giúp tự động hóa quá trình thu thập và xử lý dữ liệu, có khả năng dự báo sự biến động của thị trường, là cơ sở để doanh nghiệp đưa ra quyết định chính xác hơn. Bài viết giới thiệu kết quả ứng dụng AI tại Viện Dầu khí Việt Nam (VPI) để phát triển sản phẩm phân tích số tổng hợp thông tin thị trường, giúp nâng cao hiệu quả công tác tổng hợp, xử lý dữ liệu tự động; dự báo chỉ tiêu kinh tế và cung cấp trợ lý ảo hỗ trợ hỏi đáp thông tin thị trường.

Từ khóa: Kinh tế vĩ mô, trí tuệ nhân tạo, thị trường, xử lý dữ liệu, trợ lý ảo.

1. Giới thiệu

Những thay đổi trong nền kinh tế tác động mạnh đến nhiều lĩnh vực, bao gồm năng lượng nói chung và dầu khí nói riêng. Việc phân tích dữ liệu kinh tế không chỉ cung cấp bức tranh toàn cảnh nhằm hỗ trợ theo dõi, đánh giá diễn biến nền kinh tế mà còn là dữ liệu đầu vào quan trọng phục vụ công tác nghiên cứu về thị trường năng lượng và dầu khí.

Phân tích dữ liệu trong lĩnh vực kinh tế đang chuyển từ các phương pháp truyền thống sang hiện đại dựa trên công nghệ như trí tuệ nhân tạo (AI). Việc tích hợp AI vào phân tích dữ liệu đã tạo ra sự đổi mới mang tính cách mạng trong hoạt động của các tổ chức tài chính, chủ yếu nhờ vào khả năng xử lý, phân tích và diễn giải các bộ dữ liệu khổng lồ [1]. AI đã được sử dụng trong kinh tế học cho các mục đích dự báo, phân tích thị trường và đánh giá tác động của các chính sách [2]; hoạt động tài chính và quy trình phân tích thông tin cho các giao dịch thương mại [3]; hoặc phân tích thông tin tài chính [4]. Xu hướng này khẳng định vai trò ngày càng quan trọng của AI trong việc nâng cao hiệu quả và độ chính xác trong phân tích dữ liệu kinh tế.

Thay vì tra cứu thủ công từ 3 - 5 nguồn báo cáo khác nhau, AI có thể hỗ trợ tìm kiếm thông tin, giúp tiết kiệm 50 - 70% thời gian thực hiện. Đặc biệt, sản phẩm số tổng hợp, xử lý và lưu trữ dữ liệu vào cơ sở dữ liệu có cấu trúc. Sau khi dữ liệu được thu thập, AI hỗ trợ xử lý theo nhu cầu người dùng và lưu trữ vào cơ sở dữ liệu có cấu trúc, giúp dễ dàng truy xuất và phân tích sau này.

Với việc trực quan hóa thông tin trên báo cáo, sản phẩm cho phép người dùng dễ dàng theo dõi các chỉ số kinh tế vĩ mô, được thiết kế với giao diện dễ sử dụng, có thể tùy chỉnh để hiển thị các biểu đồ và số liệu theo yêu cầu. Đặc biệt, sản phẩm tích hợp chatbot giúp người dùng dễ dàng tra cứu dữ liệu và thông tin. Tính năng hỏi đáp thông minh cho phép người dùng đặt câu hỏi trực tiếp bằng ngôn ngữ tự nhiên và nhận lại câu trả lời trong thời gian dưới 30 giây.

Tính khả thi của sản phẩm được đảm bảo nhờ khả năng tự động hóa cao, giúp giảm đáng kể thời gian vận hành và chi phí bảo trì. Với thiết kế linh hoạt và khả năng mở rộng, sản phẩm có tiềm năng duy trì và phát triển lâu dài, đáp ứng nhu cầu ngày càng phức tạp của người dùng. Người dùng được hỗ trợ không chỉ là các nhà quản lý, chuyên gia làm việc trong lĩnh vực kinh tế mà còn cả trong lĩnh vực quản lý ngành năng lượng. Sản phẩm có thể hỗ trợ thu thập, phân tích dữ liệu vĩ mô liên quan đến thị trường năng lượng như: cung - cầu, xu hướng giá cả; tối ưu hóa quá trình quản lý dữ liệu và phân tích thị trường; xây dựng báo cáo phân tích các yếu tố ảnh hưởng đến ngành năng lượng cho các cơ quan quản lý nhà nước

Ngày nhận bài: 20/11/2024
Ngày đánh giá và sửa chữa: 20/11 - 18/12/2024
Ngày duyệt đăng: 18/12/2024.

và tổ chức nghiên cứu; các công ty tư vấn và tổ chức phân tích năng lượng…

Bài viết giới thiệu sản phẩm số phục vụ phân tích dữ liệu kinh tế vĩ mô tích hợp trí tuệ nhân tạo giúp tự động tìm kiếm và thu thập dữ liệu từ nhiều nguồn, bao gồm báo cáo kinh tế, dữ liệu từ các tổ chức quốc tế, dữ liệu tài chính từ các sàn giao dịch và các nguồn dữ liệu công khai khác.

2. Đánh giá các sản phẩm phân tích kinh tế vĩ mô

Các sản phẩm phân tích dữ liệu kinh tế trên thị trường cung cấp nguồn thông tin phong phú với khối lượng lớn, đáng tin cậy và được cập nhật định kỳ (theo ngày, tháng, quý hoặc năm). Tuy nhiên, khối lượng và tính cập nhật của dữ liệu trong các sản phẩm này phụ thuộc đáng kể vào mục tiêu sử dụng và nguồn lực của đơn vị phát triển. Các sản phẩm thường có ưu điểm trong việc cung cấp thông tin tổng hợp nhưng đôi khi gặp hạn chế về độ trễ trong cập nhật dữ liệu hoặc mức độ chi tiết phù hợp cho các nhu cầu phân tích sâu.

Cơ sở dữ liệu của Tổng cục Thống kê (GSO) là nguồn thông tin quan trọng và chính thức, cung cấp dữ liệu về các lĩnh vực kinh tế - xã hội của Việt Nam. Ưu điểm của nguồn dữ liệu này là tính chính xác cao và đáng tin cậy, với các chỉ số vĩ mô như GDP, lạm phát, tỷ lệ thất nghiệp, xuất nhập khẩu, dân số và các chỉ tiêu phát triển bền vững. Dữ liệu được phân nhóm, hỗ trợ người dùng tra cứu hoặc yêu cầu phân tích toàn diện và chuyên sâu, với khối lượng lớn và phong phú. Dữ liệu thường được công bố định kỳ theo tháng, quý hoặc năm, tùy thuộc vào loại chỉ tiêu, giúp duy trì tính ổn định và đáng tin cậy của thông tin. Tuy nhiên, nhược điểm của cơ sở dữ liệu này là chậm đáp ứng nhu cầu phân tích thời gian thực hoặc biến động ngắn hạn trên thị trường, do dữ liệu được công bố định kỳ thay vì liên tục.

Trong khi đó, Ngân hàng Thế giới (World Bank) cung cấp một khối lượng lớn dữ liệu về các chỉ số kinh tế vĩ mô và xã hội của hầu hết các quốc gia, với các lĩnh vực như phát triển bền vững, xóa đói giảm nghèo và giáo dục. Ưu điểm của nguồn dữ liệu này là phong phú và toàn diện, phù hợp với các phân tích toàn cầu và khu vực. Dữ liệu được cập nhật định kỳ hàng năm hoặc hàng quý, giúp cung cấp cái nhìn tổng quan về các vấn đề kinh tế và xã hội. Tuy nhiên, nhược điểm của cơ sở dữ liệu này cũng liên quan đến độ trễ trong việc cập nhật dữ liệu mới.

Ngược lại, Financial Times có khả năng cung cấp dữ liệu tài chính và kinh tế theo thời gian thực với khả năng cập nhật liên tục các biến động trên thị trường toàn cầu. Ưu điểm của nguồn dữ liệu này là tính kịp thời và chi tiết, cho phép theo dõi các thay đổi thị trường tức thời, đặc biệt trong lĩnh vực chứng khoán, tiền tệ, hàng hóa… Khối lượng dữ liệu lớn và được phân nhóm, giúp người dùng dễ dàng truy cập và phân tích. Tuy nhiên, nhược điểm của nguồn dữ liệu này là phạm vi chủ yếu tập trung vào các yếu tố tài chính, dẫn đến hạn chế khi sử dụng trong phân tích dài hạn và toàn diện về tình hình kinh tế - xã hội.

Với xu hướng trực quan hóa dữ liệu như hiện nay, các sản phẩm phân tích kinh tế thường được biểu diễn thông qua các dashboard. Ưu điểm của các sản phẩm dashboard trên thị trường là sự đa dạng về tính năng, giao diện và mức độ thân thiện với người dùng, đáp ứng nhiều mục tiêu phát triển khác nhau của các tổ chức. Tính năng nổi bật bao gồm khả năng tùy chỉnh biểu đồ, truy vấn dữ liệu theo thời gian thực và tích hợp các công cụ phân tích chuyên sâu. Về giao diện, hầu hết các sản phẩm có thiết kế hiện đại, tập trung vào trải nghiệm người dùng với bố cục hợp lý và khả năng tương tác cao. Tuy nhiên, các sản phẩm trên thị trường vẫn tồn tại một số hạn chế liên quan đến độ thân thiện. Dashboard hướng đến người dùng phổ thông thường dễ tiếp cận nhưng giới hạn về chức năng tùy chỉnh. Trong khi đó, các công cụ chuyên sâu dành cho nhà phân tích hoặc nhà đầu tư đòi hỏi kỹ năng sử dụng nhất định, khiến các sản phẩm trở nên khó tiếp cận đối với người dùng không chuyên.

Ví dụ, infographic của GSO có ưu điểm là tập trung vào việc truyền tải thông tin dưới dạng biểu đồ, hình ảnh và số liệu trực quan, với các chỉ số được trình bày dễ hiểu, phù hợp với người dùng phổ thông. Giao diện của sản phẩm đơn giản, dễ tiếp cận, giúp người dùng nhanh chóng nắm bắt thông tin cơ bản. Tuy nhiên, hạn chế của sản phẩm là chưa tích hợp nhiều công cụ tương tác và phần lớn nội dung mang tính tĩnh, khó đáp ứng các nhu cầu tùy chỉnh hoặc khai thác dữ liệu linh hoạt của người dùng.

Ngược lại, dashboard của World Bank [5] hoặc Tổ chức Hợp tác và Phát triển Kinh tế (OECD) [6] có ưu điểm là cung cấp các công cụ trực quan hóa, cho phép tùy chỉnh biểu đồ và truy vấn dữ liệu theo quốc gia, khu vực và các chỉ số cụ thể. Giao diện được thiết kế hợp lý với hệ thống lựa chọn rõ ràng, dễ sử dụng, phù hợp cho cả người dùng không chuyên. Tuy nhiên, các dashboard này chưa tích hợp công cụ hỏi đáp ứng dụng AI, làm giảm khả năng hỗ trợ thông minh và hiệu quả đối với người dùng cần giải pháp có tính tương tác cao hơn. Ngoài ra, sự đa dạng trong các loại dashboard còn hạn chế khi chỉ có một số

chỉ số được biểu diễn, chưa đáp ứng được nhu cầu phân tích phong phú ở nhiều lĩnh vực khác nhau của nền kinh tế.

Tương tự, dashboard của Financial Times [7] có ưu điểm nổi bật ở khả năng tương tác, cho phép người dùng điều chỉnh biểu đồ theo thời gian hoặc danh mục đầu tư. Đây là công cụ hỗ trợ hữu ích cho việc phân tích thị trường và theo dõi liên tục biến động tài chính. Tuy nhiên, sản phẩm này cũng có hạn chế tương tự một số dashboard khác khi chưa tích hợp công cụ AI giúp tối ưu hóa trải nghiệm người dùng.

Những hạn chế trên đã mở ra nhu cầu cấp thiết cho việc tích hợp công nghệ mới như AI vào quy trình phân tích dữ liệu kinh tế, khi tốc độ và khả năng thích ứng là yếu tố then chốt.

3. Ứng dụng AI phát triển sản phẩm số phân tích kinh tế vĩ mô

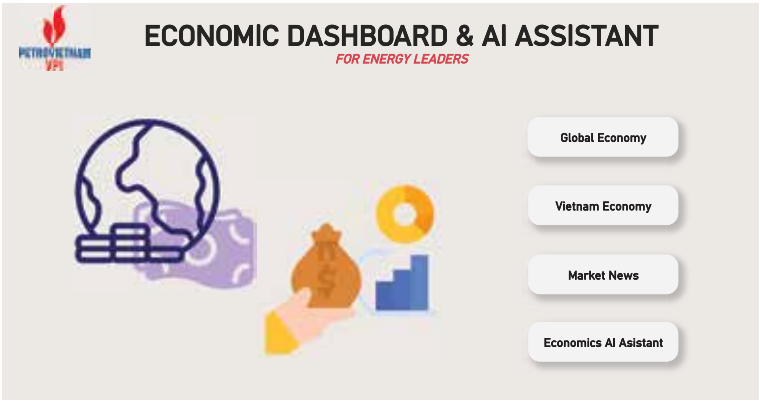
Sản phẩm "Phân tích số ứng dụng AI/ML tổng hợp thông tin thị trường và dự báo chỉ tiêu kinh tế" được VPI triển khai trên nền tảng Power BI và Microsoft Fabric, đảm bảo cho việc bảo mật, lưu trữ và cộng tác trong quá trình xây dựng sản phẩm, khả năng tự động hóa và tái sử dụng đối với sản phẩm cũng như khả năng truy cập của người dùng cuối cùng.

Đối với tính năng tra cứu dữ liệu, sản phẩm gồm dashboard được xây dựng dựa trên các thông tin liên quan đến kinh tế thế giới và Việt Nam, phân bố tại các báo cáo khác nhau. Do đó, để tìm kiếm thông tin cụ thể, người dùng có thể truy cập vào trang báo cáo đã được đặt tên. Tại mỗi trang báo cáo đều bố trí các biểu đồ thể hiện thông tin khác nhau, người dùng có thể truy cập ngay. Hoặc nhờ vào khả năng liên kết dữ liệu và hiển thị nâng cao của Power BI, người dùng có thể trỏ con trỏ chuột vào phân vùng biểu đồ mong muốn và thông tin liên quan sẽ được hiển thị, gồm: kinh tế toàn cầu, kinh tế Việt Nam, thông tin thị trường và trợ lý ảo.

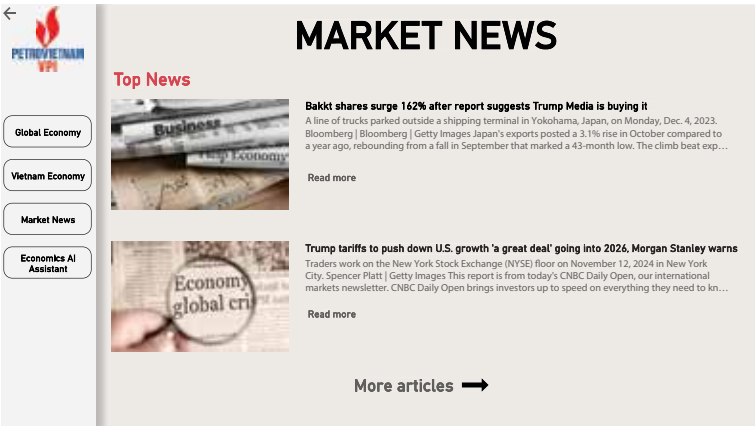
Đối với tính năng tùy chỉnh xem báo cáo, sản phẩm giúp người dùng lựa chọn kiểu xem (view) báo cáo theo khoảng thời gian nhất định, theo một hoặc nhiều nguồn thông tin. Ngoài ra, sản phẩm được thiết kế để người dùng có thể lựa chọn xem theo: thông tin tóm tắt, thông tin đầy đủ và các thông tin liên quan khác...

Đối với tính năng hỏi đáp với trợ lý ảo, sản phẩm cho phép người dùng đặt các câu hỏi liên quan đến thị trường kinh tế ngoài phạm vi dữ liệu được lưu trữ của sản phẩm. Tính năng này được xây dựng dựa trên trí tuệ nhân tạo AI, giúp nhận biết các yêu cầu tra cứu của người dùng.

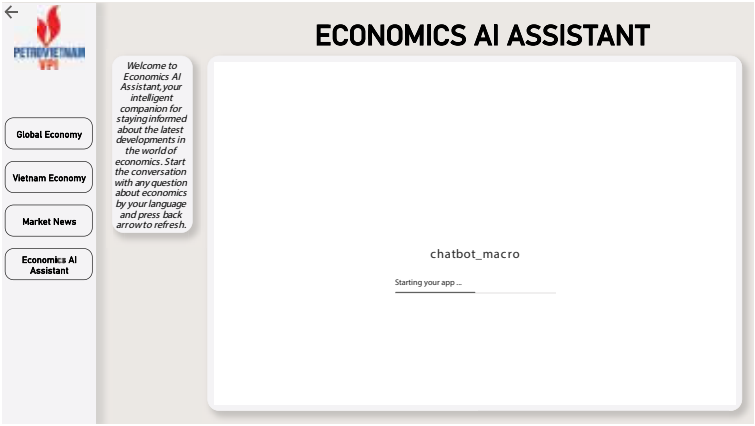
Về giao diện, cấu trúc của giao diện sản phẩm được tạo bởi các trang báo cáo; mỗi trang báo cáo có các biểu đồ tương ứng với tính năng hoặc nhóm thông tin có liên quan với nhau về nền kinh tế.



Hình 1. *Giao diện điều hướng báo cáo.*



Hình 2. *Tính năng điều chỉnh view và xem báo cáo.*

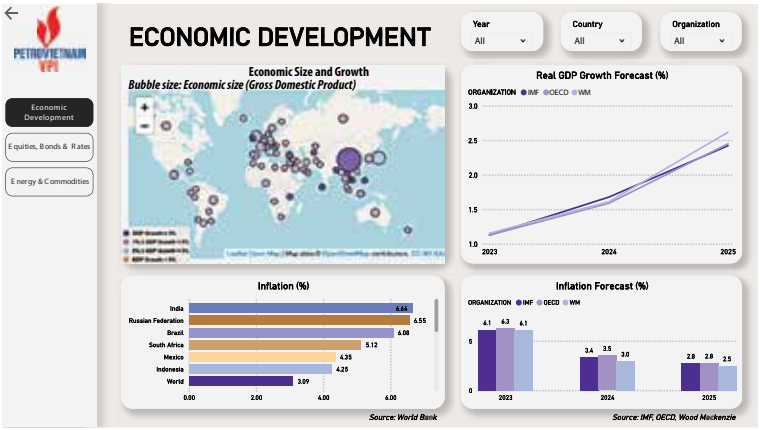


Hình 3. *Tính năng hỏi đáp với trợ lý ảo.*

Cấu trúc giao diện gồm 6 trang báo cáo: Economic Development; Equities, Bonds & Rates; Energy & Commodities; Vietnam Economy; Market News và Economics AI Assistant.

Để xây dựng sản phẩm, dữ liệu cấu trúc và phi cấu trúc từ kho dữ liệu VPI và các tổ chức (World Bank, Wood Mackenzie, Reuters,…) được thu thập, tổng hợp và xử lý với Python Notebook trên nền tảng Microsoft Fabric phục vụ truy vấn dữ liệu tự động hóa tối đa. Một số đặc điểm dữ liệu được mô tả trong Bảng 1.

Với mục tiêu nâng cao hiệu quả tốc độ xử lý, chi phí duy trì và khả năng tái sử dụng của sản phẩm, AI đã được ứng dụng trong một số quy trình kỹ thuật.



Hình 4. *Giao diện biểu diễn các chỉ tiêu tăng trưởng kinh tế.*

- Tổng hợp và xử lý thông tin

Các công cụ AI như Exa [8] và Cohere [9] được triển khai để tự động hóa các quy trình này, giảm thiểu sự can thiệp của con người và tối ưu hóa độ chính xác của dữ liệu thu thập được. Exa đảm nhận việc thu thập và xử lý dữ liệu thô, trong khi Cohere chịu trách nhiệm tổng hợp và tóm tắt thông tin.

Để đảm bảo chất lượng dữ liệu đầu vào, nhóm tác giả dựa vào kiến thức chuyên môn nhằm lựa chọn những nguồn thông tin uy tín: Reuters, CNN… và thiết lập các tham số tìm kiếm bao gồm: domain (nguồn), category (danh mục), query (truy vấn) và ngày cập nhập. Trong phạm vi sản phẩm này, các nguồn cung cấp tin như Reuters, CNN…, danh mục "news" (tin tức) và ngày cập nhật gần nhất được ưu tiên đưa vào tìm kiếm. Cách tiếp cận này đảm bảo dữ liệu đầu vào liên quan trực tiếp đến chủ đề phân tích, có tính kịp thời và độ chính xác cao.

Thông tin được thu thập đảm bảo không chứa các yếu tố nhiễu hoặc không liên quan đến chủ đề của sản phẩm. Để giảm thiểu kết

Bảng 1. *Mô tả đặc điểm các dữ liệu trong cơ sở dữ liệu*

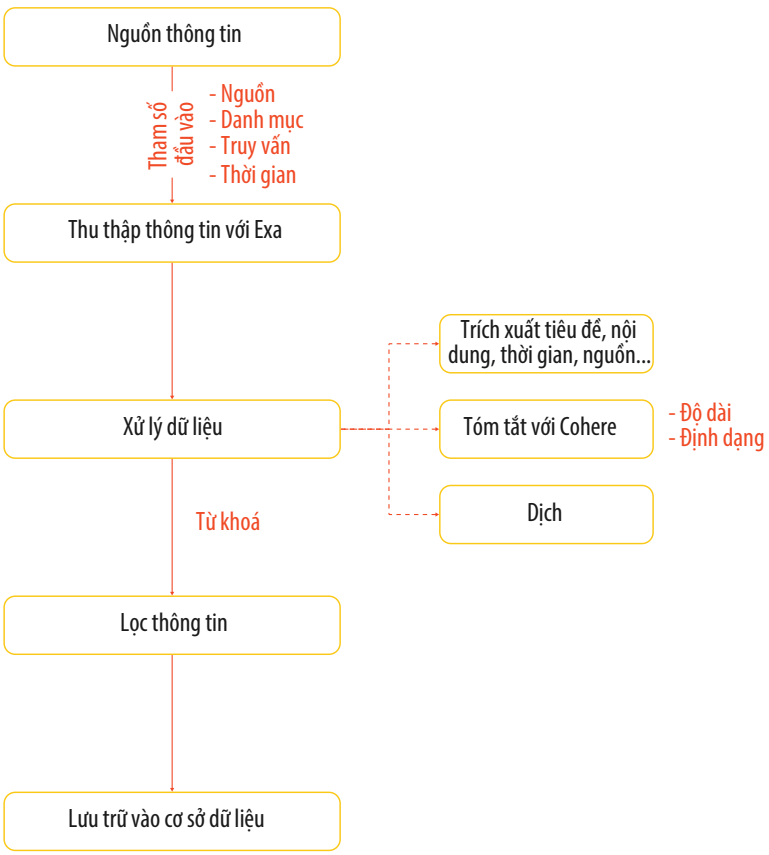
Chỉ số/chỉ tiêu	Nguồn	Giai đoạn	Tần suất
Tổng sản phẩm quốc nội (90 quốc gia)	World Bank	2010 - Hiện tại	Năm
Lạm phát (20 quốc gia)	World Bank	2010 - Hiện tại	Năm
Dự báo tổng sản phẩm quốc nội toàn cầu	IMF, OECD, Wood Mackenzie	2023 - 2025	Năm
Dự báo lạm phát toàn cầu	IMF, OECD, Wood Mackenzie	2023 - 2025	Năm
Lãi suất các Ngân hàng Trung ương (Mỹ, Anh, châu Âu)	Cục Dự trữ Liên bang, Ngân hàng Trung ương Anh - châu Âu	2018 - Hiện tại	Tháng
Lợi suất trái phiếu Chính phủ Mỹ 2 năm	Cục Dự trữ Liên bang	2018 - Hiện tại	Ngày
Lợi suất trái phiếu Chính phủ Mỹ 10 năm	Cục Dự trữ Liên bang	2018 - Hiện tại	Ngày
Chỉ số MSCI toàn cầu	Investing	2018 - Hiện tại	Ngày
Chỉ số MSCI châu Á (trừ Nhật Bản)	Yahoo!Finance	2018 - Hiện tại	Ngày
Giá dầu thô Dated Brent	EIA	2018 - Hiện tại	Ngày
Giá khí tự nhiên	EIA	2018 - Hiện tại	Ngày
Giá vàng	Investing	2018 - Hiện tại	Ngày
Giá hợp đồng tương lai dầu thô Dated Brent	ICE	2023 - Hiện tại	Giờ
Giá hợp đồng tương lai khí tự nhiên	CME	2023 - Hiện tại	Giờ
Giá hợp đồng tương lai vàng	CME	2023 - Hiện tại	Giờ
Chỉ số quản lý thu mua Việt Nam	S&P Global	2018 - Hiện tại	Tháng
Chỉ số giá tiêu dùng Việt Nam	Tổng cục Thống kê	2018 - Hiện tại	Tháng
Tỷ giá hối đoái Việt Nam đồng với USD, bảng Anh và EUR	Investing	2018 - Hiện tại	Ngày
Thông tin thị trường	Reuters, CNN…	2023 - Hiện tại	Ngày

quả nhiễu, nhóm tác giả đã thiết kế query cụ thể, tập trung vào nội dung liên quan đến chủ đề muốn tìm kiếm. Các query này được xây dựng với một số từ khóa và điều kiện nhằm định hướng hệ thống chỉ thu thập những thông tin cần thiết. Ví dụ query được sử dụng với Exa: Các tin tức kinh tế mới nhất bao gồm, nhưng không giới hạn trong các nội dung về tài chính, tiền tệ, GDP, lạm phát, lãi suất, triển vọng kinh tế.

Query này không chỉ giới hạn phạm vi thu thập vào các lĩnh vực quan trọng trong nền kinh tế mà còn lọc bỏ các nội dung ít liên quan hoặc không phù hợp ngay từ bước đầu tiên. Việc sử dụng query cụ thể này đảm bảo dữ liệu đầu vào đã qua bước sàng lọc sơ bộ, giảm thiểu khối lượng thông tin nhiễu phải xử lý trong các giai đoạn tiếp theo.

Thông tin thị trường kinh tế được xử lý và tóm tắt bằng công cụ xử lý ngôn ngữ tự nhiên Cohere. Để hướng dẫn Cohere xử lý thông tin chính xác và hiệu quả, nhóm tác giả đã cung cấp tham số về độ dài, định dạng. Đối với sản phẩm này, Cohere được yêu cầu tóm tắt văn bản với độ dài trung bình (medium) và định dạng đoạn (paragraph). Thông tin sẽ được dịch qua tiếng Việt với các thư viện cung cấp khả năng dịch văn bản trên ngôn ngữ lập trình Python.

Sau khi được xử lý và tóm tắt bằng Cohere, thông tin tiếp tục được kiểm tra để đánh giá nhằm nâng cao chất lượng. Để đảm bảo thông tin đưa vào báo cáo số thuộc chủ đề kinh tế, nhóm tác giả sử



Hình 5. *Quy trình tổng hợp và xử lý thông tin.*

dụng ngôn ngữ lập trình Python nhằm lọc tin theo các từ khóa cụ thể. Sau khi xác định các từ khóa liên quan đến chủ đề kinh tế, chẳng hạn như "GDP", "lạm phát", "thị trường chứng khoán", "chính sách tiền tệ", "chỉ số giá tiêu dùng (CPI)"…, các từ khóa này được sử dụng để quét và phân loại thông tin dựa vào nội dung đã tóm tắt từ Cohere. Bước này đảm bảo báo cáo tập trung vào thông tin liên quan trực tiếp đến chủ đề kinh tế, giúp nâng cao chất lượng thông tin và độ chính xác với chủ đề.

Quy trình tổng hợp và xử lý thông tin ứng dụng công cụ AI giúp tiết kiệm thời gian và chi phí do giảm thiểu được nhân lực làm việc thủ công, đồng thời đảm bảo chất lượng thông tin do có sự kiểm soát của nhóm chuyên môn. Đồng thời, quy trình này có thể được sử dụng trong một số sản phẩm có yêu cầu tương tự, giúp tiết kiệm thời gian xây dựng mới.

- Chatbot hỏi đáp thông tin kinh tế

Chatbot hỗ trợ hỏi đáp thông tin thị trường kinh tế được xây dựng trên nền tảng Copilot Studio [10], cung cấp thông tin cập nhật về nền kinh tế toàn cầu và Việt Nam. Với việc ứng dụng các công nghệ tiên tiến như trí tuệ nhân tạo và xử lý ngôn ngữ tự nhiên (NLP), chatbot giúp người dùng dễ dàng tiếp cận và tìm hiểu về các chỉ số kinh tế, tin tức thị trường và các sự kiện tài chính quan trọng.

Để đảm bảo chatbot có thể tìm kiếm thông tin chính xác và phù hợp với câu hỏi của người dùng, kỹ thuật prompting đã được phát triển. Đây là tập hợp các đoạn văn bản hoặc chỉ thị cụ thể, thiết kế dựa trên kiến thức chuyên môn của nhóm tác giả trong lĩnh vực kinh tế và cấu trúc dữ liệu chuyên ngành. Các prompt này được xây dựng với mục tiêu định hướng chatbot trong việc tìm kiếm thông tin từ các nguồn đáng tin cậy và uy tín, đồng thời đảm bảo rằng thông tin được cung cấp sát với yêu cầu của người dùng. Ngoài ra, chatbot còn được thiết kế để tìm kiếm và phản hồi theo ngôn ngữ của người dùng, giúp nâng cao trải nghiệm người dùng và đảm bảo rằng thông tin nhận được là dễ hiểu và chính xác. Quá trình thử nghiệm và cải tiến các prompt với sự tham gia của nhóm chuyên môn và

người dùng cuối giúp nâng cao chất lượng của chatbot. Các prompt được điều chỉnh và cập nhật dựa trên phản hồi và kết quả thực tế, đảm bảo rằng chatbot có thể đáp ứng các yêu cầu đa dạng, hiệu quả hơn trong việc cung cấp thông tin kinh tế quan trọng.

Ưu điểm lớn nhất của việc tích hợp AI và NLP là khả năng tạo ra giao diện thân thiện và dễ sử dụng. Cụ thể, chatbot hỏi đáp sử dụng AI và NLP cho phép người dùng tương tác với hệ thống bằng ngôn ngữ tự nhiên mà không cần phải học các thao tác phức tạp. Bên cạnh đó, việc tích hợp AI và NLP là khả năng tái sử dụng và mở rộng hệ thống. Các chatbot được xây dựng trên nền tảng AI có thể dễ dàng cập nhật và mở rộng để đáp ứng nhu cầu thay đổi của người dùng mà không cần phải thiết kế lại toàn bộ hệ thống, đồng thời có thể được kết nối và sử dụng trên một sản phẩm trong lĩnh vực tương tự.

4. Đánh giá sản phẩm phân tích số ứng dụng AI/ML tổng hợp thông tin thị trường và dự báo chỉ tiêu kinh tế

Sản phẩm "Phân tích số ứng dụng AI/ML tổng hợp thông tin thị trường và dự báo chỉ tiêu kinh tế" có ưu điểm hơn so với một số sản phẩm khác trên thị trường.

- Độ chính xác của các mô hình AI được sử dụng giúp nâng cao chất lượng sản phẩm. Exa và Cohere là các mô hình được huấn luyện trên bộ dữ liệu lớn và đa dạng, được thiết kế với khả năng tự cải thiện thông qua cơ chế học liên tục, thường xuyên được tinh chỉnh và cập nhật dựa trên phản hồi của người dùng và xu hướng phát triển công nghệ. Việc ứng dụng các mô hình này vào sản phẩm giúp đảm bảo độ chính xác của việc thu thập, tổng hợp dữ liệu và thông tin liên quan đến chủ đề tới hơn 90%. Tính chính xác của các mô hình AI không chỉ được đảm bảo nhờ khả năng vượt trội của công nghệ mà còn được củng cố bởi sự can thiệp của nhóm chuyên môn dựa trên kết quả từ mô hình. Nhờ sự kết hợp này, sản phẩm có thể đạt độ chính xác cao hơn so với các giải pháp chỉ sử dụng AI hoặc chỉ dựa vào kiến thức chuyên môn của con người.

- Khả năng xử lý dữ liệu của sản phẩm có ưu điểm là tốc độ xử lý nhanh và quản lý khối lượng dữ liệu lớn. Ngôn ngữ lập trình cũng như các công cụ AI được sử dụng hỗ trợ tự động cập nhật thường xuyên và xử lý hiệu quả dữ liệu mới trong vài chục giây đến vài phút. Khả năng này giúp tiết kiệm thời gian thực hiện và giảm sự can thiệp thủ công so với một số giải pháp trên thị trường hiện nay.

- Chi phí triển khai và vận hành được tối ưu nhờ ứng dụng các công cụ và quy trình tích hợp AI. So với một số giải pháp yêu cầu đội ngũ nhân công lớn để thu thập, xử lý và tổng hợp dữ liệu, sản phẩm này giúp tiết kiệm chi phí nhân công do tự động hóa tương đối các bước nêu trên.

- Các mô hình như Exa và Cohere hoạt động dựa trên cơ sở hạ tầng điện toán đám mây, giảm nhu cầu đầu tư ban đầu vào phần cứng. Với khả năng tự động hóa và tích hợp linh hoạt, sản phẩm giúp giảm thiểu chi phí đào tạo, thiết kế lại hoặc bổ sung nguồn lực khi mở rộng sang lĩnh vực khác, đồng thời đảm bảo tính nhất quán và khả năng mở rộng linh hoạt, dễ dàng tích hợp thêm nguồn dữ liệu hoặc mở rộng phạm vi phân tích.

Tuy nhiên, sản phẩm vẫn còn hạn chế khi phụ thuộc vào chất lượng dữ liệu đầu vào. Nếu dữ liệu đầu vào không đầy đủ, không đồng nhất, hoặc có sai sót, kết quả phân tích có thể không đáng tin cậy. Hơn nữa, AI đôi khi tạo ra những cảnh báo hoặc dự đoán khó hiểu hoặc khó xác thực, dẫn đến rủi ro trong việc ra quyết định. Do đó, việc xây dựng các cơ chế kiểm tra và đánh giá chất lượng dữ liệu cũng như kết quả phân tích là điều cần thiết để đảm bảo độ tin cậy của sản phẩm. Mặc dù AI có khả năng xử lý và phân tích khối lượng lớn dữ liệu, nhưng con người vẫn đóng vai trò quan trọng trong việc kiểm tra, đánh giá, diễn giải và bổ sung các khía cạnh mà công nghệ không thể xử lý hoàn chỉnh, ví dụ bối cảnh thực tế, hiểu biết trong lĩnh vực hoặc các yếu tố văn hóa - xã hội. Sự kiểm soát của con người đảm bảo AI không chỉ hoạt động chính xác về mặt kỹ thuật mà còn phù hợp với các mục tiêu thực tế và yêu cầu của công việc.

Trong tương lai, sản phẩm có thể được phát triển, nâng cao chất lượng thông qua áp dụng các công nghệ nâng cao, sử dụng mô hình học máy hiện đại và phức tạp hơn giúp dự báo một số chỉ tiêu kinh tế. Ngoài ra, sản phẩm cần tiếp tục cải thiện các cơ chế kiểm soát chất lượng dữ liệu, xây dựng hệ thống phản hồi tự động từ người dùng để điều chỉnh mô hình và tăng cường tính tương tác thông qua trợ lý ảo hoặc chatbot hỗ trợ hỏi đáp.

5. Kết luận

Sản phẩm "Phân tích số ứng dụng AI/ML trong tổng hợp thông tin thị trường và dự báo chỉ tiêu kinh tế" giúp nâng cao tốc độ và quy mô xử lý dữ liệu thông qua việc ứng dụng AI. Hệ thống có khả năng tự động đọc, trích xuất và tóm tắt tin tức, báo cáo và dữ liệu chính thống gần thời gian thực. Đặc biệt, công nghệ AI cho phép tổng hợp và diễn giải thông tin theo ngữ cảnh, tạo ra các bản tóm tắt có trích dẫn nguồn rõ ràng. Nhờ khả năng linh hoạt và tích hợp cao, sản phẩm có thể dễ dàng kết nối trực tiếp với các hệ thống BI, công cụ trực quan hóa

dữ liệu và tùy biến các báo cáo phân tích chuyên sâu phục vụ yêu cầu quản lý, nghiên cứu và ra quyết định. Tuy nhiên, việc ứng dụng AI cần được quản trị chặt chẽ để bảo đảm độ tin cậy và tính minh bạch của hệ thống; có cơ chế đánh giá chất lượng và kiểm thử định kỳ; duy trì, hiệu chỉnh và cập nhật mô hình thường xuyên để đảm bảo tính chính xác, ổn định và kịp thời của hệ thống phân tích.

Định hướng phát triển sản phẩm "Phân tích số ứng dụng AI/ML trong tổng hợp thông tin thị trường và dự báo chỉ tiêu kinh tế" tập trung vào tăng cường năng lực cốt lõi của hệ thống, đặc biệt là khả năng tương tác với người dùng và trích xuất tri thức chuyên sâu. Cụ thể, phát triển các mô hình AI có khả năng phân tích ngữ cảnh sâu sắc trong câu hỏi của người dùng, không chỉ dừng lại ở việc phản hồi đơn lẻ mà là quy trình phân tích toàn diện để đưa ra câu trả lời chính xác, có căn cứ và trích dẫn nguồn rõ ràng. Tiếp theo là tăng khả năng trích xuất insight, thông qua việc liên kết AI với kho tri thức vĩ mô có cấu trúc, đảm bảo từng kết luận phân tích, phát hiện bất thường hay giải thích giả định đều được minh chứng bằng nguồn dữ liệu xác thực. Mục tiêu của nhóm tác giả là xây dựng trợ lý phân tích thông minh, giúp tăng độ minh bạch và hiệu suất ra quyết định..

Tài liệu tham khảo

[1] Paulo Silva, Carolina Goncalves, Nuno Antunes, Marilia Curado, and Bogdan Walek, "Privacy risk assessment and privacy-preserving data monitoring", *Expert Systems with Applycation*, Volume 200, 2022. DOI: 10.1016/j.eswa.2022.116867.

[2] Tohid Atashbar and Rui Aruhan Shi, "AI and macroeconomic modelling: Deep reinforcement learning in an RBC model", 24/2/2023. [Online]. Available: https://www.imf.org/en/Publications/WP/Issues/2023/02/24/AI-and-Macroeconomic-Modeling-Deep-Reinforcement-Learning-in-an-RBC-model-530084.

[3] Rodolfo C. Cavalcante, Rodrigo C. Brasileiro, Victor L.F. Souza, Jarley P. Nobrega, and Adriano L.I. Oliveira, "Computational intelligence and financial markets: A survey and future directions", *Expert Systems with Applications*, Volume 55, pp. 194 - 211, 2016. DOI: 10.1016/j.eswa.2016.02.006.

[4] B. Shravan Kumar and Vadlamani Ravi, "A survey of the applications of text mining in financial domain", *Knowledge-Based Systems*, Volume 114, pp. 128 - 147, 2016. DOI: 10.1016/j.knosys.2016.10.003.

[5] World Bank, "PPI visualization dashboard". [Online]. Available: https://ppi.worldbank.org/en/visualization#sector=&status=&ppi=&investment=&region=&ida=&income=&ppp=&mdb=&year=&excel=false&map=&header=true.

[6] OECD, "Households' economic well-being: The OECD dashboard". [Online]. Available: https://www.oecd.org/en/data/dashboards/households-economic-well-being-the-oecd-dashboard.html.

[7] Financial Times, "Markets data". [Online]. Available: https://markets.ft.com/data.

[8] Exa. [Online]. Available: https://exa.ai/.

[9] Cohere. [Online]. Available: https://cohere.com/.

[10] Copilot Studio. [Online]. Available: https://copilotstudio.microsoft.com/.

AI APPLICATIONS IN MARKET INFORMATION COLLECTION AND MACROECONOMIC INDICATOR ANALYSIS

Dao Huong Giang, Do Hong Hanh
Vietnam Petroleum Institute (VPI)
Email: giang.dh@vpi.pvn.vn

Summary

Data analysis in economics is transitioning from traditional methods to AI-driven approaches, enabling automation in data collection and processing, as well as improving the ability to forecast market fluctuations, providing a foundation for more accurate business decision-making. This article presents the application of AI at the Vietnam Petroleum Institute (VPI) to develop a digital analytics solution for collecting market information, aimed at improving the efficiency of data integration and automated processing, forecasting economic indicators, and providing a virtual assistant to support interactive market information queries.

Key words: Macroeconomics, Artificial intelligence, market analysis, data processing, virtual assistant.

DầuKhí
KHOA HỌC, CÔNG NGHỆ VÀ ĐỔI MỚI SÁNG TẠO PETROVIETNAM

NGHIÊN CỨU CHẾ TẠO THIẾT BỊ CẦM TAY PHÁT HIỆN KHUYẾT TẬT CỦA HỆ THỐNG ĐƯỜNG ỐNG BẰNG PHƯƠNG PHÁP RÒ RỈ ĐƯỜNG SỨC TỪ (MFL)

Phạm Hồng Quang, Vũ Minh Hùng, Phan Minh Quốc Bình, Nguyễn Ngọc Khương, Nguyễn Quang Vinh, Bùi Minh Thảo
Trường Đại học Dầu khí Việt Nam (PVU)
Email: quangph@pvu.edu.vn, hungvm@pvu.edu.vn
https://doi.org/10.47800/PVSI.2025.03-07

Tóm tắt

Nhằm đáp ứng nhu cầu kiểm tra định kỳ tình trạng sức khỏe các hệ thống đường ống đã đấu nối trong nhà máy (như nhà máy lọc dầu hay nhà máy đạm), nhóm tác giả đã nghiên cứu chế tạo thiết bị kiểm tra khuyết tật cầm tay bằng phương pháp rò rỉ đường sức từ (magnetic flux leakage - MFL). Thiết bị này không chỉ có khả năng kiểm tra nhanh, mà còn có chức năng tạo ảnh màu nhờ mảng cảm biến 2 chiều, cơ cấu quét 2 chiều, công nghệ màu hóa và tích hợp bản đồ, chưa từng có trên thế giới.

Thiết bị được lắp đặt 64 cảm biến Hall phẳng, bố trí thành mảng 2D, gắn trên đế cong phù hợp với bán kính đường ống. Mảng cảm biến và đế cong được lắp đặt trên khung có bánh xe cùng với hệ thống từ hóa. Hệ thống từ hóa gồm 3 khối liên kết qua cơ cấu tấm trượt, cho phép điều chỉnh để ôm khít các đường ống có đường kính khác nhau. Khi sử dụng, thiết bị được đẩy dọc theo đường ống.

Để tăng độ phân giải ảnh, sau mỗi lần đo (chụp ảnh), toàn bộ mảng cảm biến được dịch chuyển một bước 1 mm theo chiều dọc ống và chu vi ống. Các hình ảnh thu được sẽ được tích hợp, tạo thành hình ảnh có độ phân giải 1 mm.

Từ khóa: Thiết bị chụp ảnh từ, thiết bị kiểm tra khuyết tật đường ống đấu nối, MFL, kiểm tra không phá hủy bằng phương pháp từ, phương pháp rò rỉ đường sức từ.

1. Giới thiệu

Hệ thống đường ống trong các nhà máy chế biến dầu khí chủ yếu được chế tạo bằng thép, đấu nối cố định và đan xen phức tạp. Sau thời gian vận hành, do tác động của môi trường và đặc tính hóa học của sản phẩm vận chuyển, các đường ống bị ăn mòn, tiềm ẩn nguy cơ mất an toàn thiết bị. Do đó, cần thực hiện kiểm tra định kỳ bằng các phương pháp không phá hủy (NDT) để có biện pháp xử lý hoặc thay thế kịp thời [1, 2].

Tại một số đơn vị thuộc Tập đoàn Công nghiệp - Năng lượng Quốc gia Việt Nam (Petrovietnam), công tác kiểm tra khuyết tật đường ống bằng phương pháp không phá hủy được thực hiện thường xuyên. Đối với hệ thống đường ống dài, liên tục và có kích thước lớn, các đơn vị sử dụng phương pháp phóng thoi (pipeline inspection gauge - PIG). Đối với các đường ống ngắn, thẳng, rời, các đơn vị sử dụng thiết bị kiểm tra điện từ (electromagnetic inspection - EMI). Tại các nhà máy, việc kiểm tra hệ thống đường ống đã đấu nối thường sử dụng phương pháp siêu âm. Tuy nhiên, phương pháp này bị hạn chế bởi tốc độ kiểm tra chậm.

Xuất phát từ thực tế trên, nhóm tác giả đã nghiên cứu chế tạo thiết bị kiểm tra khuyết tật không phá hủy với tốc độ cao. Phương pháp được chọn là phương pháp rò rỉ đường sức từ (MFL) nhờ các ưu điểm là tốc độ kiểm tra nhanh, chi phí thấp. Đặc biệt, thiết bị được giới thiệu trong bài viết này có chức năng chụp ảnh từ, cho phép tạo hình ảnh trực quan về khuyết tật trong hệ thống đường ống.

2. Nguyên lý hoạt động của thiết bị

Phương pháp MFL có nhiều ưu điểm, đặc biệt là tốc độ đo nhanh [3 - 7]. Tuy nhiên, nhược điểm của phương pháp này là thiếu thông tin trực quan về kích thước và hình dạng của khuyết tật. Việc chế tạo được thiết bị chụp ảnh từ sẽ vừa phát huy ưu điểm tốc độ kiểm tra

Ngày nhận bài: 11/9/2024
Ngày đánh giá và sửa chữa: 11/9 - 2/11/2024
Ngày duyệt đăng: 2/11/2024

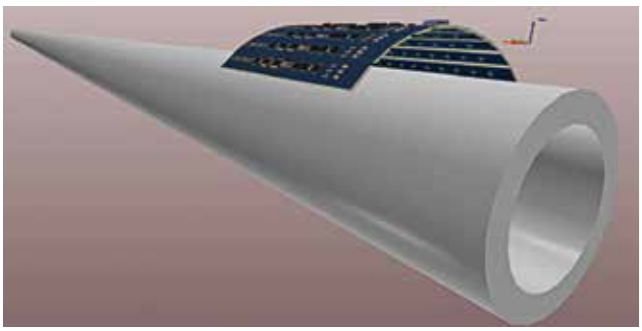
của MFL, vừa bổ sung chức năng tạo hình ảnh trực quan về khuyết tật.

Theo nguyên lý của phương pháp MFL, vật liệu sắt từ cần kiểm tra được từ hóa đến trạng thái gần bão hòa bằng nam châm điện hoặc nam châm vĩnh cửu. Tại vị trí có khuyết tật như khe nứt, vết ăn mòn, lỗ hổng, một phần đường sức từ trong vật liệu sẽ rò rỉ ra ngoài. Từ trường rò rỉ này được phát hiện bởi các cảm biến từ.

Cấu tạo của thiết bị gồm: hệ thống từ hóa, cơ cấu dịch chuyển, các cảm biến từ, mạch điện tử và hệ thống hiển thị, xử lý, lưu trữ dữ liệu. Thiết bị được thiết kế nhỏ gọn (kiểu cầm tay), dễ sử dụng để kiểm tra khuyết tật cho hệ thống đường ống đã đấu nối. Thiết bị có 2 chức năng là kiểm tra nhanh và chụp ảnh từ.

- Chức năng kiểm tra nhanh: Thiết bị được đẩy dọc theo ống với tốc độ lên đến 0,5 m/giây. Kết quả hiển thị dưới dạng đồ thị tín hiệu theo khoảng cách với 8 kênh độc lập, cho phép phát hiện vùng nghi ngờ có khuyết tật.

- Chức năng chụp ảnh từ: Thiết bị sử dụng 64 cảm biến Hall phẳng (PHR sensor) tạo thành mảng 2D để kiểm tra chi tiết các khu vực nghi ngờ có khuyết tật đã phát hiện bằng chức năng kiểm tra nhanh. Thiết bị dừng tại khu vực cần kiểm tra, thực hiện quét theo 2 chiều (dọc thân ống và theo chu vi) với khoảng quét 13 mm, bước quét 1 mm. Kết quả phân bố từ trường rò rỉ được hiển thị dưới dạng ảnh màu, thể hiện chính xác kích thước và hình dạng khuyết tật. Hình 1 mô tả nguyên lý của phương pháp chụp ảnh từ.



Hình 1. *Mô tả nguyên lý của phương pháp chụp ảnh từ.*

Các thông số cơ bản giữa thiết bị MFL cầm tay điển hình trên thị trường và thiết bị do nhóm tác giả PVU chế tạo được thể hiện trong Bảng 1.

3. Cơ sở thiết kế

Để làm cơ sở cho việc thiết kế, chế tạo, các nghiên cứu chi tiết để tối ưu hóa cấu hình, kích thước thiết bị đã được thực hiện. Cấu trúc 3 khối từ hóa: Khối từ hóa có chức năng tạo từ trường để từ hóa đối tượng cần kiểm tra.

- Kích thước ngang của khối từ hóa: Về nguyên tắc, khối từ hóa có kích thước ngang càng lớn sẽ có năng suất kiểm tra càng cao. Tuy nhiên, khung từ hóa lớn sẽ tăng khối lượng của thiết bị, gây khó khăn cho người sử dụng. Nhóm tác giả lựa chọn chiều ngang của thiết bị là 150 mm, được chia làm 3 khối, mỗi khối 50 mm. Việc chia làm 3 khối tạo cơ chế thay đổi độ ôm, thích ứng với đường ống có đường kính khác nhau. Chiều ngang 150 mm phù hợp với kích thước 100 x 100 mm của mảng cảm biến 2D.

- Độ dày gông từ, về nguyên tắc phải lớn hơn độ dày thành ống. Với độ dày của các ống từ 5 - 10 mm, nhóm tác giả lựa chọn độ dày gông từ là 15 mm.

- Chiều dài của gông từ là kích thước được tính toán cẩn thận do vừa phải tạo từ trường từ hóa đủ lớn để tạo tín hiệu, vừa phải đảm bảo vùng đồng nhất về từ trường cho mảng cảm biến 2D. Việc nghiên cứu chọn chiều dài gông từ đã được thực hiện qua mô phỏng bằng phương pháp phần tử hữu hạn (FEM) cho các độ dài 300 mm, 350 mm và 400 mm.

Phương pháp phần tử hữu hạn (finite element method - FEM) là phương pháp số dùng để phân tích các phương trình đạo hàm riêng trên miền xác định có hình

Bảng 1. *So sánh các thông số cơ bản giữa thiết bị MFL cầm tay thương phẩm và thiết bị của PVU chế tạo*

Thông số	Thiết bị MFL thương mại [8, 9]	Thiết bị MFL PVU chế tạo
Chức năng kiểm tra nhanh	Có	Có
Chức năng chụp ảnh từ	Không	Có
Công nghệ	MFL	MFL
Phương pháp dịch chuyển	Đẩy tay	Đẩy tay
Loại cảm biến	Hall thường	Hall phẳng (PHR)
Độ phân giải cao nhất	7 mm	1 mm
Khả năng phát hiện khuyết tật	30% pitting với ống dày 6 mm	20% pitting với ống dày 8 mm
Khả năng phát hiện khuyết tật qua lớp phủ	Dày 3 mm	Dày 3 mm
Đường kính ống tối thiểu	127 mm	114 mm

dạng và điều kiện biên phức tạp, mà không thể giải bằng phương pháp giải tích [7, 8]. Trong số các phần mềm ứng dụng phương pháp phần tử hữu hạn, ANSYS là công cụ phổ biến nhất để phân tích từ trường (mật độ đường sức từ, cường độ từ trường và thế vector từ) và các đặc trưng điện từ cơ bản (độ từ cảm, lực điện từ).

Phần mềm ANSYS được thiết kế để giải các phương trình Maxwell và có khả năng mô phỏng các loại trường điện từ (điện trường, từ trường, dòng xoáy hay nam châm vĩnh cửu). Phương pháp FEM mô tả chính xác trường điện từ và dễ dàng khi thay đổi cấu trúc hình học cũng như các thông số đầu vào. Trong nghiên cứu này, nhóm tác giả sử dụng phần mềm ANSYS 2021R1, trong đó ANSYS Maxwell được áp dụng để giải các bài toán điện từ.

Vì hệ từ hóa gồm 3 gông từ giống nhau, nhóm tác giả chỉ thực hiện mô phỏng trên 1 gông từ. Cấu trúc hệ từ hóa của 1 gông từ cùng mẫu kiểm tra được thể hiện trong Hình 2. Ngoài ra, sẽ có một kết quả mô phỏng cho 3 khối ở trạng thái đã kết nối.

Một số kết quả sự phân bố từ trường điển hình của gông từ với độ dài khác nhau và tín hiệu MFL được thể hiện trong Hình 3 - 6.

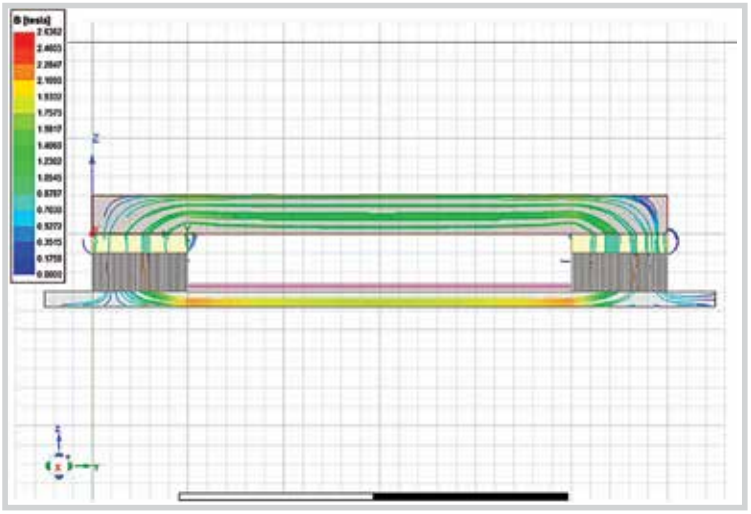
Kết quả mô phỏng như sau:

- Giá trị cảm ứng từ By tại tâm tấm thép giảm khi độ dài gông tăng, cụ thể là 1,85 T, 1,78 T và 1,74 T tương ứng độ dài gông từ 300 mm, 350 mm và 400 mm. Các giá trị này đủ đạt trạng thái gần bão hòa, đảm bảo gây tín hiệu MFL.

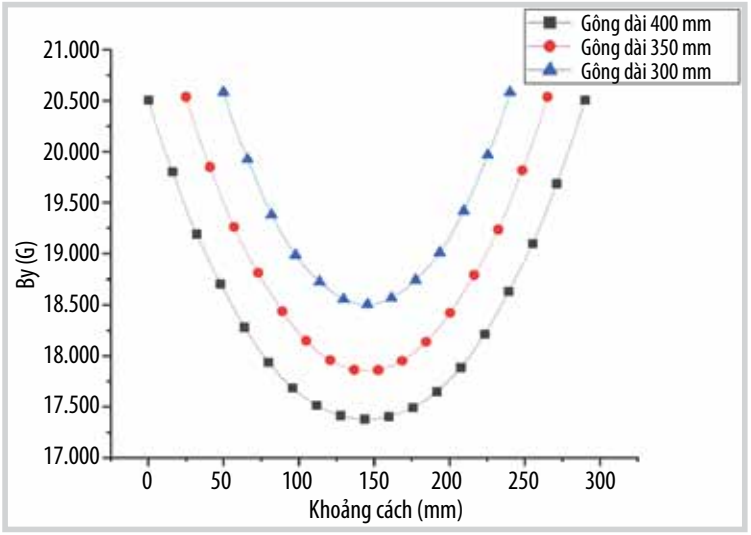
- Sự phân bố By bên ngoài tấm có dạng hình lòng chảo. Độ phẳng của lòng chảo tăng khi độ dài gông từ tăng. Xét trong giới hạn 100 mm quanh tâm (vùng đặt cảm biến đo), độ lệch By ở mép và ở tâm là 59 G, 27 G và 18 G tương ứng gông dài 300 mm, 350 mm và 400 mm. Như vậy, nên chọn cấu hình L = 400 mm để đảm bảo sự đồng nhất.

- Sự đồng nhất của từ trường tạo bởi 3 gông từ theo phương ngang ở vùng giữa của gông từ khá tốt.

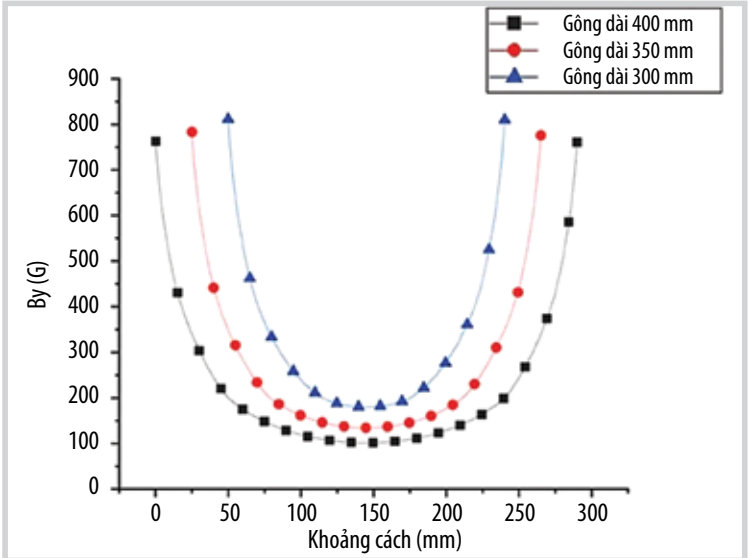
- Các khuyết tật đã tạo nên các tín hiệu MFL. Tín hiệu MFL cho thành phần song song có dạng 1 cực đại, độ lớn tăng mạnh theo độ sâu



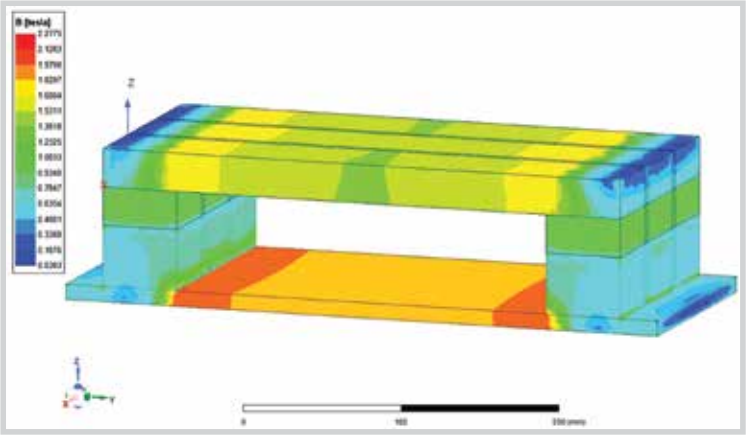
Hình 2. *Mô hình mô phỏng FEM cho 1 gông từ.*



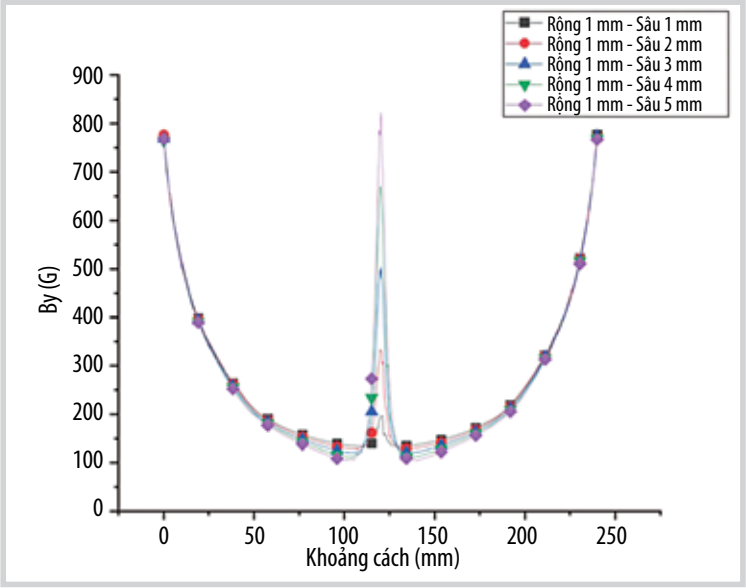
Hình 3. *Phân bố từ trường By (thành phần song song với tấm thép) bên trong tấm thép với các gông từ có độ dài khác nhau 300 mm, 350 mm và 400 mm.*



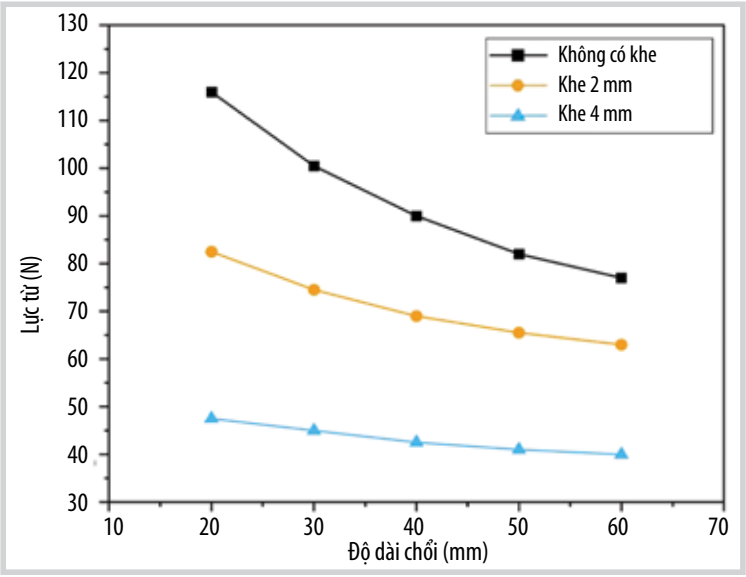
Hình 4. *Đồ thị phân bố từ trường thành phần song song (By) ở khoảng cách 3 mm trên bề mặt tấm thép của cấu hình gông từ dài 300 mm, 350 mm và 400 mm, nam châm dày 25 mm, chổi từ dài 40 mm (15 mm đế và 25 mm sợi thép).*



Hình 5. *Hình ảnh màu phân bố từ trường của bộ 3 gông từ, đặt cách nhau 5 mm.*



Hình 6. *Đồ thị tổng hợp tín hiệu MFL cho các khuyết tật rộng 1 mm thay đổi độ sâu từ 1 - 5 mm.*



Hình 7. *Kết quả mô phỏng lực hút giữa cực từ và tấm thép đối với các khe hở không khí từ 0 - 4 mm và chiều dài chổi từ 20 - 60 mm.*

khuyết tật. Với khuyết tật sâu 1 mm, tín hiệu cỡ vài chục Gauss. Các cảm biến Hall phẳng hoàn toàn có thể phát hiện trong kiểu đo nhanh.

Độ dài chổi từ: Để giảm lực hút giữa cực từ và bề mặt ống, nhóm tác giả sử dụng lớp chổi thép từ tính làm đệm. Kết quả mô phỏng lực hút giữa cực có chổi và tấm thép được thể hiện trong Hình 7 cho 3 trường hợp không có khe hở không khí, khe hở không khí bằng 2 mm và khe hở không khí bằng 4 mm. Kết quả cho thấy lực hút giảm đi nhiều lần khi có thêm một lớp chổi từ tính. Ví dụ, nếu chổi từ dài 40 mm, lực hút cho trường hợp cực từ dính sát tấm thép là 85 N và đối với trường hợp có khe hở không khí 2 mm, lực hút là 70 N. Xét trường hợp khe hở không khí 2 mm, nếu tính toán cho cả 6 cực từ, tổng lực hút là 280 N, đủ để giữ thiết bị có khối lượng khoảng 20 - 25 kg.

Kết quả đo lực hút giữa gông từ và tấm thép cho thấy lực hút của cực từ lần lượt là 120 N, 60 N và 32 N đối với khoảng cách 0 mm, 2 mm và 4 mm giữa cực từ và tấm thép. Từ kết quả này, nhóm tác giả đề xuất độ dài chổi từ tối ưu là 40 mm.

Kích thước nam châm: Kích thước nam châm phải đảm bảo đủ tạo ra mật độ đường sức từ đủ lớn, nhưng không quá lớn làm ảnh hưởng kích thước chung của thiết bị. Ngoài ra, kích thước nam châm phải phù hợp với loại nam châm có sẵn trên thị trường. Chiều ngang nam châm nên bằng 50 mm, đúng bằng chiều rộng gông từ. Chiều dọc nam châm, về nguyên tắc phải lớn hơn độ dày mẫu kiểm tra vài lần. Về chiều dày, độ mạnh của nam châm không tăng tuyến tính theo độ dày nam châm mà tuân theo công thức:

$$B_x = \frac{B_r}{\pi}\left(tan^{-1}\frac{AB}{2x\sqrt{4x^2+A^2+B^2}} - tan^{-1}\frac{AB}{2(2L+x)\sqrt{4(2L+x)^2+A^2+B^2}}\right)$$

Trong đó:

B_r: Cảm ứng từ dư của nam châm;

A, B, và L: Độ dài, rộng và dày của nam châm;

x: Khoảng cách từ mặt nam châm.

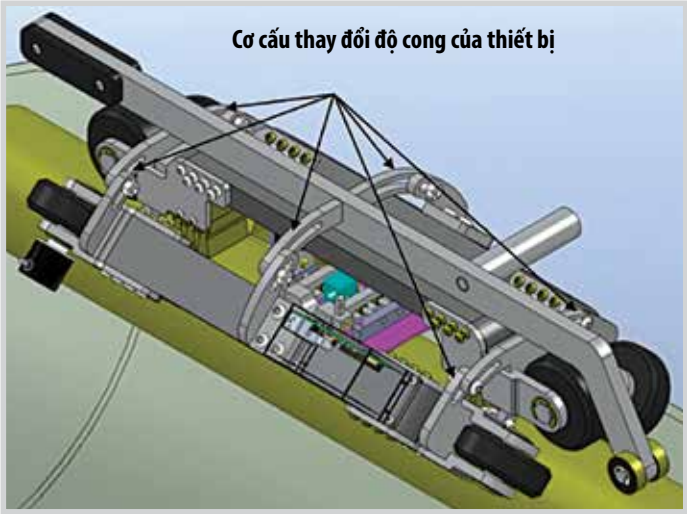
Với các tính toán trên, nhóm tác giả chọn bề dày nam châm là 25 mm. Kích thước nam châm sẽ là 50 mm x 50 mm x 25 mm.

Cơ cấu thích nghi độ cong ống kiểm tra: Do thiết bị làm việc với các ống có đường kính khác nhau nên thiết bị cần có cơ cấu thay đổi độ cong. Nhóm tác giả thiết kế cơ cấu điều chỉnh kiểu dẻ quạt với 6 cặp tấm trượt, 4 cặp ở 2 đầu và 2 cặp ở giữa. Đây là thiết kế mới có nhiều ưu điểm thay thế cho kiểu ốc vít mà các sản phẩm thương mại sử dụng. Các sản phẩm thương mại chỉ thiết kế cho các ống thay đổi trong phạm vi hẹp. Thiết bị được thiết kế cho các ống có đường kính thay đổi từ 4½ inch đến 20 inch. Nếu áp dụng kiểu ốc vít sẽ cần thêm bậc tự do, dẫn đến cơ cấu phức tạp, khó căn chỉnh và không đảm bảo độ chắc chắn. Hình 8 mô tả cơ cấu thay đổi độ cong của thiết bị.

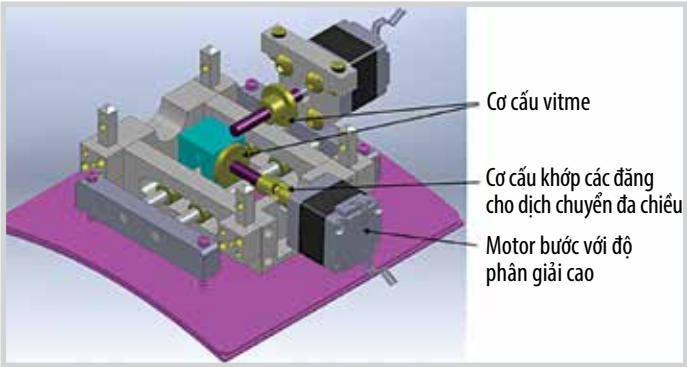
Kích thước mảng cảm biến và số lượng cảm biến: Hệ cảm biến bao gồm 8 x 8 = 64 cảm biến gắn thành mảng hình vuông 100 mm x 100 mm. Trước hết, 8 cảm biến được gắn vào một thanh PCB, sau đó 8 thanh được gắn vào một tấm nhựa dẻo, cuối cùng tấm nhựa dẻo được dán vào một guốc nhôm có độ cong phù hợp với loại ống cần kiểm tra. Bề mặt guốc cảm biến được phủ bằng một lớp vật liệu phi từ tính để bảo vệ cảm biến. Guốc cảm biến sẽ gắn với một hệ cơ khí dịch chuyển 2 chiều dọc ống và theo chu vi thành ống. Việc chọn số lượng cảm biến được quyết định bởi quy mô của nghiên cứu, tốc độ ghi nhận và xử lý tín hiệu đo của các mạch điện tử hiện hành cũng như tổng công suất điện tiêu thụ. Kích thước 100 mm x 100 mm của mảng cảm biến được tính từ kích thước ngang tổng cộng của hệ là 150 mm. Như vậy, mảng cảm biến sẽ nằm trong vùng đồng nhất của từ trường, cả theo chiều dọc và ngang.

Cơ chế chuyển động 2 chiều và tạo hình ảnh: Thiết bị sử dụng cơ chế dịch chuyển vitme 2 chiều được truyền động từ 2 động cơ bước. Cơ chế dịch chuyển dọc ống là vitme thẳng. Cơ chế dịch chuyển theo phương chu vi sử dụng trục mềm để chuyển từ chuyển động thẳng của vit me sang chuyển động cong, cho phép giá đỡ mảng cảm biến trượt trên các rãnh cong.

Chu trình quét của mảng cảm biến được thực hiện theo trình tự sau: (i) từ vị trí ban đầu, mảng cảm biến dịch chuyển 1 mm theo phương chu vi; (ii) dịch chuyển 13 bước (mỗi bước 1 mm) dọc theo thân ống; (iii) dịch chuyển tiếp 1 mm theo phương chu vi; (iv) dịch chuyển 13 bước ngược lại dọc theo thân ống. Quá trình quét hoàn thành khi mảng cảm biến đã dịch chuyển đủ 13 bước theo phương chu vi, mảng cảm biến sẽ tự động trở về vị trí ban đầu.



Hình 8. *Cơ cấu thay đổi độ cong để thích nghi đường ống có đường kính khác nhau.*

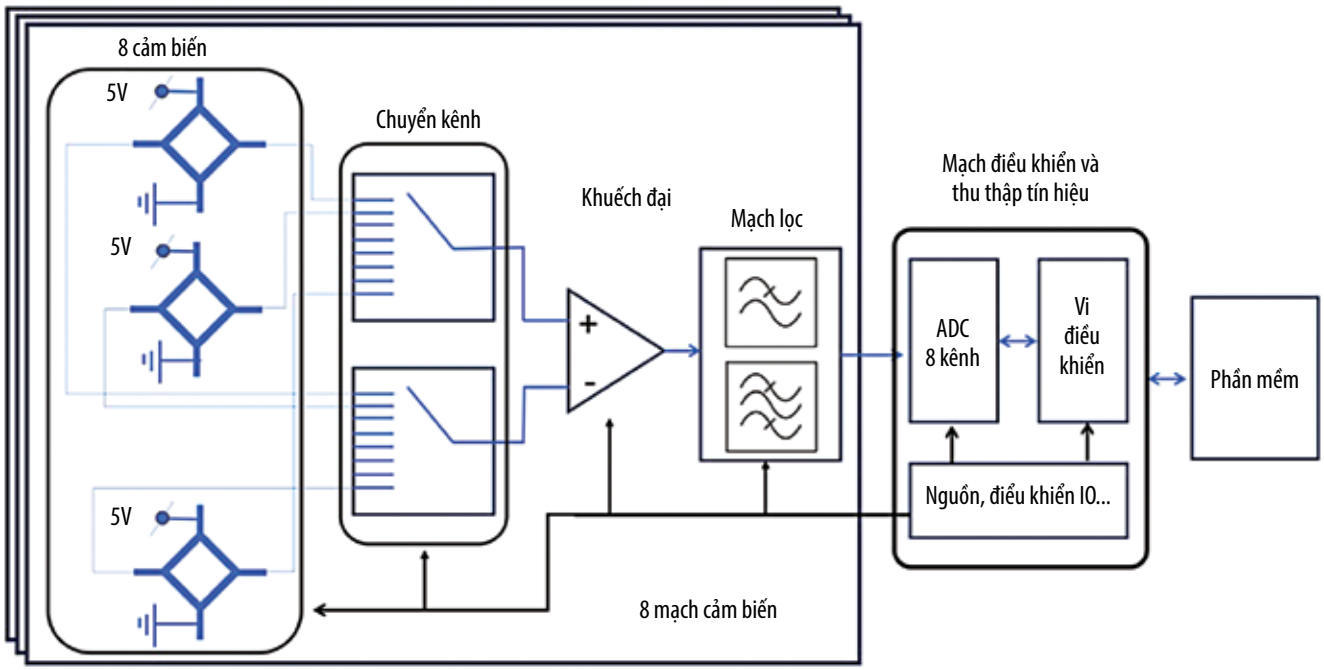


Hình 9. *Cơ cấu dịch chuyển mảng cảm biến 2 chiều.*

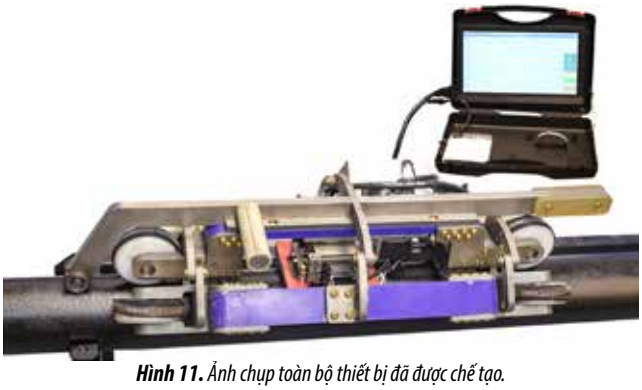
Tại mỗi vị trí, hệ thống ghi lại kết quả phân bố từ trường dưới dạng 1 ảnh gồm 64 ô vuông (mỗi ô có kích thước 1 mm x 1 mm, cách nhau 13 mm). Tổng cộng 13 x 13 = 169 hình ảnh được tích hợp thành 1 hình ảnh duy nhất kích thước 104 mm x 104 mm, phản ánh chính xác kích thước và hình dạng khuyết tật của đường ống. Hình 9 mô tả cơ chế dịch chuyển 2 chiều cùng mảng cảm biến.

Mạch điện tử: Tín hiệu tương tự thu được từ cảm biến từ tính Hall phẳng được khuếch đại, lọc nhiễu qua mạch lọc dải tần số từ 0,01 - 1.000 Hz. Mạch tín hiệu được chia làm 8 thanh, mỗi thanh cho một dãy 8 cảm biến. Tín hiệu của mỗi thanh sẽ được gửi đến một kênh chuyển đổi tín hiệu tương tự sang tín hiệu số (analog-to-digital, ADC). Như vậy 8 thanh cảm biến sẽ cần 8 kênh ADC để chuyển tín hiệu của cả mảng cảm biến lên vi điều khiển. Để giảm nhiễu sau khuếch đại tín hiệu của cảm biến, nhóm tác giả sử dụng loại mạch lọc với các thông số như dải tần số làm việc, số mắt lọc, độ phẩm chất của mạch lọc đã tối ưu hóa để lọc lựa tín hiệu tương tự trước khi gửi lên máy tính. Sơ đồ của mạch điện tử được thể hiện trong Hình 10.

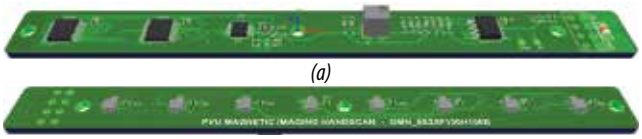
Vi điều khiển nhận tín hiệu từ 8 kênh ADC theo phương thức đồng thời hoặc tuần tự, đảm bảo tốc độ truyền tín hiệu



Hình 10. *Sơ đồ mạch điện tử cho hệ 64 cảm biến.*



Hình 11. *Ảnh chụp toàn bộ thiết bị đã được chế tạo.*



(a)

(b)

Hình 12. *Ảnh chụp mặt trên (a) và mặt dưới (b) của 1 thanh cảm biến gồm 8 cảm biến Hall phẳng.*

từ 64 cảm biến liên tục với tần số > 100 lần/giây. Để đáp ứng yêu cầu này, theo tính toán của nhóm tác giả, tốc độ truyền dữ liệu của chip ADC 8 nhân cần đạt từ 200 - 500 kBit/s.

Để tối ưu tốc độ truyền, nhóm tác giả sử dụng phương thức mã hóa tín hiệu theo hệ lục phân (hex). Tốc độ truyền số liệu cao sẽ gây ra sai số về biên độ và pha của tín hiệu nhận được trên máy tính do ảnh hưởng nhiễu của môi trường đo. Do đó, tín hiệu thô nhận được trên máy tính cần phải lọc lựa, loại bỏ nhiễu của môi trường để phép đo được chuẩn xác. Dựa trên điều kiện đo thực tế, nhóm tác giả đề xuất sử dụng 3 thuật toán lọc nhiễu bao gồm: bộ lọc thông thấp, bộ lọc thông cao và bộ lọc cấm dải (notch filter).

4. Kết quả chế tạo và thử nghiệm thiết bị kiểm tra khuyết tật cầm tay bằng phương pháp rò rỉ đường sức từ

Hệ cơ khí: Hình 11 là toàn bộ thiết bị đã được chế tạo.

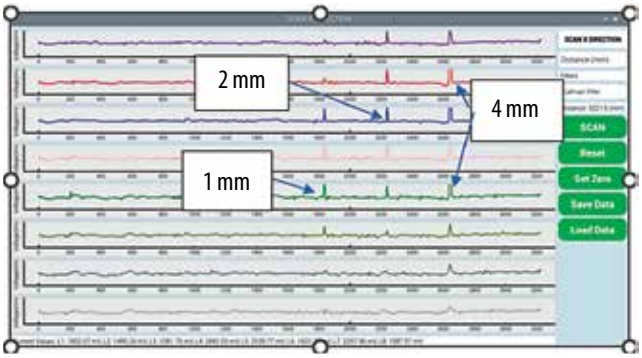
Mạch điện tử: Hình 12 là ảnh chụp mặt trên (a) và mặt dưới (b) của một thanh cảm biến gồm 8 cảm biến Hall phẳng. Mặt trên của thanh cảm biến gồm bộ chuyển kênh vi sai, khuếch đại vi sai và mạch lọc. Mặt dưới của thanh cảm biến bao gồm 8 cảm biến, khoảng cách giữa các cảm biến là 13 mm.

Hình 13 là ảnh chụp mạch thu thập tín hiệu bao gồm ADC 8 kênh (tốc độ đọc tín hiệu 500 kbps), bộ ổn áp 3.3 V và các chân thu nhận tín hiệu, các chân điều khiển bộ chuyển kênh. Mạch được thiết kế để tương thích với vi điều khiển của máy tính Raspberry PI 5. Hình 14 là ảnh chụp mảng cảm biến đã được gắn vào guốc cong.

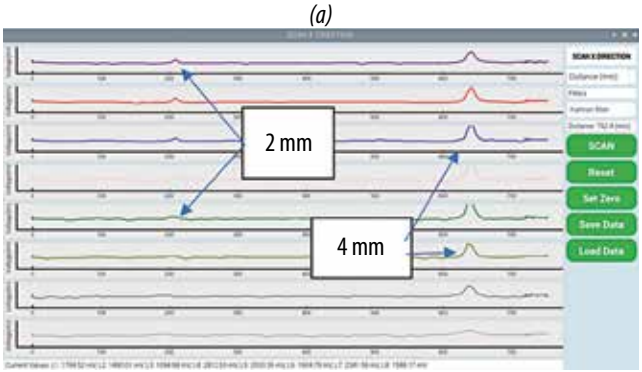
Các mẫu ống có khuyết tật: Nhóm tác giả chế tạo 2 loại ống có các khuyết tật nhân tạo, trong đó 1 loại ống có đường kính 4½ inch, dài 3.000 mm, dày 8,56 mm và 1 loại ống có đường kính 20 inch, dài 6.000 mm, dày 9,5 mm. Đây cũng là 2 loại ống có kích thước nhỏ nhất và lớn nhất đối với phạm vi của thiết bị. Trên các ống được chế tạo các khuyết tật nhân tạo dạng khe có độ rộng 1 mm, dài 40 mm và các độ sâu 1 mm, 2 mm, 4mm, cách nhau 400 mm; 2 khuyết tật dạng khe có độ sâu 2 mm và 4 mm

Hình 13. *Ảnh chụp mạch thu thập tín hiệu.*

Hình 14. *Ảnh chụp mảng cảm biến đã được gắn vào guốc cong (chưa che lớp bảo vệ).*



(a)



(b)

Hình 15. *Kết quả đo nhanh trên 8 cảm biến qua các khuyết tật dạng khe từ trái qua phải sâu 1 mm, 2 mm và 4 mm (a), ống có lớp bọc dày 3 mm với 2 khuyết tật khe sâu 2 mm và 4 mm (b).*

ở bên trong ống, cách nhau 420 mm cũng được chế tạo. Vật liệu làm ống là loại thép đúc theo tiêu chuẩn ASTM A106 có hàm lượng carbon 0,2% và độ từ thẩm trung bình khoảng 1.000. Để kiểm tra khả năng của thiết bị khi kiểm tra ống có lớp phủ bên ngoài, tác giả bọc bên ngoài một

Hình 16. *Kết quả chụp ảnh từ trên khuyết tật khe sâu 1 mm.*

đoạn ống một lớp vật liệu sợi nhựa nhôm dày 3 mm. Lưu ý là mọi loại vật liệu phi từ như không khí, nhôm, nhựa đều có độ từ thẩm bằng 1 nên có ảnh hưởng như nhau đến tín hiệu đo.

Kết quả đo bằng chức năng quét nhanh: Phép đo nhanh được thực hiện bằng cách đẩy thiết bị dọc ống. Kết quả phép đo nhanh được thể hiện trong Hình 15a và 15b ở đó có 8 đường đồ thị tín hiệu phụ thuộc vị trí ứng với 8 cảm biến. Sự xuất hiện các khuyết tật được thể hiện bằng các đỉnh nhọn trên đồ thị.

- Kết quả đo nhanh đã phát hiện được các khuyết tật sâu từ 1 mm trở lên mặc dù độ lớn từ trường MFL cực nhỏ (chỉ vài Gauss).

- Ống có lớp bọc dày 3 mm, thiết bị có thể phát hiện khuyết tật 2 mm trở lên

Kết quả đo bằng chức năng tạo ảnh: Kết quả đo bằng chức năng tạo ảnh được thể hiện trong Hình 16. Kết quả cho thấy trên nền màu vàng ứng với từ trường nền có 1 vùng màu đỏ ứng với từ trường rò rỉ gây bởi 1 khuyết tật dạng khe nằm ngang sâu 1 mm.

- Kết quả quét 2D tạo ảnh màu đã thực hiện thành công với tốc độ 1 phút 30 giây trên các khuyết tật với độ phân giải 1 mm. Ảnh màu phản ánh chính xác sự phân bố từ trường MFL gây bởi khuyết tật, cho hình ảnh trực quan của khuyết tật.

- Việc tăng hệ số khuếch đại kết hợp thuật toán lọc nhiễu thích hợp có thể tăng khả năng phát hiện khuyết tật nhỏ.

5. Kết luận

Thiết bị kiểm tra khuyết tật đường ống kiểu cầm tay sử dụng phương pháp rò rỉ đường sức từ (MFL) đã được nhóm tác giả chế tạo và thử nghiệm thành công. Thiết bị có chức năng chụp ảnh màu được thực hiện bởi mảng cảm biến 2D kết hợp cơ cấu quét 2 chiều, công nghệ màu

hóa và tích hợp bản đồ, tạo ra hình ảnh trực quan về kích thước và hình dạng khuyết tật. Đây là giải pháp mới so với các thiết bị MFL thương mại hiện nay.

Ngoài ra, thiết bị còn có các ưu điểm: cơ cấu thích nghi cho phép kiểm tra đường ống có đường kính khác nhau; có độ nhạy cao và tiêu thụ điện năng thấp nhờ sử dụng cảm biến Hall phẳng; khả năng tạo ảnh với độ phân giải 1 mm nhờ công nghệ quét bước nhỏ. Thiết bị đã đạt các tiêu chuẩn kỹ thuật đề ra và nhóm tác giả sẽ tập trung hoàn thiện sản phẩm và mở rộng khả năng ứng dụng của thiết bị.

Tài liệu tham khảo

[1] D.C. Jiles, "Review of magnetic methods for nondestructive evaluation (Part 2)", *NDT International*, Volume 23, Issue 2, pp. 83 - 92, 1990. DOI: 10.1016/0308-9126(90)91892-W.

[2] D.L. Atherton, "Magnetic inspection is key to ensuring save pipelines", *Oil and Gas Journal*, Volume 87, Issue 2, pp. 52 - 61, 1989. DOI: 10.1016/0963-8695(93)90206-a.

[3] Yong Zhang, Zhongfu Ye, and Chong Wang, "A fast method for rectangular crack sizes reconstruction in magnetic flux leakage testing", *NDT & E International*, Volume 42, Issue 5, pp. 369 - 375, 2009. DOI: 10.1016/j.ndteint.2009.01.006.

[4] H.A.Vanaei, A. Eslami, and A. Egbewande, "A review on pipeline corrosion, in-line inspection (ILI), and corrosion growth rate models", *International Journal of Pressue Vessels Piping*, Volume 149, pp. 43 - 54, 2017. DOI: 10.1016/j.ijpvp.2016.11.007.

[5] Qingshan Feng, Rui Li, Baohua Nie, Shucong Liu, Lianyu Zhao, and Hong Zhang, "Literature review: Theory and application of in-line inspection technologies for oil and gas pipeline girth weld defection", *Sensors*, Volume 17, Issue 1, 2017. DOI: 10.3390/s17010050.

[6] Eddyfi Technologies, "Pipescan HD". [Online]. Available: https://www.eddyfi.com/en/product/pipescan-hd.

[7] Silverwing, "Pipescan - Adjustable magnetic flux leakage pipe scanner". [Online]. Available: https://www.ndt-instruments.com/wp-content/uploads/2018/01/pipescan-mfl-pipe-inspection.pdf.

[8] Slideshare, "Ansoft maxwell 3D field simulator v11 user's guide". [Online]. Available: https://www.slideshare.net/EraBrown/ansoft-maxwell-3d-v11-user-guide.

[9] Huang Zuoying, Que Peiwen, and Chen Liang, "3D FEM analysis in magnetic flux leakage method", *NDT & E International*, Volume 39, Issue 1, pp. 61 - 66, 2006. DOI: 10.1016/j.ndteint.2005.06.006.

RESEARCH AND MANUFACTURE OF PORTABLE DEVICE TO DETECT PIPELINE DEFECTS USING THE MAGNETIC LEAKAGE FLUX METHOD

Pham Hong Quang, Vu Minh Hung, Phan Minh Quoc Binh, Nguyen Ngoc Khuong, Nguyen Quang Vinh, Bui Minh Thao
Petrovietnam University (PVU)
Email: quangph@pvu.edu.vn, hungvm@pvu.edu.vn

Summary

To meet the needs of periodic health inspections of interconnected pipeline systems in factories such as refineries and ammonia plants, the research team developed a portable defect inspection device using the magnetic flux leakage (MFL) method. This device not only enables rapid inspection, but also features color imaging capability through a two-dimensional sensor array, 2D scanning mechanism, colorization technology and integrated mapping - features not previously available worldwide. The device is equipped with 64 flat Hall sensors arranged on a 2D array mounted on a curved base that conforms to the pipe's radius. The sensor array and curved base are installed on a wheeled frame integrated with the magnetization system. The magnetization system consists of 3 blocks connected by a sliding plate mechanism, allowing adjustment to tightly fit pipes of different diameters. During operation, the device is manually pushed along the pipeline.

To enhance image resolution, the entire sensor is shifted by 1 mm along both the longitudinal and circumferential directions of the pipe after each measurement (image capture). The captured images are then integrated to produce a final image with 1 mm resolution.

Key words: Magnetic imaging device, pipeline defect inspection, MFL method, non-destructive magnetic testing, magnetic flux leakage method.

DầuKhí
KHOA HỌC, CÔNG NGHỆ VÀ ĐỔI MỚI SÁNG TẠO PETROVIETNAM

NGHIÊN CỨU CHẾ TẠO ROBOT PHÁT HIỆN ĂN MÒN BỒN NỔI HÌNH TRỤ ĐỨNG CHỨA XĂNG DẦU

Nguyễn Thị Lan[1, 2], Trần Dương Tấn Quyền[1], Phạm Quốc Huy[3], Nguyễn Tấn Tiến[1, 4], Lê Văn Sỹ[2, 5]
[1]Trường Đại học Bách khoa - Đại học Quốc gia Tp. Hồ Chí Minh
[2]Trường Đại học Dầu khí Việt Nam (PVU)
[3]Trường Cao đẳng Dầu khí (PV College)
[4]Phòng Thí nghiệm trọng điểm Quốc gia Điều khiển số và Kỹ thuật hệ thống (DCSELab)
[5]Tổng công ty Bảo dưỡng - Sửa chữa Công trình Dầu khí - CTCP (PVMR)
Email: ntlan.sdh231@hcmut.edu.vn
https://doi.org/10.47800/PVSI.2025.03-08

Tóm tắt

Bài viết giới thiệu kết quả nghiên cứu, thiết kế và chế tạo nguyên mẫu robot phát hiện ăn mòn cho thân bồn nổi hình trụ đứng chứa xăng dầu, dựa trên phương pháp kiểm tra rò rỉ đường sức từ (magnetic flux leakage - MFL). Thiết kế tổng thể, hệ thống động lực, điện, điều khiển và bộ kiểm tra rò rỉ từ đã được tính toán, thiết kế tối ưu bằng phần mềm chuyên dụng, đảm bảo các yêu cầu về an toàn cháy nổ theo quy định của IEC. Sau khi thiết kế, chế tạo, nguyên mẫu robot có trọng lượng 39 kg, vận tốc di chuyển từ 10 - 50 mm/giây.

Nguyên mẫu robot sau khi lắp ráp hoàn chỉnh đã được thử nghiệm tại xưởng trên mô hình tấm thép đứng, mô phỏng điều kiện làm việc thực tế của các bồn chứa xăng dầu. Kết quả thử nghiệm cho thấy robot đáp ứng các chỉ tiêu quan trọng: khả năng bám dính trên bề mặt thẳng đứng; di chuyển ổn định qua đường hàn có chiều cao tối đa 10 mm ở vận tốc tối đa 50 mm/giây; có khả năng phát hiện khuyết tật kích thước nhỏ với độ sâu 2 mm trên tấm thép dày đến 14 mm.

Từ khóa: Kiểm tra ăn mòn, kiểm tra không phá hủy, phương pháp rò rỉ từ thông, robot leo bồn chứa.

1. Giới thiệu

Trong lĩnh vực tồn trữ xăng dầu, sử dụng các phương pháp phù hợp trong kiểm tra ăn mòn bồn chứa là rất quan trọng để bảo vệ thiết bị, nhân sự, môi trường và cộng đồng xung quanh ngoài việc giảm thời gian ngừng hoạt động và chi phí bảo trì của các bể chứa.

Trong phương pháp kiểm tra thủ công, người kiểm tra tiếp cận bề mặt cần đánh giá với sự hỗ trợ của hệ thống giàn giáo quy mô lớn. Quá trình đánh giá tình trạng các bể chứa được thực hiện theo hướng dẫn của API 653 [1], với sự hỗ trợ của thiết bị siêu âm cầm tay. Đánh giá này gồm thông tin chung về bể chứa, đánh giá hiện tượng lún và đo mức độ ăn mòn và độ dày của vỏ, đáy, mối hàn và các bộ phận khác của bể chứa.

Công việc này mất nhiều thời gian và nguy hiểm đến tính mạng con người. Thêm vào đó, phương pháp kiểm tra truyền thống chỉ kiểm tra được ở những vị trí dưới đáy bồn và phần thân kề đáy bồn. Đối với phần trên cao của thân bồn, việc kiểm tra sẽ khó khăn và càng trở nên nguy hiểm hơn. Do vậy, việc tự động hóa quá trình kiểm tra bồn bể bằng robot nhằm tiết kiệm thời gian, tăng năng suất và đảm bảo an toàn cho con người là hết sức cần thiết.

Một số nước phát triển đã đưa vào vận hành các robot hiện đại phục vụ kiểm tra tự động trên các bồn nổi chứa xăng dầu như Neptune [2], Maverick [3] và Robavatar [4]. Các robot này được trang bị các module phát hiện ăn mòn dựa trên công nghệ kiểm tra không phá hủy (NDT) [5]. Các kỹ thuật NDT phổ biến trong các vật liệu thép là kiểm tra siêu âm (ultrasonic testing - UT); kiểm tra phát xạ âm thanh (acoustic emission testing - AE); kiểm tra dòng điện xoáy (eddy currents testing - ECT); kiểm tra hạt từ (magnetic particle testing - MT) và kiểm tra rò rỉ từ thông (magnetic flux leakage testing - MFL) [5].

Trong ứng dụng kiểm tra ăn mòn cấu kiện thép từ tính, phương pháp kiểm tra rò rỉ từ thông MFL được đánh giá là thích hợp hơn so với phương pháp kiểm tra NDT khác, nhờ

Ngày nhận bài: 4/9/2024
Ngày đánh giá và sửa chữa: 4/9 - 27/11/2024
Ngày duyệt đăng: 27/11/2024

các ưu điểm sau: Phương pháp này không đòi hỏi quá trình chuẩn bị (pre-processing) phức tạp, dễ thu nhận tín hiệu, dễ thực hiện trực tuyến quá trình đo, phát hiện đa dạng các loại khuyết tật như khuyết tật bề mặt, lỗ rỗng, sẹo, vết nứt, ăn mòn bên trong và bên ngoài bề mặt, với khả năng phát hiện khuyết tật rất nhỏ, trên bề dày tấm thép lớn. Điều này đã được minh chứng rõ trong nghiên cứu của Li và cộng sự [6], theo đó, nhóm tác giả đã đề xuất một đầu dò từ độ nhạy cao được sử dụng trong MFL để phát hiện các vết nứt nhỏ, với mẫu được thử nghiệm là thép kết cấu carbon Q235A và các khuyết tật có chiều dài 6 mm, chiều rộng từ 60 - 80 µm và chiều sâu từ 7 - 60 µm. Kết quả thí nghiệm cho thấy tất cả các khuyết tật đều được phát hiện.

Chính vì vậy, kỹ thuật kiểm tra MFL đã được áp dụng phổ biến trong kiểm tra ăn mòn trên các hệ thống đường ống dẫn khí đốt [7 - 11]. Từ tín hiệu MFL thu được, việc diễn giải các đặc trưng của mẫu tín hiệu giúp đánh giá kích thước cơ bản (dài, rộng, sâu), thậm chí là hình dáng khuyết tật, đây là chủ đề nghiên cứu được thảo luận nhiều trong những năm gần đây [12 - 17].

Về nghiên cứu phát triển robot leo bồn, nhiệm vụ quan trọng nhất là phát triển một cơ chế thích hợp để đảm bảo robot bám chặt vào thành bồn ở các bề mặt có độ cong khác nhau một cách đáng tin cậy mà không làm mất khả năng di chuyển và tính cơ động của nó.

Phương pháp bám dính - di chuyển thường được sử dụng nhất trong phát triển robot leo tường đứng là kết hợp nam châm vĩnh cửu với cơ cấu bánh xích [18 - 21]. Kiểu kết cấu vững chắc và diện tích tiếp xúc lớn đã làm tăng tính ổn định của robot, nhưng lại thiếu sự linh động và khó thích ứng với các bề mặt không bằng phẳng hoặc cong. Để khắc phục hạn chế này, một số nhóm nghiên cứu đã thêm hệ thống treo vào cơ cấu bánh xích, cho phép uốn cong phần bánh xích và thích ứng với các bề mặt cong mà không làm mất đi tính ổn định của robot [22].

Mặc dù cơ cấu bánh xích từ đã được sử dụng rộng rãi trong robot leo tường trên các bề mặt sắt từ, nhưng khả năng linh động, tốc độ di chuyển vẫn là thách thức lớn. Để giải quyết vấn đề này, một số nhóm nghiên cứu đã đề xuất sử dụng cơ chế bám dính và di chuyển bằng các bánh xe từ tính có trọng lượng nhẹ [23]. Các bánh xe từ tính này mở rộng phạm vi ứng dụng tiềm năng của robot, cho phép bánh xe hoạt động trên các bề mặt cong [24] và không bằng phẳng [25].

Trong thực tiễn công nghiệp, các bồn nổi chứa xăng dầu dạng trụ thường được làm bằng thép carbon (vật liệu có từ tính). Vì vậy, các nghiên cứu về robot leo bồn thường sử dụng cơ chế bám dính bằng nam châm, đặc biệt là nam châm vĩnh cửu. Mặc dù đã có một số robot leo bồn thương mại hóa được trang bị bánh xích từ tính, tuy nhiên các nghiên cứu thử nghiệm cho thấy hệ thống động lực bánh xích làm gia tăng đáng kể trọng lượng robot và làm giảm khả năng cơ động. Từ các hạn chế này, nhóm tác giả đề xuất sử dụng hệ thống động lực bánh xe từ tính, dùng nam châm vĩnh cửu cho thiết kế robot leo bồn, do tính cơ động cao và trọng lượng nhỏ. Do robot phải hoạt động trong thời gian dài, nên phương pháp bám dính dựa trên nam châm vĩnh cửu là lựa chọn phù hợp nhằm tiết kiệm năng lượng.

2. Thiết kế robot phát hiện ăn mòn bồn nổi hình trụ đứng chứa xăng dầu

2.1. Thiết kế robot mang bộ từ hóa

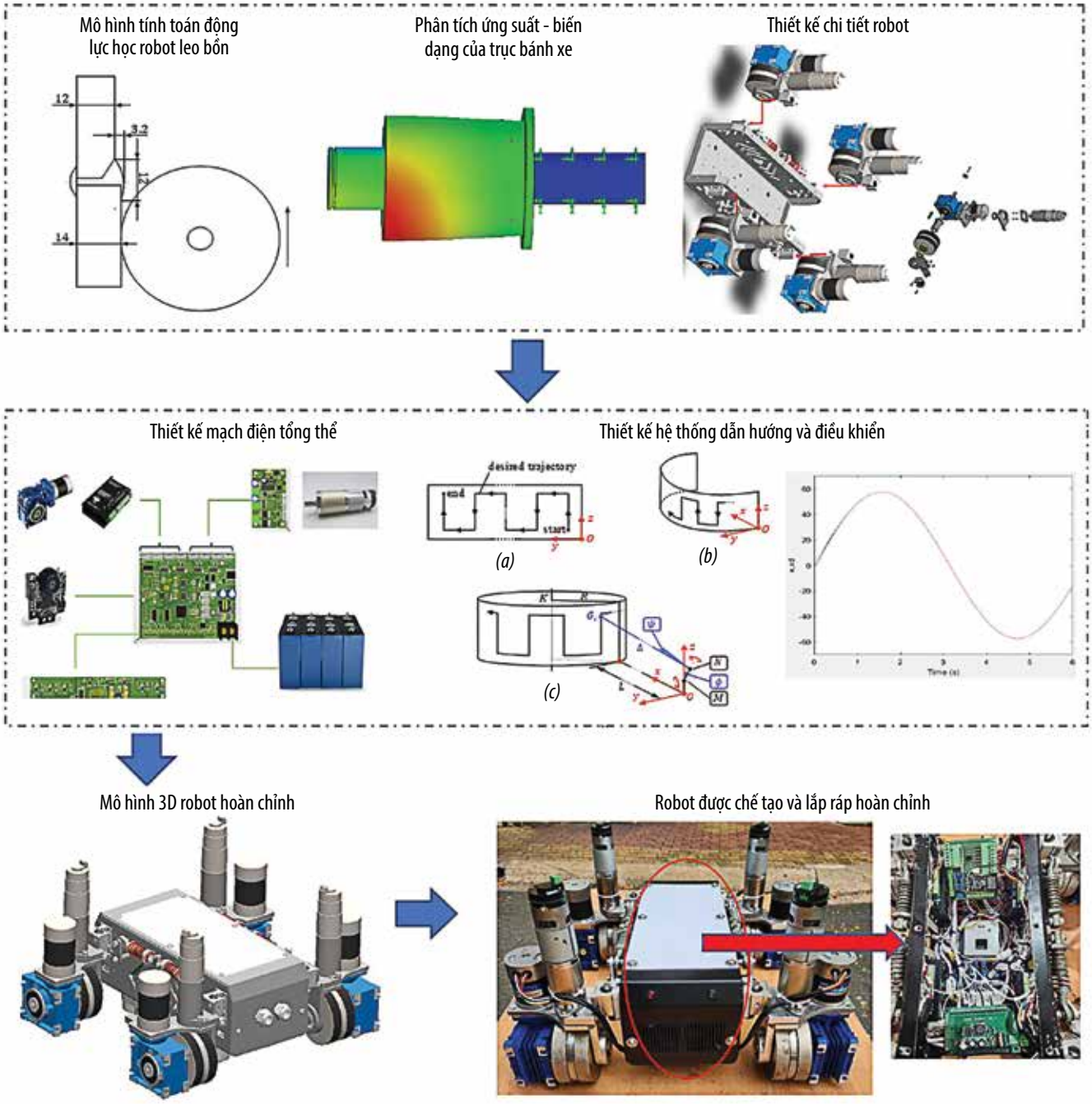
Ở giai đoạn thiết kế nguyên lý, có 2 vấn đề lớn cần giải quyết. Thứ nhất là tối ưu hóa thiết kế hình dáng, kích thước cơ bản của robot và hệ thống nam châm cung cấp

Hình 1. *Thực nghiệm robot di chuyển trên bề mặt cong.*

Hình 2. *Thực nghiệm robot di chuyển trên thành bồn đứng, vượt qua đường hàn.*

lực bám dính, sao cho bảo đảm đủ lực bám dính, đồng thời giảm tối đa trọng lượng robot. Thứ hai là lựa chọn phương thức di chuyển hợp lý, đảm bảo robot có khả năng di chuyển linh hoạt trên bề mặt thẳng đứng, bán kính cong khác nhau và có thể vượt qua các đường hàn có chiều cao đến 5 mm (theo khảo sát thực tế). Hình 1 và 2 trình bày quá trình thực nghiệm tại xưởng nhằm hiệu chỉnh các phương án thiết kế sơ bộ robot.

Sau khi thử nghiệm, phương án di chuyển bằng 4 bánh xe từ, truyền động độc lập để thiết kế hệ thống di chuyển cho robot leo bồn đã được lựa chọn. Để hoàn thiện mô hình nguyên mẫu robot, nhóm tác giả đã tính toán, thiết kế chi tiết toàn bộ robot với sự hỗ trợ của các phần mềm chuyên dụng. Cụ thể, mô hình toán động lực học được xây dựng nhằm tối ưu hóa thiết kế tổng thể robot dựa trên phần mềm Matlab, gồm kích thước cơ bản, lực từ và moment trục bánh xe, bố trí chung và trọng tâm robot...; các cơ cấu truyền động như trục bánh xe, cơ cấu giảm chấn được mô hình hóa và phân tích ứng suất - biến dạng dựa trên phương pháp phần tử hữu hạn (FEM) trên phần mềm Ansys; mạch điện tổng được thiết kế và



Hình 3. *Tối ưu hóa thiết kế robot leo bồn mang bộ từ hóa thông qua mô phỏng và thực nghiệm.*

mô phỏng trên phần mềm OrCAD; giải thuật điều khiển robot cũng được xây dựng và mô phỏng trên phần mềm Matlab nhằm đánh giá độ ổn định và tốc độ đáp ứng trước khi thử nghiệm. Ở giai đoạn thiết kế chi tiết và thiết kế thi công, toàn bộ mô hình 3D và bản vẽ thi công đều được xây dựng với sự hỗ trợ của phần mềm Solidworks. Toàn bộ quy trình tính toán, thiết kế và chế tạo robot leo bồn được mô tả trong Hình 3.

Giai đoạn này, nhóm nghiên cứu đã bước đầu hoàn thiện nguyên mẫu robot leo bồn theo phương thức bám dính bằng từ tính, với 4 bánh xe truyền động độc lập. Các thông số cơ bản của robot được trình bày trong Bảng 1.

2.2. Thiết kế bộ kiểm tra rò rỉ từ thông

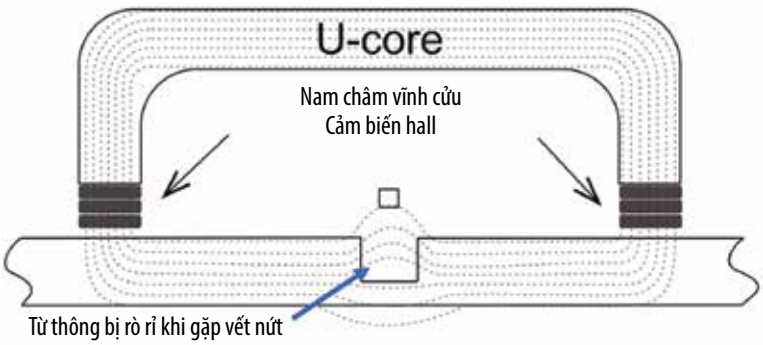
Bộ từ hóa đóng vai trò quan trọng trong phương pháp kiểm tra rò rỉ từ thông. Trong đó, kích thước và vật liệu cấu thành bộ từ hóa cần phải được thiết kế tối ưu sao cho giảm trọng lượng của toàn bộ robot trong khi tối đa hóa diện tích bao phủ (diện tích khảo sát bề mặt) và đủ khả năng cung cấp từ trường đạt trạng thái bão hòa trên bề mặt cần kiểm tra (nhằm đảm bảo tín hiệu rò rỉ từ thông được ghi nhận từ cảm biến luôn "rõ nét").

Hình 4 mô tả nguyên lý của phương pháp kiểm tra rò rỉ từ thông, trong đó, hệ thống gồm 1 bộ từ hóa dạng thanh (backing bar) và 1 bộ cảm biến từ trường, với 2 cực là nam châm vĩnh cửu. Bộ từ hóa tạo ra một vùng từ trường bão hòa, phân bố gần như đều và song song với bề mặt kết cấu thép. Khi gặp vết nứt, dù rất nhỏ, từ thông sẽ bị rò rỉ ra khỏi bề mặt kết cấu thép và được cảm biến ghi nhận. Dữ liệu này chính là biểu đồ mô tả cường độ rò rỉ từ thông dọc theo chiều dài của vết nứt. Qua nghiên cứu từ tài liệu [26, 27], kết hợp với mô phỏng số và thí nghiệm trên các mẫu thử (Hình 5), nhóm tác giả nhận thấy tín hiệu rò rỉ từ thông bị ảnh hưởng bởi các thông số công nghệ của bộ kiểm tra rò rỉ từ thông, đặc điểm kết cấu - vật liệu bề mặt kiểm tra và kích thước khuyết tật.

Trong thực tiễn, việc thử nghiệm số lượng lớn tổ hợp tham số thiết kế hệ gông từ là rất khó khăn. Do vậy, nhóm nghiên cứu đã tiếp

Bảng 1. *Thông số kỹ thuật cơ bản của robot được thiết kế*

Đặc điểm/Đặc tính kỹ thuật của robot	Thông số kỹ thuật/Thông tin
Trọng lượng	39 kg
Kích thước (dài x rộng x cao)	450 mm x 350 mm x 150 mm
Vận tốc di chuyển	10 - 50 mm/giây
Moment tối đa trên trục bánh xe	17 N.m
Bán kính bánh xe	50 mm



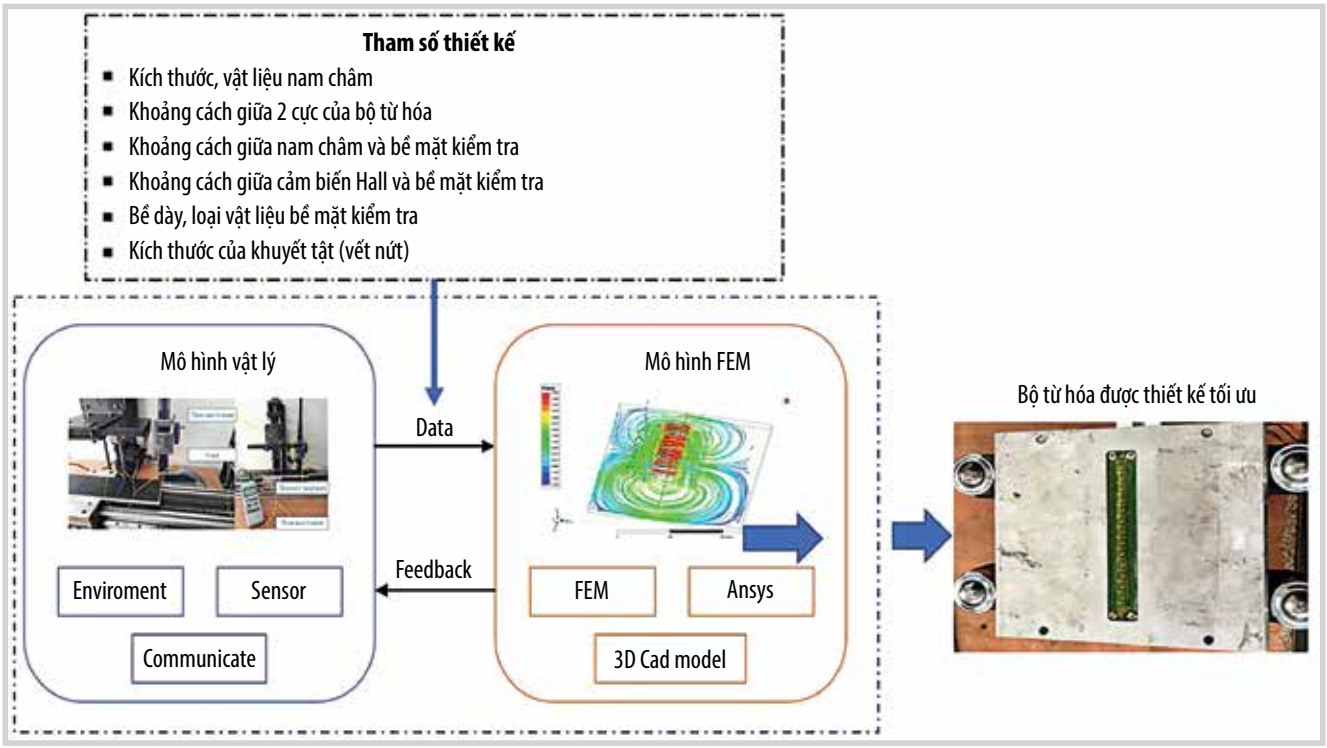
Hình 4. *Mô tả phương pháp kiểm tra rò rỉ từ thông MFL.*

cận với phương pháp mô phỏng số, dựa trên phương trình chủ đạo Maxwell và giải bằng phương pháp phần tử hữu hạn FEM trên module Ansys Maxwell, nhằm dự đoán ảnh hưởng của các thông số công nghệ đến cường độ tín hiệu rò rỉ từ thông. Các kết quả mô phỏng được xác nhận thông qua thực nghiệm trên mẫu thử với các khuyết tật được gia công chính xác. Cuối cùng, tập hợp thông số công nghệ tối ưu theo hướng giảm trọng lượng bộ từ hóa và tối đa hóa cường độ rò rỉ từ thông đã được ghi nhận. Toàn bộ phương pháp tiếp cận này được mô tả trong Hình 6, cấu tạo của bộ từ hóa hoàn chỉnh được trình bày trong Hình 7. Phương pháp luận và mô hình toán thiết kế tối ưu bộ từ hóa của nhóm tác giả đã được kiểm chứng trong một nghiên cứu tiền khả thi và được mô tả đầy đủ trong tài liệu [28].

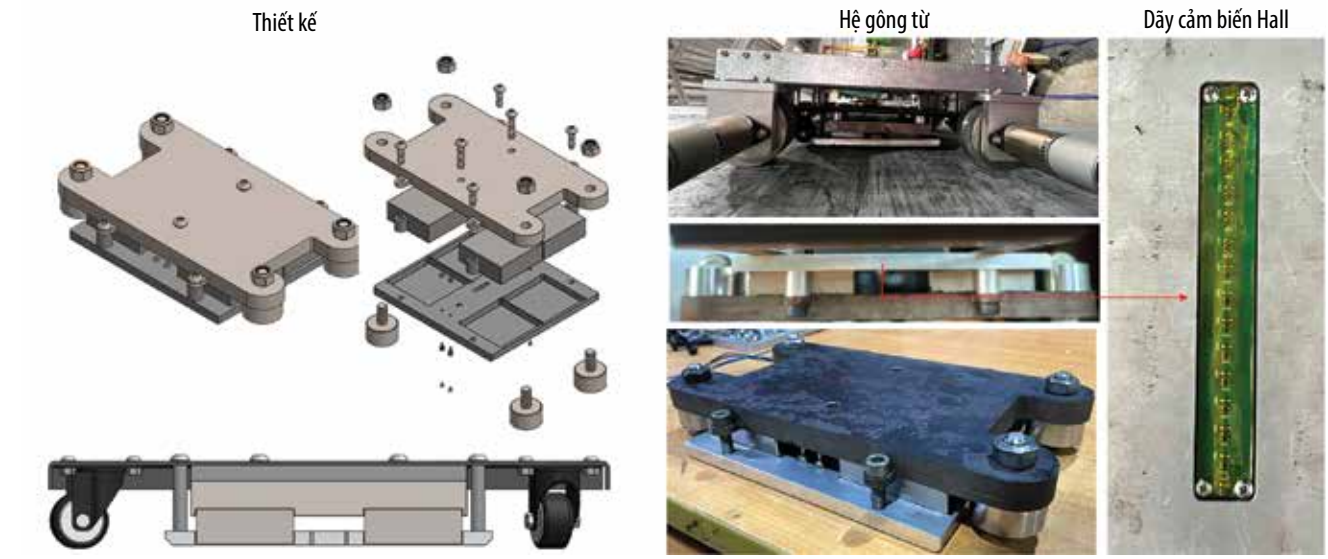
Trong quá trình thử nghiệm, nhóm tác giả phát hiện ra dữ liệu MFL bị ảnh hưởng bởi nhiễu hệ thống và các tác nhân môi trường trong quá trình vận hành. Bên cạnh đó, do hệ thống vận hành sử dụng nguyên lý MFL cho việc phát hiện khuyết tật dẫn tới sự phát sinh dòng điện xoáy khi robot di chuyển. Việc phát sinh dòng điện xoáy này có thể dẫn tới sự ảnh hưởng lên cường độ và sự thay đổi trong tín hiệu MFL, gây khó khăn trong việc đánh giá dữ liệu. Do vậy, việc xử lý dữ liệu thô từ cảm biến trả về là việc cơ bản cần thực hiện.

Quy trình xử lý tín hiệu MFL để xác định khuyết tật trên tấm kim loại gồm 4 bước chính (Hình 8). Đầu tiên, bước "Thu thập dữ liệu" là quá trình sử dụng cảm biến để thu thập tín hiệu MFL từ bề mặt tấm kim loại. Dữ liệu thu được thường gồm cả tín hiệu hữu ích liên quan đến khuyết tật và tín hiệu nhiễu gây ra bởi hệ thống, môi trường hay các yếu tố khác. Tiếp theo là bước "Lọc nhiễu", nơi các tín hiệu không mong muốn được loại bỏ dựa trên biến đổi wavelet theo kinh nghiệm (EWT). Quá trình này giúp giữ lại các thông tin MFL quan trọng, đồng thời làm mịn tín hiệu để giảm ảnh hưởng của nhiễu, giúp phát hiện

Hình 5. *Mô hình tham số hóa và thực nghiệm trong thiết kế hệ từ hóa.*



Hình 6. *Tối ưu hóa thiết kế bộ từ hóa dựa trên mô phỏng FEM và thực nghiệm mô hình thu nhỏ.*



Hình 7. *Cấu tạo hệ gông từ được thiết kế và chế tạo hoàn chỉnh.*

Hình 8. *Quy trình xử lý dữ liệu MFL.*

Bảng 2. *Bảng thông số kỹ thuật của bồn trụ đứng chứa xăng dầu khảo sát tại Công ty cổ phần Lọc hóa dầu Bình Sơn*

Thông số bồn trụ đứng chứa xăng dầu được khảo sát	
Lưu chất	MOGAS 92/95
Tiêu chuẩn bồn chứa xăng dầu	API 650 10th Sept 2003, Add.3 (SI Unit)
Đường kính trong của bồn	28.000 mm
Chiều cao của bồn	19.700 mm
Vật liệu làm bồn	A36M
Chiều cao đường hàn (ngang và đứng): Max. 5 mm	
Bề rộng đường hàn (ngang và đứng): 15 - 20 mm	
Sơn bề mặt ngoài 3 lớp: Có + Lớp trong cùng: Chống rỉ + Lớp giữa: Bám dính + Lớp ngoài cùng: Sơn che phủ	
Cách điện (insulation): Không có	

Bảng 3. *Thông số khuyết tật mẫu*

Tên khuyết tật	Chiều dài (mm)	Độ sâu (mm)	Bề rộng (mm)	Phương	Vị trí
KT1	25	50	2	45°	S1
KT2	37,5	5	2	Thẳng đứng	S2
KT3	50	4	2	Ngang	S3
KT4	65	3	2	Xiên góc 60°	S1
KT5	40	3,5	2	Xiên góc 120°	S2
KT6	40	2	2	Xiên góc bất kỳ	S3
KT7	25	3	2	Thẳng đứng	S4
KT8	65	2	2	Ngang	S4

rõ hơn các dấu hiệu của khuyết tật. Tuy nhiên, tín hiệu sau khi lọc vẫn có thể chịu ảnh hưởng của dòng điện xoáy (eddy current), gây ra sự không đồng nhất về cường độ. Do đó, trong bước "Xử lý, phân tích dữ liệu", một thuật toán được sử dụng để đồng nhất và chuẩn hóa dữ liệu thu được, giúp cải thiện độ chính xác trong việc xác định vị trí khuyết tật. Cuối cùng, toàn bộ dữ liệu sau khi xử lý sẽ được "Vẽ bản đồ 2D" trên hệ trục tọa độ của tấm kim loại, qua đó định vị chính xác khu vực có khả năng tồn tại khuyết tật.

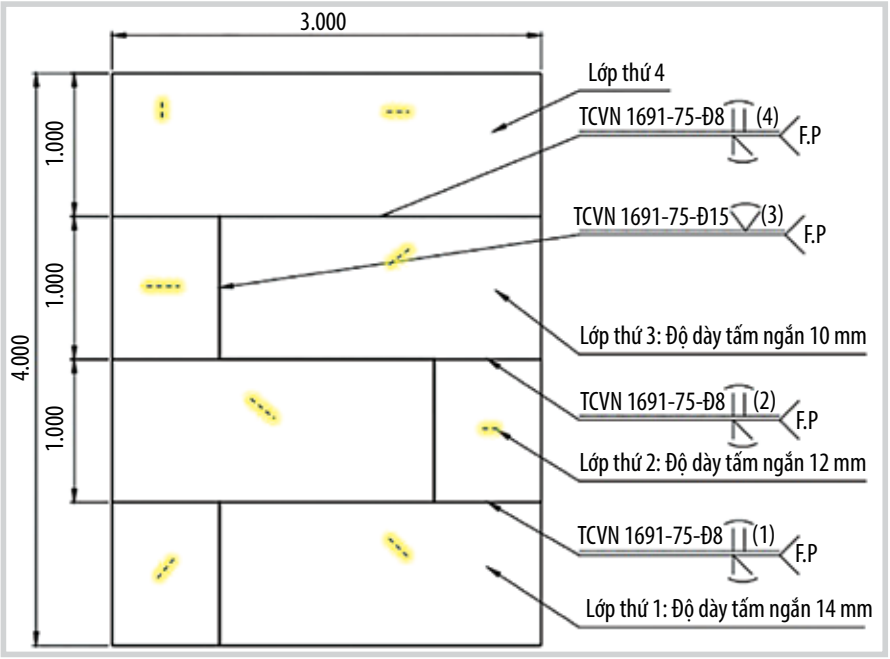
3. Kết quả nghiên cứu chế tạo robot phát hiện ăn mòn bồn nổi hình trụ đứng chứa xăng dầu

Để đánh giá khả năng đáp ứng động lực học (độ bám dính, khả năng di chuyển, tính cơ động) và khả năng phát hiện ăn mòn của robot được thiết kế chế tạo, nhóm tác giả trình bày kết quả thực nghiệm tại xưởng, khảo sát đặc trưng kết cấu của các bồn trụ đứng chứa xăng dầu, nhằm cung cấp dữ liệu đầu vào phục vụ thiết kế thí nghiệm; Mục 3.2 sẽ trình bày kết quả ghi nhận được sau thí nghiệm và đưa ra một số nhận xét.

3.1. Thiết kế thí nghiệm

Sau khi thực hiện khảo sát các loại bồn trụ đứng chứa xăng dầu tại Công ty cổ phần Lọc hóa dầu Bình Sơn (BSR) tại khu P1 và P3, nhóm tác giả đã chọn bồn trụ đứng (thông số kỹ thuật như Bảng 2) làm chuẩn để gia công tấm thép thử nghiệm, vì dữ liệu thiết kế, bảng thông số kỹ thuật và phạm vi chiều dày của tấm thép thân bồn hoàn toàn phù hợp với đặc tính kỹ thuật của robot được thiết kế.

Từ dữ liệu khảo sát này, 1 tấm thép đứng được chế tạo nhằm mô phỏng lại các đặc trưng vật lý của thành bồn trong thực tế. Cụ thể, tấm thép được cấu thành từ vật liệu thép A36M, kích thước (A x L) = 4.000 mm x 3.000 mm, được hàn ghép từ các tấm thép nhỏ có độ dày khác nhau, phạm vi độ dày 8 - 14 mm, cùng với các khuyết tật nhân tạo với kích thước và hình dáng khác nhau, nhằm đánh giá toàn diện khả năng phát hiện ăn mòn của robot trong thực tế. Hình 9 và Bảng 3 trình bày chi tiết các đặc trưng vật lý của tấm thép thử nghiệm đã được mô tả ở trên. Hình 10 trình bày hình ảnh thực tế của tấm thép được gia công theo thiết kế.



Hình 9. *Bản thiết kế tấm thép thử nghiệm.*

Hình 10. *Tấm thép thử nghiệm.*

3.2. Kết quả

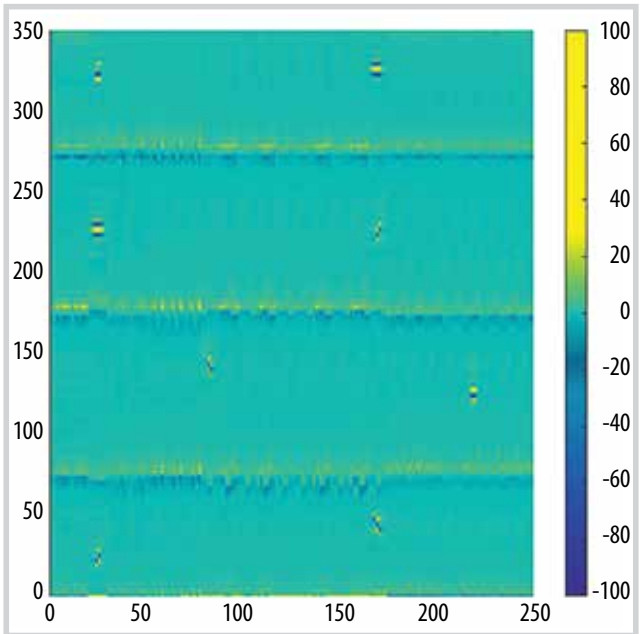
Kết quả thí nghiệm thực hiện tại xưởng trên tấm thép đặt theo phương đứng (Hình 11) như sau: Lực hút của bánh xe từ: 900 N; lực hút của gông từ: 2.000 N; chiều sâu khuyết tật nhỏ nhất có thể phát hiện: 2 mm.

Kết quả này chứng tỏ hệ thống MFL có độ nhạy cao, vượt trội so với phương pháp kiểm tra bằng dòng điện xoáy. Hình 12 trình bày biểu đồ cường độ tín hiệu MFL thu được sau thí nghiệm với phạm vi quét 2,5 m × 3,5 m. Cụ thể, các khuyết tật với kích thước và hướng khác nhau đều được phát hiện và biểu thị rõ ràng trong biểu đồ. Lưu ý rằng 3 đường tín hiệu MFL nằm ngang phản ánh các đường hàn trên tấm thép.

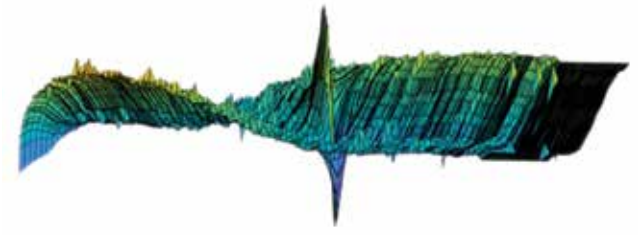
Trong quá trình thử nghiệm, lộ trình quét của robot đã được quy hoạch tối ưu nhằm giảm thời gian quét, đồng thời đảm bảo các khuyết tật mẫu nằm trong phạm vi quét của mạch cảm biến Hall. Do đó, các đường hàn dọc không nằm trong phạm vi quét của các cảm biến Hall.

Một mẫu tín hiệu ghi nhận được từ bộ cảm biến khi đi qua một khuyết tật được trình bày trong Hình 13a. Đây là tín hiệu MFL khi robot quét qua khuyết tật có chiều sâu nhỏ nhất (2 mm) ở khu vực tấm thép có bề dày 8 mm, tương ứng với khuyết tật KT8 trong Bảng 3. Mặc dù dữ liệu xuất hiện nhiễu trong quá trình thu thập, nhưng không ảnh hưởng đến việc phát hiện khuyết tật. Các tín hiệu nhiễu xuất hiện ở lân cận tín hiệu chính do rò rỉ từ thông khi gặp khuyết tật, nguyên nhân là do mật độ từ thông phân bố giữa 2 cực nam châm của hệ từ hóa quá

Hình 11. *Robot đang thử nghiệm trên tấm thép đứng tại xưởng.*



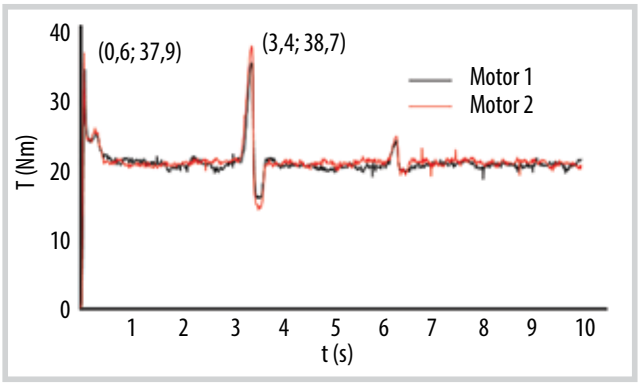
Hình 12. *Kết quả kiểm tra khuyết tật.*



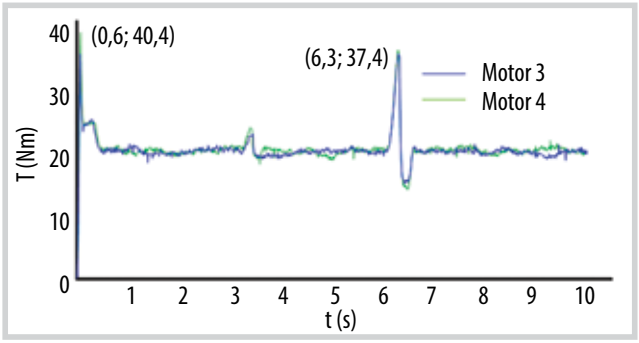
(a) Tín hiệu MFL ban đầu

(b)Tín hiệu MFL sau khi khử nhiễu

Hình 13. *Tín hiệu rò rỉ từ thông (MFL) khi robot đi qua khuyết tật có chiều sâu 2 mm, tại tấm thép có bề dày 8 mm.*



(a) Hai bánh trước



(b) Hai bánh sau

Hình 14. *Đồ thị moment động cơ khi đi qua đường hàn.*

lớn (khoảng 300 Gauss). Trên thực tế, các tín hiệu nhiễu này dễ dàng được loại bỏ bằng các bộ lọc (xem kết quả lọc nhiễu ở Hình 13b).

Về khả năng bám dính và tính cơ động, kết quả thí nghiệm cho thấy robot có thể di chuyển theo phương thẳng đứng trên tấm thép với tốc độ tối đa 0,05 m/giây. Nhờ lực từ rất lớn, cùng với cơ cấu giảm chấn và hệ động lực độc lập của 4 bánh xe từ, robot có khả năng vượt qua các đường hàn có chiều cao tối đa 10 mm. Trong thực tế, các mối hàn thường có chiều cao rất nhỏ so với 10 mm. Hình 14 trình bày đồ thị moment xoắn tương ứng với 4 động cơ khi robot vượt qua đường hàn trên tấm thép thử nghiệm. Do đó, với nguyên mẫu hiện tại, robot hoàn toàn đủ khả năng bám dính, di chuyển và vượt qua đường hàn an toàn trên các bồn trụ đứng chứa xăng dầu.

4. Kết luận

Kết quả nghiên cứu, thiết kế và chế tạo robot phát hiện ăn mòn trên bồn hình trụ đứng chứa xăng dầu. Robot và hệ thống kiểm tra rò rỉ từ thông (MFL) được thiết kế tối ưu bằng các phần mềm chuyên dụng như Matlab, Ansys, SolidWorks và OrCAD nhằm giảm trọng lượng và tăng hiệu quả kiểm tra ăn mòn.

Để đánh giá khả năng di chuyển, tính cơ động và khả năng phát hiện khuyết tật, nhóm tác giả đã chế tạo một mô hình thành bồn trụ đứng với tấm thép kích thước 4 m (cao) × 3 m (rộng), bán kính cong 14 m. Tấm thép này được hàn từ các tấm thép nhỏ có bề dày từ 8 - 14 mm, cùng với các khuyết tật nhân tạo trên mặt sau, nhằm mô phỏng đặc trưng vật lý của thành bồn chứa xăng dầu và hóa chất trong thực tế. Kết quả thử nghiệm trên tấm thép đứng tại xưởng cho thấy robot đã đạt được các chỉ tiêu chất lượng chính như vận tốc di chuyển tối đa 0,05 m/giây; khả năng bám dính và vượt qua đường hàn cao tối đa 10 mm; chất lượng tín hiệu MFL ghi nhận có khả năng phát hiện khuyết tật có độ sâu 2 mm trên tấm thép dày đến 14 mm.

Trong các nghiên cứu tiếp theo, nhóm tác giả sẽ tiếp tục hoàn thiện nguyên mẫu robot, tối ưu hóa thiết kế hệ thống định vị, dẫn hướng và điều khiển; phát triển bộ dữ liệu và các thuật toán trí tuệ nhân tạo để tái dựng chính xác hình dạng và kích thước khuyết tật từ tín hiệu MFL, nhằm cung cấp dữ liệu hữu ích cho thống kê, đánh giá tuổi thọ bồn chứa và ra quyết định bảo trì sớm.

Tài liệu tham khảo

[1] ATS, "Tank inspection protocol API 653". [Online]. Available: https://www.atsenvironmental.com/commercial/aboveground-storage-tanks/api-inspection/api-653/#:~:text=What is API 653%3F,650 or API 12C standards.

[2] Neptune Robotics, "A world class automated biofouling management and robotic hull cleaning system". [Online]. Available: https://neptune-robotics.com/.

[3] Maverick Inspection, [Online]. Available: https://www.maverickinspection.com/.

[4] Robavatar, "Crawler robot chassis". [Online]. Available: https://www.robavatar.com/products/basic/crawler/.

[5] Maria Inês Silva, Evgenii Malitckii, Telmo G. Santos, and Pedro Vilaça, "Review of conventional and advanced non-destructive testing techniques for detection and characterization of small-scale defects", *Progress in Materials Science*, Volume 138, 2023. DOI: 10.1016/j.pmatsci.2023.101155.

[6] Erlong Li, Yihua Kang, Jian Tang, and Jianbo Wu, "A new micro magnetic bridge probe in magnetic flux leakage for detecting micro-cracks", *Journal of Nondestructive Evaluation*, Volume 37, 2018. DOI: 10.1007/s10921-018-0499-8.

[7] Huaguang Zhang, Lei Wang, Jianfeng Wang, Fengyuan Zuo, Jifeng Wang, and Jinhai Liu, "A pipeline defect inversion method with erratic MFL signals based on cascading abstract features", *IEEE Transactions on Instrumentation and Measurement*, Volume 71, pp. 1 - 11, 2022. DOI: 10.1109/TIM.2022.3152243.

[8] Bin Liu, Zihan Wu, Peng Wang, Luyao He, Lijian Yang, Zheng Lian, and Tong Liu, "Quantization of magnetic flux leakage internal detection signals for composite defects of gas and oil pipelines", *Energy Reports*, Volume 9, pp. 5899 - 5914, 2023. DOI: 10.1016/j.egyr.2023.05.025.

[9] Lin Jiang, Huaguang Zhang, Jinhai Liu, Xiangkai Shen, and Lei Wang, "Pipeline irregular defect inversion for magnetic flux leakage detection system based on heterogeneous multiclass feature fusion", *IEEE Transactions on Instrumentation and Measurement*, Volume 72, pp. 1 - 9, 2023. DOI: 10.1109/TIM.2023.3265110.

[10] Bingyi Mao, Yi Lu, Peiliang Wu, Bingzhi Mao, and Pengfei Li, "Signal processing and defect analysis of pipeline inspection applying magnetic flux leakage methods", *Intelligent Service Robotics*, Volume 7, Issue 4, pp. 203 - 209, 2014. DOI: 10.1007/s11370-014-0158-6.

[11] Jiatong Ling, Ke Feng, Teng Wang, Min Liao, Chunsheng Yang, and Zheng Liu, "Data modeling techniques for pipeline integrity assessment: A state-of-the-art survey", *IEEE Transactions on Instrumentation and Measurement*, Volume 72, pp. 1 - 17, 2023. DOI: 10.1109/TIM.2023.3279910.

[12] Jie Yuan, Mengtian Qiao, Chun Hu, Yufei Cheng, Zhen Wang, and Dezhi Zheng, "A classification and quantitative assessment method for internal and external surface defects in pipelines based on ASTC-Net", *Advanced Engineering Informatics*, Volume 61, 2024. DOI: 10.1016/j.aei.2024.102492.

[13] Rongbiao Wang, Yongzhi Chen, Haozhi Yu, Zhiyuan Xu, Jian Tang, Bo Feng, YihuKang, and Kai Song, "Defect classification and quantification method based on AC magnetic flux leakage time domain signal characteristics", *NDT E International*, Volume 149, 2025. DOI: 10.1016/j.ndteint.2024.103250.

[14] Haotian Wei, Shaohua Dong, Lushuai Xu, Fan Chen, Hang Zhang, and Xingtao Li, "Internal inspection method for crack defects in ferromagnetic pipelines under remanent magnetization", *Measurement*, Volume 242, 2025. DOI: 10.1016/j.measurement.2024.115907.

[15] Lin Jiang, Huaguang Zhang, Jinhai Liu, Qi Xiao, Xiangkai Shen, and Hang Xu, "A physics-guided MFL deformed defect recovery method", *IEEE Transactions on Automation Science Engineering*, Volume 21, Issue 2, pp. 2113 - 2124, 2024. DOI: 10.1109/TASE.2023.3260281.

[16] Pengpeng Shi, Pengcheng Zhang, Shuai Hao, Wenshuai Wang, and Xiaofan Gou, "Classification and evaluation for nearside/backside defect via magnetic flux leakage: A dual probe design with SVM and PSO intelligence algorithms", *NDT E International*, Volume 144, Issue 1, 2024. DOI: 10.1016/j.ndteint.2024.103100.

[17] Lei Wang, Huaguang Zhang, Jinhai Liu, Xiangkai Shen, and Fengyuan Zuo, "Irregular defect size estimation for the magnetic flux leakage detection system via expertise-informed collaborative network", *IEEE Transactions on Industrial Electronics*, Volume 71, Issue 10, pp. 13189 - 13200, 2024. DOI: 10.1109/TIE.2024.3357845.

[18] Zeliang Xu and Peisun Ma, "A wall-climbing robot for labelling scale of oil tank's volume", *Robotica*, Volume 20, Issue 2, 2002. DOI: 10.1017/S0263574701003964.

[19] Weimin Shen, J. Gu, and Yanjun Shen, "Permanent magnetic system design for the wall-climbing robot", *IEEE International Conference on Mechatronics and Automation, Niagara Falls, ON, Canada, 29 July - 1 August 2005*. DOI: 10.1533/abbi.2006.0024.

[20] Patrick Schoeneich, Frederic Rochat, Olivier Truong-Dat Nguyen, Roland Moser, and Francesco Mondada, "TRIPILLAR: A miniature magnetic caterpillar climbing robot with plane transition ability", *Robotica*, Volume 29, Issue 7, pp. 1075 - 1081, 2011. DOI: 10.1017/S0263574711000257.

[21] Yanheng Liu, HyunGyu Kim, and TaeWon

Seo, "AnyClimb: A new wall-climbing robotic platform for various curvatures", *IEEE/ASME Transactions on Mechatronics*, Volume 21, Issue 4, 2016. DOI: 10.1109/TMECH.2016.2529664.

[22] Junyu Hu, Xu Han, Yourui Tao, and Shizhe Feng, "A magnetic crawler wall-climbing robot with capacity of high payload on the convex surface", *Robotics and Autonomous Systems*, Volume 148, 2022. DOI: 10.1016/j.robot.2021.103907.

[23] Wolfgang Fischer, Fabien Tâche, and Roland Siegwart, "Magnetic wall climbing robot for thin surfaces with specific obstacles", *Springer Tracts in Advanced Robotics*, 2008. DOI: 10.1007/978-3-540-75404-6_53.

[24] Mansi S. Chabukswar, Ravikant K. Nanwatkar, and Aparna M. Bagde, "Design and experimental analysis of magnetic climbing robot", *Intnational Journal of Research in Engineering Science and Management*, Volume 2, Issue 12, 2019.

[25] Raihan Enjikalayil Abdulkader, Prabakaran Veerajagadheswar, Nay Htet Lin, Selva Kumaran, Suresh Raj Vishaal, and Rajesh Elara Mohan, "Sparrow: A magnetic climbing robot for autonomous thickness measurement in ship hull maintenance", *Journal of Marine Science and Engineering*, Volume 8, Issue 6, 2020. DOI: 10.3390/JMSE8060469.

[26] Yue Long, Songling Huang, Lisha Peng, Shen Wang, and Wei Zhao, "A new dual magnetic sensor probe for lift-off compensation in magnetic flux leakage detection", *IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Dubrovnik, Croatia, 25 - 28 May 2020*. DOI: 10.1109/I2MTC43012.2020.9129204.

[27] Peng-Peng Shi, Shuai Hao, and Tian-Shou Liang, "The defect depth evaluation based on the dual-sensor strategy: Resisting the lift-off disturbance in magnetic flux leakage testing", *Journal of Magnetism and Magnetic Materials*, Volume 582, 2023. DOI: 10.1016/j.jmmm.2023.171039.

[28] Quoc Minh Lam, Van Tu Duong, Hoang Long Phan, Tan Tien Nguyen, and Huy Hung Nguyen, "Parametric design methodology for yoke-magnetization in magnetic flux leakage detection systems", *Results Engineering*, Volume 23, 2024. DOI: 10.1016/j.rineng.2024.102669.

[29] Nguyễn Thị Lan, Huỳnh Khắc Tâm và Thái Lâm Cường Quốc, "Nghiên cứu ứng dụng robot và phương pháp kiểm tra không phá hủy trong đánh giá ăn mòn bồn chứa nhiên liệu", *Dầu khí - Khoa học, Công nghệ và Đổi mới sáng tạo*, Số 5, trang 69 - 78, 2024. DOI: 10.47800/PVSI.2024.05-08.

DEVELOPMENT OF A WALL-CLIMBING ROBOT FOR CORROSION DETECTION OF VERTICAL OIL TANKS

Nguyen Thi Lan[1, 2], Tran Duong Tan Quyen[1], Pham Quoc Huy[3], Nguyen Tan Tien[1, 4], Le Van Sy[2, 5]
[1]Ho Chi Minh City University of Technology - Vietnam National University Ho Chi Minh City
[2]Petrovietnam University (PVU)
[3]Petrovietnam College (PV College)
[4]National key Laboratory of Digital Control and System Engineering (DCSELab)
[5]Petrovietnam Maintenance and Repair Corporation (PVMR)
Email: ntlan.sdh231@hcmut.edu.vn

Summary

This paper presents the research, design, and fabrication of a prototype robot for corrosion detection on vertical cylindrical oil tanks using the magnetic flux leakage (MFL) inspection technique. The robot's overall design, propulsion system, electrical components, control architecture, and MFL inspection unit were meticulously optimized using specialized software, in compliance with IEC explosion safety standards. The completed prototype has a total mass of 39 kg and operates within a speed range of 10 - 50 mm/s.

The prototype then underwent testing on a vertical steel plate fabricated to simulate the physical properties of actual gasoline and oil storage tanks. Preliminary test results demonstrate the robot's adherence to key performance criteria, including strong adhesion, efficient mobility on vertical surfaces, and high sensitivity to micro-defects. Specifically, the robot can traverse weld seams up to 10 mm in height, maintain stable movement at a maximum speed of 50 mm/s, and detect defects as small as 2 mm in depth on steel plates up to 14 mm thick.

Key words: Corrosion inspection, non-destructive testing (NDT), magnetic flux leakage, wall-climbing robot.

DầuKhí
KHOA HỌC, CÔNG NGHỆ VÀ ĐỔI MỚI SÁNG TẠO PETROVIETNAM

KẾT HỢP SỬ DỤNG ẢNH VIỄN THÁM QUANG HỌC VÀ RADAR TRONG PHÂN LOẠI VÀ GIÁM SÁT VẾT DẦU TRÊN BIỂN

Trịnh Lê Hùng, Lê Văn Phú
Học viện Kỹ thuật Quân sự
Email: trinhlehung@lqdtu.edu.vn
https://doi.org/10.47800/PVSI.2025.03-09

Tóm tắt

Dữ liệu viễn thám được sử dụng phổ biến trên thế giới trong nghiên cứu ô nhiễm tràn dầu trên biển. Bài viết giới thiệu kết quả kết hợp sử dụng ảnh viễn thám quang học Sentinel 2 MSI và ảnh radar Sentinel 1 giúp phát hiện và phân loại vết dầu trên biển. Dữ liệu Sentinel 2 MSI được sử dụng để tính chỉ số OSI (oil spill index) trên cơ sở các kênh trong dải sóng nhìn thấy, trong khi dữ liệu Sentinel 1 dùng để tính giá trị tán xạ ngược, từ đó phân loại vết dầu bằng phương pháp phân ngưỡng. Việc kết hợp dữ liệu viễn thám đa chủng loại cho phép tăng dày bộ dữ liệu đầu vào, nâng cao hiệu quả công tác giám sát ô nhiễm tràn dầu trên biển.

Từ khóa: Ảnh viễn thám quang học, ảnh viễn thám radar, sự cố tràn dầu, phân loại.

1. Giới thiệu

Quá trình khai thác, vận chuyển dầu mỏ cũng như hoạt động giao thông hàng hải tiềm ẩn những nguy cơ gây ô nhiễm môi trường biển, đặc biệt là ô nhiễm tràn dầu. Do đặc điểm khu vực biển rộng lớn, việc tiếp cận bằng các phương pháp quan trắc truyền thống gặp rất nhiều khó khăn. Các sự cố ô nhiễm tràn dầu thường chỉ được phát hiện khi vết dầu đã lan vào khu vực gần bờ, làm giảm hiệu quả công tác ứng phó với sự cố [1].

Công nghệ viễn thám có ưu điểm là diện tích phủ trùm rộng, có thể thu nhận dữ liệu ảnh đa thời gian ở các khu vực khó tiếp cận, được sử dụng phục vụ phát hiện sớm và giám sát ô nhiễm tràn dầu trên biển. Các nghiên cứu chủ yếu sử dụng dữ liệu ảnh viễn thám radar (ảnh SAR) trong quan trắc vết dầu do ảnh SAR ít chịu ảnh hưởng của điều kiện thời tiết [2, 3]. Một số nghiên cứu sử dụng dữ liệu viễn thám quang học như MODIS [4, 5], Sentinel 2 [6] và Landsat [7] để phân loại vết dầu trên biển. Tuy nhiên, dữ liệu viễn thám quang học Sentinel 2 và Landsat có một số hạn chế nhất định, đặc biệt là sự phụ thuộc vào điều kiện ánh sáng và dễ bị ảnh hưởng bởi mây che phủ, làm giảm hiệu quả phát hiện vết dầu trong điều kiện thời tiết bất lợi. Để khắc phục hạn chế này, Sentinel 1 với cảm biến SAR cung cấp dữ liệu không bị ảnh hưởng bởi mây và ánh sáng, cho phép giám sát liên tục và hiệu quả hơn các sự cố tràn dầu trên biển. Nhiều nghiên cứu cho thấy các phương pháp nhận dạng và phân loại vết dầu trên ảnh viễn thám được sử dụng phổ biến gồm: phương pháp phân ngưỡng tự động [8, 9], phương pháp sử dụng mô hình trí tuệ nhân tạo như học máy, học sâu [10 - 13], phương pháp sử dụng các đặc trưng (texture) của dữ liệu ảnh [14]. Các nghiên cứu này đã chứng minh tính hiệu quả của phương pháp viễn thám so với phương pháp nghiên cứu truyền thống, giúp phát hiện nhanh và giám sát sự lan truyền của vết dầu tràn trên biển.

Bài viết giới thiệu kết quả kết hợp sử dụng dữ liệu viễn thám quang học (Sentinel 2) và radar (Sentinel 1) trong phân loại và đánh giá sự thay đổi vết dầu theo thời gian. Phương pháp phân ngưỡng tự động được áp dụng đối với cả 2 loại dữ liệu viễn thám để phân loại vết dầu. Việc áp dụng dữ liệu viễn thám đa chủng loại giúp tăng dày nguồn dữ liệu và rút ngắn thời gian thu thập thông tin về ô nhiễm tràn dầu, từ đó nâng cao hiệu quả công tác giám sát ô nhiễm tràn dầu trên biển. Các thế hệ vệ tinh tiếp theo của chùm vệ tinh Sentinel 1 và Sentinel 2 được đưa vào hoạt động đã rút ngắn thời gian thu nhận dữ liệu tại một vị trí trên trái đất.

2. Dữ liệu và khu vực nghiên cứu

2.1. Khu vực nghiên cứu

Khu vực thử nghiệm trong nghiên cứu là vùng biển

Ngày nhận bài: 15/4/2025
Ngày đánh giá và sửa chữa: 15/4 - 6/5/2025
Ngày duyệt đăng: 6/5/2025

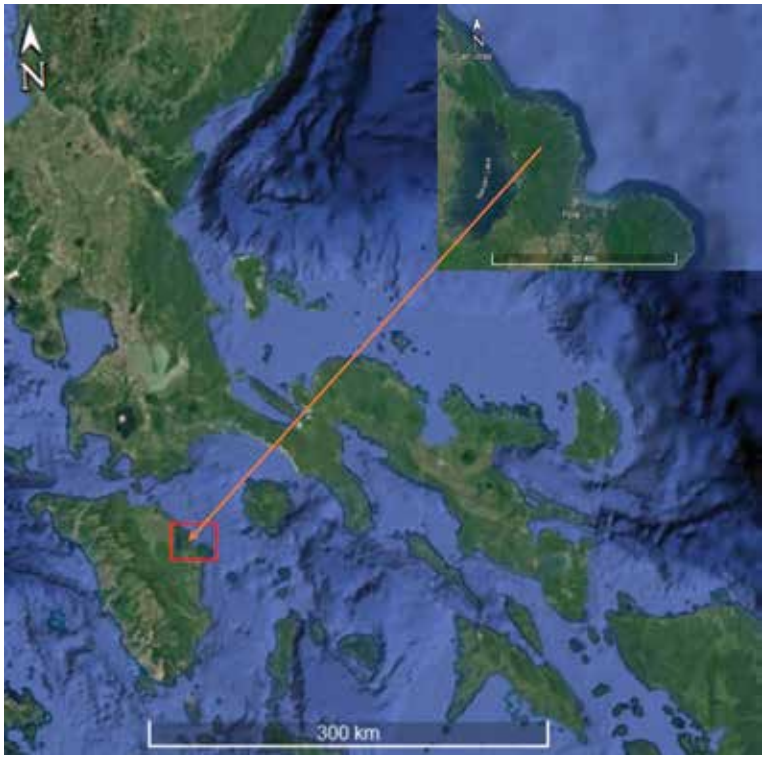
ngoài khơi Naujan, tỉnh Oriental Mindoro (Philippines). Đây là khu vực xảy ra sự cố tràn dầu do tàu chở nhiên liệu bị chìm vào ngày 28/2/2023. Vệt dầu loang sau đó đã tiến gần hơn đến bãi biển Polacay (thành phố Pola) và bãi biển Tagumpay. Bộ ảnh từ Cơ quan Vũ trụ Philippines (PhilSA) cho thấy sự di chuyển của vết dầu từ phía Đông của quần đảo Mindoro về vùng biển phía Đông Bắc của tỉnh [15].

2.2. Dữ liệu viễn thám

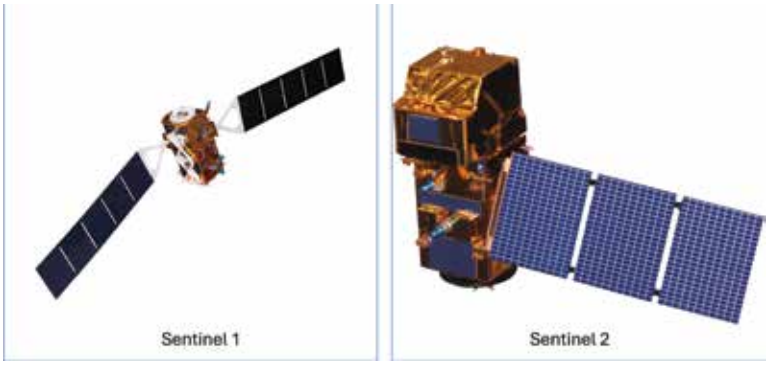
Sentinel là tên của các vệ tinh quan sát trái đất thuộc Chương trình Copernicus của Cơ quan Hàng không Vũ trụ châu Âu (ESA), bao gồm các vệ tinh từ Sentinel 1 đến Sentinel 6, trong đó có cả vệ tinh quang học và vệ tinh radar. Dữ liệu vệ tinh Sentinel có độ phân giải không gian đa dạng, thời gian cập nhật ngắn và được cung cấp hoàn toàn miễn phí. Dữ liệu vệ tinh Sentinel được sử dụng rất phổ biến trong nhiều lĩnh vực, đặc biệt là trong giám sát tài nguyên, môi trường.

Sentinel 1 là chùm vệ tinh radar gồm 3 vệ tinh (Sentinel 1A, Sentinel 1B và Sentinel 1C) có đặc điểm giống nhau. Sentinel 1A được phóng lên quỹ đạo ngày 3/4/2014, Sentinel 1B ngày 25/4/2016 và vệ tinh mới nhất - Sentinel 1C đã được phóng thành công lên quỹ đạo vào ngày 5/12/2024. Bộ cảm biến trên vệ tinh Sentinel 1 thu nhận ảnh radar khẩu độ mở tổng hợp, băng C (tần số 5.405 GHz) với các phân cực và độ phân giải không gian khác nhau. Thời gian chụp lặp lại của từng vệ tinh là 12 ngày. Như vậy, với 3 vệ tinh Sentinel 1 cho phép thu nhận ảnh tại một vị trí trên bề mặt trái đất trong 4 ngày.

Chùm vệ tinh quang học Sentinel 2 cũng gồm 3 vệ tinh có đặc điểm giống nhau (Sentinel 2A, Sentinel 2B và Sentinel 2C), trong đó Sentinel 2C được phóng thành công lên quỹ đạo vào ngày 5/9/2024. Vệ tinh Sentinel 2 sử dụng bộ cảm biến đa phổ MSI (MultiScanner Instrument), thu nhận ảnh ở 13 kênh phổ trong dải sóng nhìn thấy và hồng ngoại (Bảng 1). Độ phân giải không gian ảnh Sentinel 2 cao nhất lên đến 10 m ở các kênh trong dải sóng nhìn thấy (visible) và cận hồng ngoại (NIR-near infrared).



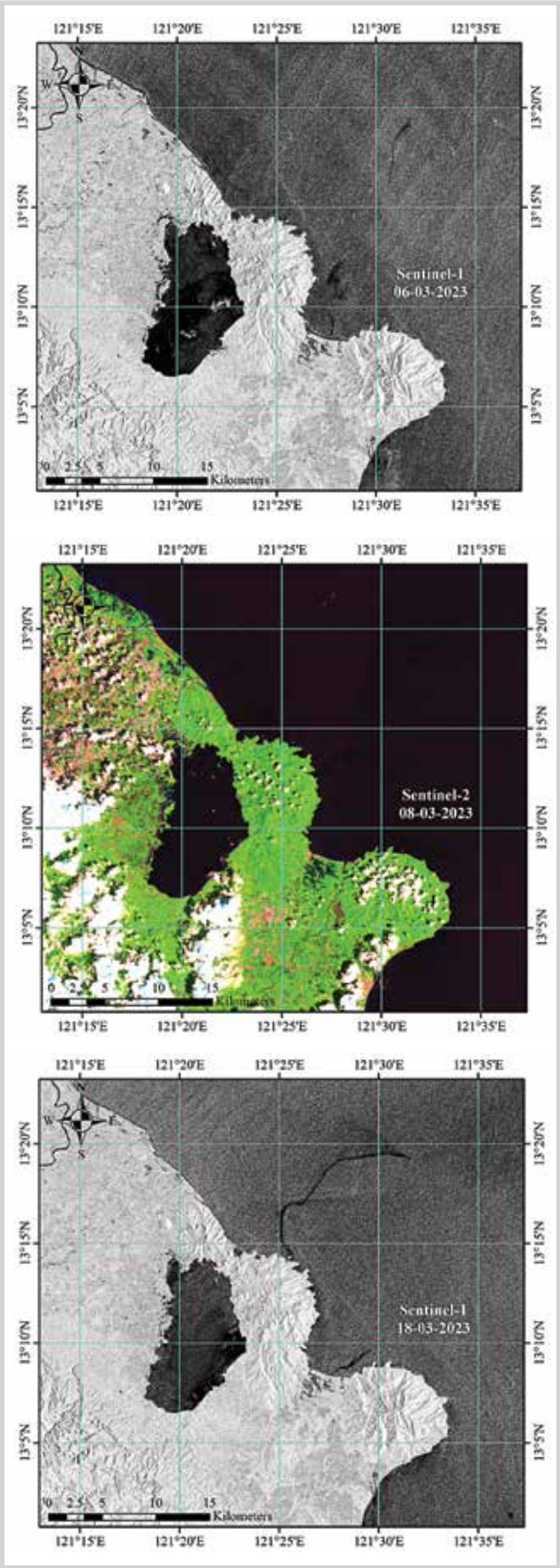
Hình 1. *Khu vực nghiên cứu tại vùng biển ngoài khơi Naujan, tỉnh Oriental Mindoro, Philippines.*



Hình 2. *Hình ảnh vệ tinh radar Sentinel 1 và vệ tinh quang học Sentinel 2.*

Bảng 1. *Đặc điểm ảnh vệ tinh Sentinel-2*

Kênh	Bước sóng (µm)	Độ phân giải không gian (m)
1	0,421 - 0,457	60
2	0,439 - 0,535	10
3	0,537 - 0,582	10
4	0,646 - 0,685	10
5	0,694 - 0,714	20
6	0,731 - 0,749	20
7	0,768 - 0,796	20
8	0,767 - 0,908	10
8A	0,848 - 0,881	20
9	0,931 - 0,958	60
10	1,338 - 1,414	60
11	1,539 - 1,681	20
12	2,072 - 2,312	20



Hình 3. *Dữ liệu viễn thám sử dụng trong nghiên cứu.*

Trong nghiên cứu này, ảnh radar Sentinel 1 chụp ngày 6/3/2023 và 18/3/2023 cùng ảnh quang học Sentinel 2 chụp ngày 8/3/2023 được sử dụng để phân loại vết dầu, từ đó đánh giá sự lan truyền của vết dầu trên vùng biển khu vực Naujan. Dữ liệu viễn thám sử dụng trong nghiên cứu được thể hiện trên Hình 3. Hình 3 cho thấy trên ảnh radar Sentinel 1, vết dầu tương phản rõ rệt với khu vực xung quanh, trong khi trên ảnh quang học Sentinel 2 rất khó phân biệt vết dầu và vùng biển.

3. Phương pháp nghiên cứu

Dữ liệu ảnh viễn thám Sentinel 1 và Sentinel 2 sau khi được thu thập tại cơ sở dữ liệu Copernicus (https://scihub.copernicus.eu/dhus/#/home) bằng nền tảng điện toán đám mây Google Earth Engine (GEE) được tiền xử lý và cắt theo ranh giới khu vực nghiên cứu.

3.1. Xử lý dữ liệu ảnh quang học Sentinel 2

Để phát hiện và phân loại vết dầu trên biển từ ảnh vệ tinh quang học Sentinel 2, trong nghiên cứu sử dụng chỉ số vết dầu OSI (oil spill index). Chỉ số OSI được đề xuất bởi Rajendran và cộng sự (2021) trên cơ sở sử dụng các kênh phổ ở dải sóng nhìn thấy (blue, green và red) theo công thức sau [16, 17]:

$$OSI = \frac{\rho_{RED} + \rho_{GREEN}}{\rho_{BLUE}} \quad (1)$$

Đối với ảnh Sentinel 2, chỉ số OSI có thể được xác định như sau:

$$OSI_{Sentinel2} = \frac{\rho_{band4} + \rho_{band3}}{\rho_{band2}} \quad (2)$$

Để loại bỏ ảnh hưởng của các đối tượng đất liền tới kết quả phân loại vết dầu bằng chỉ số OSI, trong nghiên cứu cũng sử dụng chỉ số nước MNDWI (modified normalized difference water index) xác định theo công thức sau [18]:

$$MNDWI = \frac{\rho_{GREEN} - \rho_{SWIR1}}{\rho_{GREEN} + \rho_{SWIR1}} \quad (3)$$

Trong đó, các kênh GREEN và SWIR1 tương ứng là kênh 2 và kênh 10 ảnh Sentinel 2.

Do đặc điểm vết dầu thường có hình dạng kéo dài, màu tối, việc phát hiện dựa trên điểm ảnh riêng lẻ thường thiếu chính xác và dễ gây nhiễu do tính biến động cao của mặt nước biển. Để khắc phục hạn chế này, trong nghiên cứu tiến hành phân đoạn ảnh (segmatation) thành các siêu điểm ảnh sử dụng thuật toán SNIC (simple non-iterative clustering) - một thuật toán phân cụm không lặp,

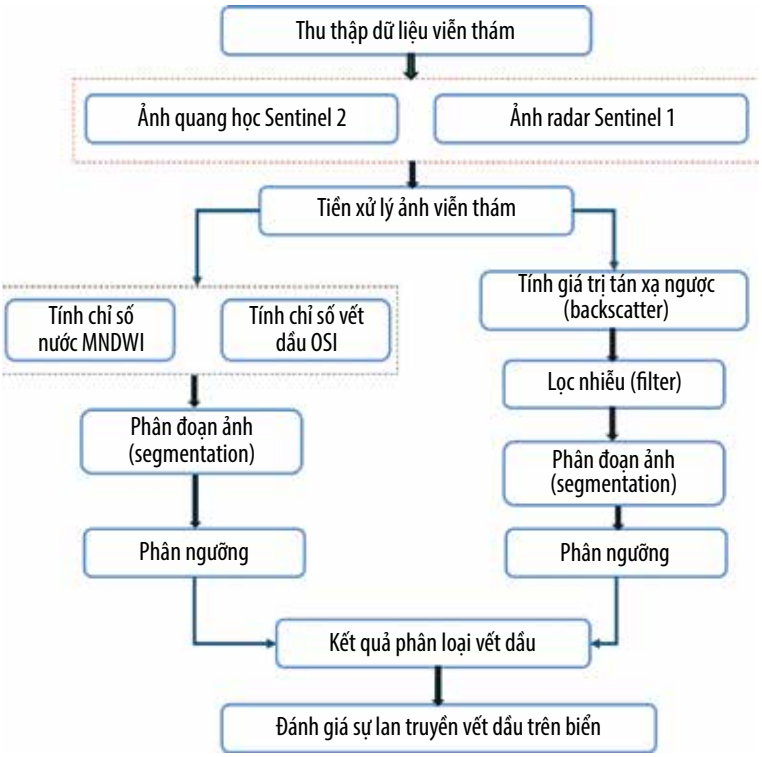
được tối ưu hóa để hoạt động hiệu quả trong môi trường xử lý dữ liệu lớn GEE. Phân đoạn ảnh giúp gom nhóm các điểm ảnh có đặc điểm tương đồng, cụ thể là có giá trị tán xạ ngược (đối với ảnh Sentinel 1) hoặc giá trị phổ (đối với ảnh Sentinel 2) tương tự nhau. SNIC được sử dụng trong nghiên cứu do thuật toán được tích hợp sẵn trong GEE, giúp triển khai thuận tiện và hiệu quả. Các tham số cài đặt của thuật toán là size: 10; compactness: 0,3; connectivity: 8; neighborhood size: 128. Quá trình phân đoạn được thực hiện cho cả ảnh Sentinel 1 và Sentinel 2. Mỗi cụm sau phân đoạn ảnh được đại diện bởi một tâm cụm - là giá trị trung bình của các điểm ảnh trong cụm và được sử dụng trong các bước phân ngưỡng tiếp theo. Cách tiếp cận này giúp tận dụng ngữ cảnh không gian, giảm nhiễu do điểm ảnh đơn lẻ và tăng độ chính xác trong việc phân loại các đối tượng phức tạp trên bề mặt biển, đặc biệt là khi xử lý các hiện tượng có biên mờ như vết dầu loang [19]. Sau khi phân đoạn, ở bước tiếp theo, phương pháp phân ngưỡng tự động Otsu [20] được áp dụng để phân loại vết dầu từ ảnh quang học Sentinel 2.

3.2. Xử lý dữ liệu ảnh radar Sentinel 1

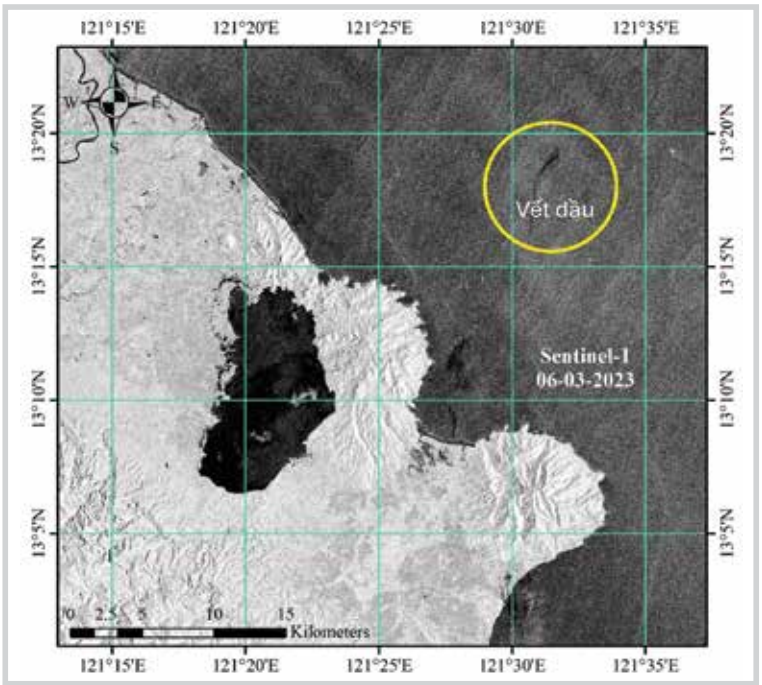
Đối với ảnh radar Sentinel 1, sau khi tiền xử lý, dữ liệu ảnh được dùng để tính giá trị tán xạ ngược (backscatter), sau đó tiến hành lọc nhiễu để loại bỏ ảnh hưởng của nhiễu đốm trên ảnh bằng phương pháp lọc trung bình (mean, cửa sổ lọc 3 × 3). Phương pháp phân đoạn ảnh và phân ngưỡng tự động Otsu cũng được sử dụng để phân loại vết dầu trên ảnh Sentinel 1.

Để nâng cao hiệu quả giám sát và tính toán diện tích vết dầu trên biển, việc kết hợp sử dụng dữ liệu từ Sentinel 1 và Sentinel 2 là một phương pháp hiệu quả. Do cả hai loại dữ liệu đều có độ phân giải không gian 10 m, việc kết hợp sử dụng các ảnh này giúp tạo ra tập dữ liệu đa ảnh, cải thiện độ phân giải thời gian cho bài toán phát hiện và giám sát tràn dầu trên biển.

Sơ đồ quy trình xử lý dữ liệu viễn thám quang học và radar trong phân loại vết dầu trên biển được thể hiện trên Hình 4.



Hình 4. *Sơ đồ quy trình xử lý ảnh.*

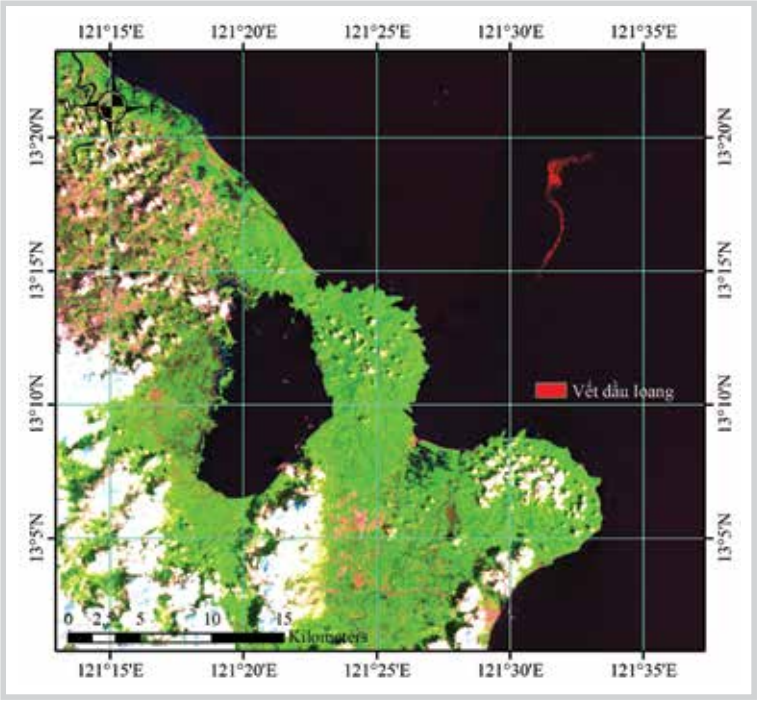


Hình 5. *Vết dầu trên ảnh Sentinel 1 ngày 6/3/2023.*

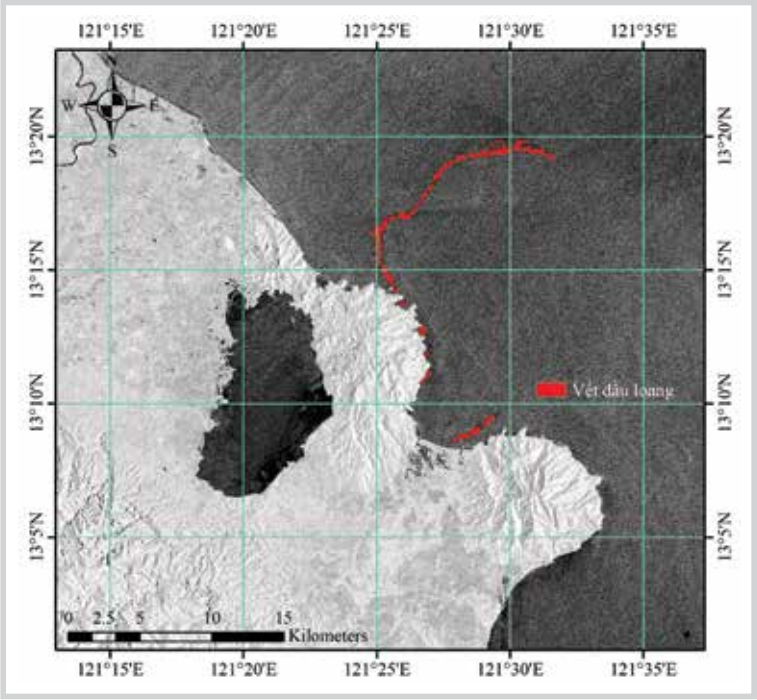
4. Kết quả và thảo luận

Hình 5 thể hiện ảnh vệ tinh Sentinel 1 chụp ngày 6/3/2023, trong đó vết dầu có màu đen, dạng hình tuyến. Vào thời điểm này, vết dầu còn ở ngoài khơi khu vực Naujan, diện tích khoảng 1,2 km^2.

Kết quả phân loại vết dầu bằng chỉ số OSI xác định từ ảnh vệ tinh quang học Sentinel 2 ngày 8/3/2023, sử dụng phương pháp phân



Hình 6. *Kết quả phân loại vết dầu bằng ảnh Sentinel 2 ngày 8/3/2023 sử dụng chỉ số OSI.*



Hình 7. *Kết quả phân loại vết dầu trên ảnh Sentinel 1 ngày 18/3/2023.*

ngưỡng được trình bày trên Hình 6, trong đó vết dầu được thể hiện là màu đỏ. Có thể nhận thấy, vết dầu có sự mở rộng đáng kể so với ngày 6/3/2023 và di chuyển gần hơn về phía đất liền. Diện tích vết dầu xác định trên ảnh Sentinel 2 ngày 8/3/2023 đạt 2,19 km^2.

Đến ngày 18/3/2023, vết dầu đã di chuyển và ảnh hưởng trực tiếp đến khu vực ven biển Naujan, trong đó vết dầu kéo dài từ khu vực xảy ra sự cố vào đến bờ biển (Hình 7). Diện tích vết dầu tính được từ ảnh Sentinel 1 vào ngày 18/3/2023 đạt 6,84 km^2, cao gấp hơn 3 lần so với ngày 8/3/2023.

Để đánh giá độ chính xác kết quả phân loại vết dầu trên ảnh Sentinel 2 bằng chỉ số OSI, trong nghiên cứu đã xây dựng 90 mẫu bao gồm các pixel thuộc lớp vết dầu và các pixel không phải là vết dầu. Các mẫu được lựa chọn dựa trên kinh nghiệm của người xử lý, nhằm đảm bảo tính đại diện và phân bố đồng đều trong toàn ảnh. Kết quả nhận được cho thấy, số lượng điểm mẫu được phân loại đúng bằng phương pháp phân ngưỡng tự động đạt 85,56%. Tương tự, với ảnh Sentinel 1, độ chính xác phân loại vết dầu đạt 91,11%. Có thể nhận thấy, việc sử dụng ảnh radar Sentinel 1 cho phép phân loại vết dầu với độ chính xác cao hơn so với ảnh quang học. Điều này có thể giải thích do đặc điểm vết dầu được phân biệt rõ nét trên ảnh radar, khác với trên ảnh quang học, vết dầu thường bị lẫn với vùng biển xung quanh.

Mặc dù là những công cụ phổ biến trong việc phát hiện vết dầu trên biển, chỉ số OSI và phương pháp phân ngưỡng Otsu vẫn có những hạn chế nhất định. Cụ thể, hiệu quả của các phương pháp này có thể bị ảnh hưởng bởi điều kiện ánh sáng, thời tiết và sự hiện diện của các yếu tố gây nhiễu như mây, sương mù hoặc các hiện tượng tự nhiên khác. Để khắc phục những hạn chế này và nâng cao độ chính xác trong việc phát hiện vết dầu, việc kết hợp dữ liệu từ các cảm biến khác nhau, chẳng hạn như Sentinel 1 và Sentinel 2 cùng với việc áp dụng các phương pháp xử lý ảnh tiên tiến là một hướng đi cần thiết và hiệu quả. Đồng thời, việc kết hợp dữ liệu từ nhiều vệ tinh thương mại khác nhau có thể nâng cao khả năng giám sát liên tục và kịp thời các sự cố tràn dầu trên biển.

5. Kết luận

Dữ liệu viễn thám đa nguồn, gồm ảnh quang học Sentinel 2 và ảnh radar Sentinel 1 được sử dụng để phân loại vết dầu trên biển. Chỉ số vết dầu OSI xác định trên cơ sở các kênh phổ ở dải sóng nhìn thấy ảnh Sentinel 2 được sử dụng để phân loại vết dầu bằng phương pháp phân ngưỡng tự động. Phương pháp phân ngưỡng Otsu cũng được áp dụng cho

ảnh radar Sentinel 1 để phân loại vết dầu trên biển. Kết quả cho thấy vết dầu được phân loại với độ chính xác cao trên 85%.

Việc kết hợp sử dụng dữ liệu viễn thám đa nguồn cho phép tăng dày nguồn dữ liệu đầu vào, từ đó nâng cao hiệu quả trong phát hiện và giám sát ô nhiễm tràn dầu trên biển. Với 3 vệ tinh trong chùm vệ tinh Sentinel 1 và 3 vệ tinh trong chùm vệ tinh Sentinel 2, đặc biệt dữ liệu được cung cấp miễn phí, ảnh Sentinel là nguồn dữ liệu đầu vào quý giá trong nghiên cứu môi trường biển nói chung, ô nhiễm tràn dầu nói riêng.

Tài liệu tham khảo

[1] Nguyễn Đình Dương, *Ô nhiễm dầu trên biển và quan trắc bằng viễn thám siêu cao tần*. Nhà xuất bản Khoa học và Kỹ thuật, 2011, trang 107 - 137.

[2] Damián Mira Martínez, Pablo Gil, Beatriz Alacid, and Fernando Torres, "Oil spill detection using segmentation-based approaches", *6th International Conference on Pattern Recognition Applications and Methods, Porto (Portugal), February 2017*. DOI:10.5220/0006191504420447.

[3] Yonglei Fan, Xiaoping Rui, Guangyuan Zhang, Tian Yu, Xijie Xu, and Stefan Posld, "Feature merged network for oil spill detection using SAR images", *Remote Sensing*, Volume 13, Issue 16, 2021. DOI: 10.3390/rs13163174.

[4] Fahad A.M. Alawadi, *"Detection and classification of oil spills in MODIS satellite imagery"*, University of Southampton, School of Ocean and Earth Science, Doctoral Thesis, 2011.

[5] Teodosio Lacava, Emanuele Ciancia, Irina Coviello, Carmine Di Polito, Caterina S.L. Grimaldi, Nicola Pergola, Valeria Satriano, Marouane Temimi, Jun Zhao, and Tramutoli Valerio, "A MODIS-based robust satellite technique (RST) for timely detection of oil spilled areas", *Remote Sensing*, Volume 9, Issue 2, 2017. DOI: 10.3390/rs9020128.

[6] Polychroni Kolokoussis and Vassilia Karathanassi "Oil spill detection and mapping using Sentinel 2 imagery", *Journal of Marine Science and Engineering*, Volume 6, Issue 1, 2018. DOI: 10.3390/jmse6010004.

[7] Alireza Taravat and Fabio Del Frate, "Development of band rationing algorithm and neural networks to detection of oil spills using Landsat ETM+ data", *Eurasip Journal on Advances in Signal Processing*, 2012. DOI: 10.1186/1687-6180-2012-107.

[8] Alaa Sheta, Mouhammd Alkasassbed, Malik Braik, and Hafsa Abu Ayyash, "Detection of oil spills in SAR images using threshold segmentation algorithms", *International Journal of Computer Applications*, Volume 57, Issue 7, pp. 10 - 15, 2012.

[9] Fangjie Yu, Wuzi Sun, Jiaojiao Li, Yang Zhao, Yanmin Zhang, and Ge Chen, "An improved Otsu method for oil spill detection from SAR images", *Oceanologia*, Volume 59, Issue 3, pp. 311 - 317, 2017. DOI: 10.1016/j.oceano.2017.03.005.

[10] Alaa Akram Huby, Rafid Sagban, and Raaid Alubady, "Oil spill detection based on machine learning and deep learning: A review", *5th International Conference on Engineering Technology and its Applications (IICETA), Al-Najaf, Iraq, 2022*. DOI: 10.1109/IICETA54559.2022.9888651.

[11] Ngoc An Bui, Youngon Oh, and Impyeong Lee, "Oil spill detection and classification through deep learning and tailored data augmentation", *International Journal of Applied Earth Observation and Geoinformation*, Volume 129, 2024. DOI: 10.1016/j.jag.2024.103845.

[12] Ujjawal Singh, Divya Acharya, and Saurabh Mishra, "Oil spill detection and monitoring with artificial intelligence: A futuristic approach", *Workshop on Control and Embedded Systems, Chennai, India, 22 - 24 April 2022*.

[13] Zhen Sun, Qingshu Yang, Nanyang Yan, Siyu Chen, Jianhang Zhu, Jun Zhao, and Shaojie Sun, "Utilizing deep learning algorithms for automated oil spill detection in medium resolution optical imagery", *Marine Pollution Bulletin*, Volume 206, 2024. DOI: 10.1016/j.marpolbul.2024.116777.

[14] Rodrigo N. Vasconcelos, Carlos A.D. Lentini, André T. Cunha Lima, Luís F.F. Mendonça, Garcia V. Miranda, and Elaine C.B. Cambuí, "Oil spill detection based on texture analysis: How does feature importance matter in classification", *International Journal of Remote Sensing*, Volume 43, Issue 11, pp. 4045 - 4064, 2022. DOI:10.1080/01431161.2022.2106163.

[15] Gabriel Pabico Lalu, "Oil spill nearing shoreline of Pola, Oriental Mindoro based on satellite photos - PhilSA", 17/3/2023. [Online]. Available: https://newsinfo.inquirer.net/1744654/philsa-says-satellite-photos-show-oil-spill-nearing-shoreline.

[16] Sankaran Rajendran, Ponnumony Vethamony, Fadhil N. Sadooni, Hamad Al-Saad Al-Kuwari, Jassim A. Al-Khayat, Himanshu Govil, and Sobhi Nasir, "Sentinel-2

image transformation methods for mapping oil spill - A case study with Wakashio oil spill in the Indian Ocean, off Mauritius", *MethodsX*, Volume 8, 2021. DOI: 10.1016/j.mex.2021.101327.

[17] Sankaran Rajendran, Ponnumony Vethamony, Fadhil N. Sadooni, Hamad Al Saad Al-Kuwari, Jassim A. Al-Khayat, Vashist O. Seegobin, Himanshu Govil, and Sobhi Nasir, "Detection of Wakashio oil spill off Mauritius using Sentinel-1 and 2 data: Capability of sensors, image transformation methods and mapping", *Environmental Pollution*, Volume 274, 2021. DOI: 10.1016/j.envpol.2021.116618.

[18] Hanqiu Xu, "Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery", *International Journal of Remote Sensing*, Volume 27, Issue 14, pp. 3025 - 3033, 2006.

[19] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 34, Issue 11, pp. 2274 - 2282, 2012. DOI: 10.1109/TPAMI.2012.120.

[20] Nobuyuki Otsu, "A threshold selection method from gray-level histograms", *IEEE Transactions on Systems, Man, and Cybernetics*, Volume 9, Issue 1, pp. 62 - 66, 1979. DOI: 10.1109/TSMC.1979.4310076.

INTEGRATION OF OPTICAL AND RADAR REMOTE SENSING IMAGES FOT CLASSIFYING AND MONITORING OIL SPILLS AT SEA

Trinh Le Hung, Le Van Phu
Military Technical Academy
Email: trinhlehung@lqdtu.edu.vn

Summary

Remote sensing data has been widely used in the world in the study of oil spill pollution at sea. This paper presents a study combining the use of Sentinel 2 MSI optical remote sensing and Sentinel 1 radar images to detect and classify oil spills. Sentinel 2 MSI data is used to calculate the OSI (oil spill index) based on visible bands, while Sentinel 1 data is used to calculate the backscatter value, from which oil spills are classified by the thresholding method. The integration of multi-type remote sensing data allows to enhance the density of the input dataset, helping to improve the effectiveness of monitoring marine oil spill pollution.

Key words: Optical remote sensing image, radar remote sensing image, oil spill, classification.

DầuKhí
KHOA HỌC, CÔNG NGHỆ VÀ ĐỔI MỚI SÁNG TẠO PETROVIETNAM

TRÍ TUỆ NHÂN TẠO TRONG TIÊU THỤ VÀ TỐI ƯU HÓA HỆ THỐNG NĂNG LƯỢNG

Tóm tắt

Sự phát triển của trí tuệ nhân tạo (AI) đang định hình lại nhiều ngành công nghiệp, trong đó có lĩnh vực năng lượng. AI không chỉ là công cụ công nghệ mà còn là nhân tố chuyển đổi chiến lược, mang đến thách thức và cơ hội kép. Một mặt, việc đào tạo và vận hành các mô hình AI, đặc biệt là AI tạo sinh (Generative AI), đòi hỏi lượng điện năng khổng lồ từ các trung tâm dữ liệu (data center - DC), tạo áp lực lớn lên nguồn cung và lưới điện toàn cầu. Mặt khác, AI sở hữu tiềm năng lớn để tối ưu hóa, nâng cao hiệu suất và tăng cường khả năng chống chịu của hệ thống năng lượng đang ngày càng phức tạp, phi tập trung hóa và được số hóa.

Cơ quan Năng lượng Quốc tế (IEA) đưa ra các kịch bản dự báo tiêu thụ điện của AI đến năm 2030, đánh giá lợi ích tiềm năng của AI trong các lĩnh vực trọng yếu (lưới điện thông minh, hiệu suất năng lượng) và đề xuất chính sách cần thiết nhằm đạt được sự cân bằng tối ưu giữa "năng lượng cho AI" và "AI cho năng lượng".

Từ khóa: Trí tuệ nhân tạo, AI cho năng lượng.

1. Giới thiệu

Ngành công nghiệp năng lượng toàn cầu đang đối mặt với 3 vấn đề lớn gồm: an ninh năng lượng, chi phí hợp lý và sự bền vững. Để giải quyết thách thức này, hệ thống năng lượng đang chuyển dịch theo 3 xu hướng chính: khử carbon (decarbonisation) thông qua năng lượng tái tạo, phi tập trung hóa (decentralisation) với sự tham gia của các nguồn năng lượng phân tán (DERs) và số hóa (digitalisation).

Đối với ngành năng lượng, AI được xem là công cụ quan trọng để quản lý tính phức tạp của lưới điện tích hợp năng lượng tái tạo biến đổi (VRE) và hàng triệu điểm kết nối nhỏ lẻ. Tuy nhiên, sự phát triển của AI không phải là không có chi phí. Hạ tầng tính toán của AI, chủ yếu là các trung tâm dữ liệu, đang trở thành một trong những phụ tải điện tăng trưởng nhanh nhất trên thế giới.

Một trung tâm dữ liệu tập trung vào AI điển hình tiêu thụ lượng điện tương đương 100.000 hộ gia đình, nhưng những trung tâm lớn nhất đang được xây dựng ngày nay sẽ tiêu thụ gấp 20 lần.

Tại các thị trường mới nổi như Đông Nam Á, nhất là Việt Nam, trí tuệ nhân tạo đang vươn lên như một trụ cột chiến lược trong hành trình chuyển đổi số và phát triển kinh tế quốc gia. Theo dự báo, thị trường AI tại Việt Nam sẽ tăng trưởng trung bình 15,8% mỗi năm, đạt quy mô 1,52 tỷ USD vào năm 2030. AI có thể đóng góp tới 130 tỷ USD cho nền kinh tế Việt Nam vào năm 2040.

AI đang thúc đẩy nhu cầu dữ liệu tăng vọt chưa từng có. Đặc biệt, dữ liệu đang trở thành tài sản chiến lược, việc lưu trữ, xử lý và bảo mật dữ liệu trong nước có vai trò quan trọng trong bối cảnh địa chính trị biến động.

Tuy nhiên, sự bùng nổ AI đi kèm mức tiêu thụ năng lượng tăng vọt. Theo IEA, dự báo đến năm 2028, AI có thể chiếm 15 - 20% tổng mức tiêu thụ điện của các trung tâm dữ liệu - tăng đáng kể so với mức 8% vào năm 2023. Điều này đồng nghĩa với việc cần hệ thống làm mát hiệu quả hơn và mật độ tính toán cao hơn so với các hệ thống công nghệ thông tin truyền thống.

Cùng lúc đó, mô hình xử lý dữ liệu đang chuyển dịch từ tập trung sang phân tán, với khoảng 50% tác vụ AI dự kiến sẽ được xử lý theo mô hình hybrid, nghĩa là kết hợp giữa trung tâm dữ liệu và xử lý tại biên, giúp tăng tốc độ xử lý, giảm độ trễ.

Nghiên cứu của IEA tập trung phân tích định lượng và định tính toàn diện về mối quan hệ cộng sinh giữa AI và năng lượng: quy mô, tốc độ và các yếu tố thúc đẩy nhu cầu điện năng từ AI/DC trên toàn cầu; tiềm năng của AI trong việc tối ưu hóa chuỗi giá trị năng lượng và thúc đẩy các mục tiêu bền vững; đề xuất các biện pháp chính sách để quản lý thách thức về nguồn cung và tận dụng tối đa lợi ích của AI.

2. AI trong tiêu thụ và tối ưu hóa hệ thống năng lượng

2.1. AI và an ninh năng lượng

An ninh năng lượng được đặc trưng bởi một số yếu tố bao gồm, nhưng không giới hạn ở (i) khả năng tiếp cận năng lượng đáng tin cậy để đáp ứng nhu cầu của nền kinh tế; (ii) khả năng chi trả năng lượng này với mức biến động giá hạn chế và (iii) khả năng phục hồi trước các "cú sốc" của thị trường năng lượng - hoặc khả năng phục hồi nhanh chóng của hệ thống năng lượng từ các "cú sốc" đó. Các ứng dụng AI giải quyết 1 hoặc nhiều khía cạnh này bao gồm:

- Giảm chi phí năng lượng: Các ứng dụng AI đang được sử dụng trong nhiều lĩnh vực khác nhau, bao gồm đánh giá tài nguyên và tối ưu hóa các quy trình, giúp rút ngắn thời gian phát triển và giảm chi phí. Ví dụ, việc áp dụng mô phỏng dựa trên AI được ước tính giảm chi phí gần 10% trong các hoạt động dầu khí ngoài khơi. Các kết quả tương tự cũng được quan sát thấy đối với năng lượng tái tạo, ví dụ như khi các mô hình AI được triển khai để tối ưu hóa hoạt động của trang trại điện gió, giúp giảm chi phí vận hành.

- Bảo vệ cơ sở hạ tầng năng lượng quan trọng: AI được ứng dụng trong việc đảm bảo an ninh của cơ sở hạ tầng năng lượng quan trọng ở những khu vực con người khó tiếp cận. Ví dụ, sau vụ phá hoại đường ống Nord Stream năm 2022, NATO đã sử dụng các hệ thống hàng hải không người lái được hỗ trợ bởi AI có thể giúp xác định hoạt động đáng ngờ dưới nước và ngăn chặn gián đoạn nguồn cung năng lượng (WSJ, 2025).

- Khả năng phục hồi của hệ thống năng lượng thông qua dự báo thời tiết tốt hơn: Dự báo thời tiết chính xác và phân tích các mô hình thời tiết thay đổi trong bối cảnh thế giới đang nóng lên là điều cần thiết để tối ưu hóa hoạt động, quy hoạch và khả năng phục hồi của các hệ thống năng lượng. Thời gian tính toán dự báo thời tiết có thể được rút ngắn từ vài giờ xuống chỉ 1 phút nhờ các ứng dụng AI, sử dụng một phần nghìn lượng điện.

- Bảo trì dự đoán để tăng cường độ tin cậy: Bảo trì dự đoán dựa trên AI đang cách mạng hóa quản lý cơ sở hạ tầng năng lượng bằng cách đảm bảo giảm thời gian ngừng hoạt động và cải thiện hiệu quả vận hành.

- Phân tích dự báo sự ổn định của lưới điện và tích hợp năng lượng tái tạo: Khi tỷ trọng phát điện năng lượng tái tạo biến đổi tăng lên, các thuật toán AI có thể cải thiện việc điều phối các nguồn năng lượng, quan trọng để xử lý các hệ thống điện với tỷ trọng năng lượng tái tạo cao. Việc tích hợp tăng cường năng lượng tái tạo được tạo ra trong nước cũng làm giảm sự phụ thuộc vào nhiên liệu nhập khẩu.

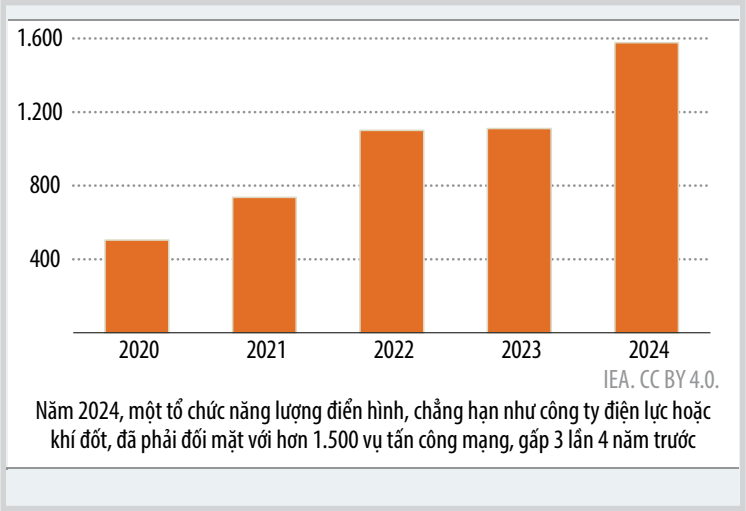
- Tăng cường an ninh mạng để bảo vệ cơ sở hạ tầng quan trọng: Khi các hệ thống năng lượng ngày càng được điện khí hóa, tích hợp và kết nối, tính dễ bị tổn thương trước các cuộc tấn công mạng cũng tăng lên. Các tính năng an ninh mạng được hỗ trợ bởi AI, ví dụ như phát hiện mối đe dọa sớm hơn, có thể giúp bảo vệ các hệ thống năng lượng kịp thời hơn.

Đây chỉ là một số ứng dụng AI trong tăng cường an ninh năng lượng, đáp ứng khả năng chi trả, khả năng phục hồi và độ tin cậy cao hơn, đảm bảo nguồn cung năng lượng cho nhu cầu trong nước.

2.2. AI và an ninh mạng trong ngành năng lượng

Khi ngành năng lượng đã được điện khí hóa, số hóa và kết nối nhiều hơn, cũng đồng nghĩa với việc dễ bị tổn thương trước các mối đe dọa an ninh mạng. Nguyên nhân do cơ sở hạ tầng công nghệ thông tin kế thừa, tự động hóa, điện toán đám mây và sự phụ thuộc vào các nhà cung cấp bên thứ ba có thể không có hệ thống bảo mật. Thực tế đã có nhiều trường hợp hệ thống năng lượng bị tấn công kể từ trường hợp đầu tiên được biết đến khi xảy ra cuộc tấn công mạng dẫn đến mất điện ở Ukraine năm 2015 ảnh hưởng đến 225.000 người (IEA, 2020), 1 cuộc tấn công bằng phần mềm độc hại vào lưới điện Mumbai dẫn đến mất điện ở Ấn Độ năm 2020 (India Today, 2021) và cuộc tấn công năm 2021 dẫn đến gián đoạn hoạt động tại hệ thống đường ống dầu lớn nhất thế giới, cung cấp 40 - 45% nhiên liệu ở miền Đông nước Mỹ (IEA, 2021). Phân tích cho thấy dự án khí đốt và điện điển hình phải đối mặt với hơn 1.500 cuộc tấn công mỗi tuần vào năm 2024 (Checkpoint, 2025), gấp 3 lần con số 4 năm trước đó (IEA, 2023a).

AI tăng cường phát hiện mối đe dọa và cho phép bảo vệ phản ứng nhanh hơn, trong khi đó các công cụ tấn công ngày càng tinh vi hơn. Các ứng dụng AI có thể cho phép phát hiện mối đe dọa theo thời gian thực, phản ứng tự động với các sự cố và tăng cường phòng thủ. Mặt khác, các công cụ dựa trên AI cũng có thể bị khai thác để tự động hóa các cuộc tấn công và trốn tránh phát hiện. Các công cụ AI tạo sinh được sử dụng bởi các tác nhân độc hại để truy cập sâu hơn vào mạng mục tiêu, với kịch bản và kỹ thuật trốn tránh độc hại (Google, 2025). Trước những mối đe dọa đang phát triển này, việc triển khai các hệ thống



Hình 1. *Số vụ tấn công mạng mỗi tuần của một tổ chức năng lượng điển hình giai đoạn 2020 - 2024.*

an ninh mạng được hỗ trợ bởi AI chủ động hơn, phản ứng nhanh với các mối đe dọa là rất quan trọng để đảm bảo khả năng phục hồi của ngành năng lượng.

2.3. An ninh của chuỗi cung ứng ngành năng lượng cho AI

Đảm bảo nguồn cung cấp điện giá cả phải chăng và đáng tin cậy cho các trung tâm dữ liệu là trọng tâm của thách thức năng lượng cho AI. IEA nghiên cứu an ninh của chuỗi cung ứng cho AI, bao gồm phát điện, truyền tải, thiết bị điện và cung cấp các loại khoáng sản quan trọng.

Trong khi năng lượng tái tạo cung cấp ¼ sản lượng điện cho trung tâm dữ liệu, khí tự nhiên và than vẫn đóng vai trò quan trọng, đặc biệt là ở Mỹ và Trung Quốc. Để đáp ứng nhu cầu điện ngày càng tăng trong tương lai, các công ty công nghệ đã hỗ trợ các lựa chọn cung cấp mới, bao gồm hạt nhân, địa nhiệt tiên tiến và lưu trữ điện trong thời gian dài.

Trong khi đó, các doanh nghiệp năng lượng đã chủ động lập kế hoạch xây dựng các cơ sở phát điện chuyên dụng hoặc cung cấp năng lượng để đáp ứng nhu cầu trung tâm dữ liệu. Ví dụ, Chevron đã hợp tác với Engine No. 1 để phát triển các nhà máy điện khí công suất 4 GW với turbine từ GE Vernova tại Mỹ, bỏ qua lưới truyền tải. Sáng kiến này cũng để ngỏ lựa chọn kết hợp thu giữ và lưu trữ carbon và năng lượng tái tạo. Exxon Mobil đang xem xét mô hình tương tự, với việc phát triển nhà máy điện khí công suất 1,5 GW để cung cấp cho các trung tâm dữ liệu quy mô lớn, cùng với kế hoạch bổ sung thu giữ và lưu trữ carbon có khả năng thu giữ hơn 90% lượng phát thải.

Nhu cầu khí tự nhiên để cung cấp điện cho các trung tâm dữ liệu dự kiến sẽ tăng gần 35 tỷ m^3 trên toàn cầu từ năm 2024 đến 2035 trong Kịch bản cơ sở và cao tới 55 tỷ m^3 trong Kịch bản cất cánh. Nhu cầu bổ sung trong cả 2 trường hợp chủ yếu phát sinh ở Mỹ, với tài nguyên khí đá phiến và dầu chặt sít. Sản lượng gia tăng từ các bể Permian, Haynesville và Marcellus đã đưa sản lượng khí tự nhiên của Mỹ đạt gần 1.200 tỷ m^3 vào năm 2024. Khoảng 80% nguồn cung này được tiêu thụ trong nước, với 10% xuất khẩu dưới dạng khí tự nhiên hóa lỏng, còn lại là xuất khẩu qua đường ống.

IEA cho rằng các doanh nghiệp cung cấp khí phải có tầm nhìn rõ ràng về quy mô tăng trưởng nhu cầu trung tâm dữ liệu. Ví dụ, trong Kịch bản cất cánh, nếu điện khí đáp ứng toàn bộ sự gia tăng từ nhu cầu điện của trung tâm dữ liệu ở Mỹ trong thập kỷ tới, sẽ cần hơn 100 tỷ m3 vào năm 2035. Do đó, tác động về giá có thể lớn hơn nhiều nếu nhu cầu bổ sung này không được lập kế hoạch, công suất vận chuyển qua đường ống hoặc thỏa thuận cung cấp với các tiện ích và nhà điều hành trung tâm dữ liệu.

2.4. Cơ sở hạ tầng lưới điện cho AI

Ngoài cung cấp năng lượng cho các trung tâm dữ liệu, tính khả dụng của cơ sở hạ tầng truyền tải điện cũng là yếu tố quyết định chính của an ninh năng lượng cho AI. Các trung tâm dữ liệu chậm kết nối với lưới điện, với sự chậm trễ lên đến 10 năm ở một số thị trường chính. Khoảng 1/5 trong tổng số dự án xây dựng trung tâm dữ liệu toàn cầu trong Kịch bản cơ sở có nguy cơ bị trì hoãn do các điểm nghẽn của lưới điện.

2.5. Chuỗi cung ứng thiết bị điện cho AI

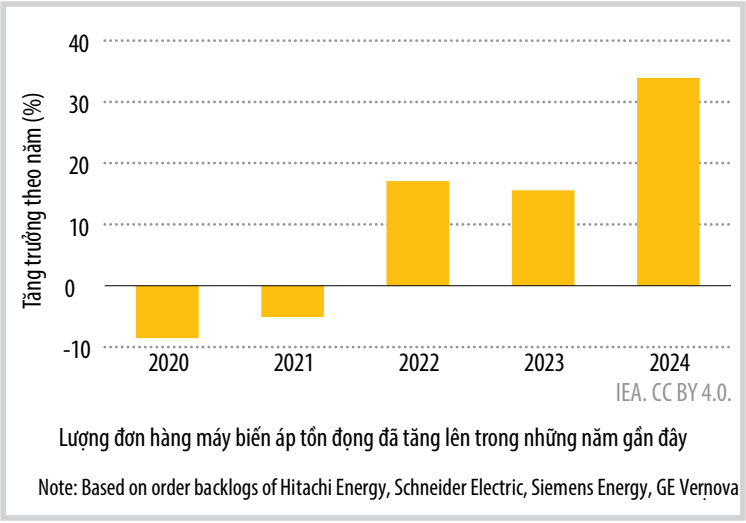
Sự mở rộng của các trung tâm dữ liệu AI đã làm tăng tính cấp bách của việc giải quyết các ràng buộc chuỗi cung ứng thiết bị điện. Mở rộng cơ sở hạ tầng trên nhiều khu vực đã gây áp lực đáng kể lên chuỗi cung ứng. Nhu cầu gia tăng mở rộng ra ngoài thiết bị cho truyền tải điện áp cao để bao gồm các giải pháp điện áp thấp, tích hợp các nguồn năng lượng biến đổi và nhu cầu người tiêu dùng mới, khiến khả năng phục hồi của chuỗi cung ứng trở nên quan trọng hơn bao giờ hết.

Kết quả khảo sát của IEA cho thấy giá cáp điện đã tăng gần gấp đôi trong 5 năm qua, trong khi đó giá máy biến áp cũng tăng vọt kể từ năm 2022, tùy theo độ phức tạp và thiết kế, có loại tăng gấp 2,6 lần so với mức giá trước đại

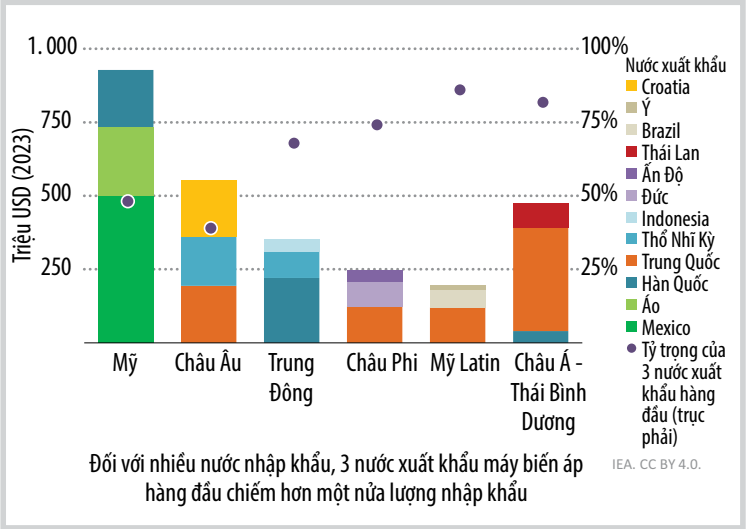
dịch. Các điều kiện lắp đặt khó khăn càng làm tăng thêm chi phí, gây thêm áp lực lên các chuỗi cung ứng vốn đã căng thẳng và các kế hoạch đầu tư cho cơ sở hạ tầng truyền tải. Thời gian thực hiện máy biến áp đã tăng gần gấp đôi trong 2 - 3 năm qua, các nhà sản xuất lớn phải đối mặt với số lượng đơn hàng tồn đọng kỷ lục.

Thị trường toàn cầu đã phản ứng tích cực với sự gia tăng nhu cầu thiết bị điện, công bố các kế hoạch mở rộng công suất và đầu tư mới. Tuy nhiên, việc mở rộng quy mô công suất cần có thời gian, ví dụ thường cần 3 - 4 năm cho 1 cơ sở sản xuất cáp.

Nhu cầu về cơ sở hạ tầng lưới điện đã đẩy chi phí thành phần lên cao, cùng với các yếu tố khác như lạm phát, gián đoạn logistics toàn cầu, biến động giá nguyên liệu và chi phí năng lượng tăng ở một số thị trường. Do đó, việc đảm bảo nguồn cung chiến lược và khả năng phục hồi của chuỗi cung ứng sẽ là chìa khóa để đáp ứng nhu cầu năng lượng ngày càng tăng.



Hình 2. *Gia tăng đơn hàng máy biến áp điện tồn đọng tại một số công ty sản xuất giai đoạn 2020 - 2024.*



Hình 3. *Giá trị nhập khẩu máy biến áp theo quốc gia hoặc khu vực nhập khẩu năm 2024.*

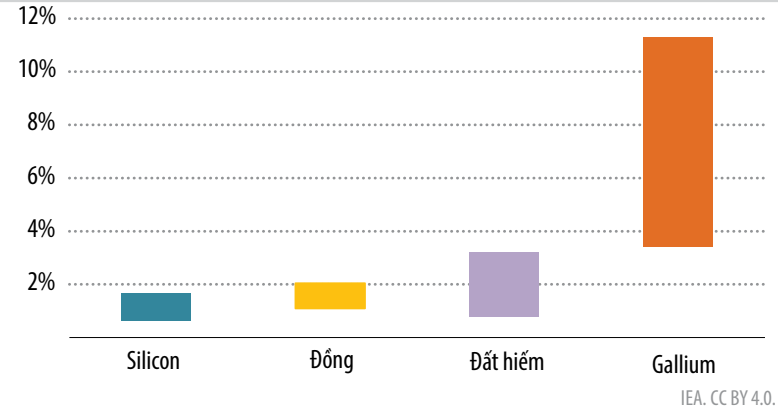
2.6. Chuỗi cung ứng khoáng sản quan trọng cho AI

Một số loại khoáng sản quan trọng như: đồng, nhôm, silicon, gallium, nguyên tố đất hiếm… cần thiết để xây dựng các trung tâm dữ liệu mới và các công nghệ năng lượng (IEA, 2024).

Đồng được sử dụng trong các hệ thống phân phối điện (cáp, thanh dẫn và thiết bị đóng cắt), trong mạng hiệu suất cao và cáp dữ liệu, và trong cơ sở hạ tầng làm mát cho bộ trao đổi nhiệt và ống. Silicon, đặc biệt là silicon siêu tinh khiết, là vật liệu bán dẫn chính được sử dụng trong bộ xử lý và các thành phần bộ nhớ và lưu trữ tốc độ cao. Gallium, thường ở dạng gallium nitride hoặc gallium arsenide, được sử dụng cho các bộ chuyển đổi điện tần số cao và hiệu suất cao và các thành phần tần số vô tuyến. Nhôm được sử dụng cho giá máy chủ, vỏ bọc và cấu trúc lắp đặt, cũng như trong các bộ tản nhiệt và tấm làm mát trong hệ thống làm mát nhờ trọng lượng nhẹ và khả năng dẫn nhiệt vượt trội. Các nguyên tố đất hiếm, đặc biệt là neodymium, praseodymium, dysprosium và terbium, tìm thấy ứng dụng trong nam châm hiệu suất cao cho động cơ trong quạt làm mát, bộ truyền động chính xác, cụm ổ cứng.

IEA dự báo nhu cầu khoáng sản cho việc mở rộng công suất trung tâm dữ liệu vào năm 2030 so với năm 2024 có thể đạt tăng 2% cho đồng và silicon, hơn 3% cho các nguyên tố đất hiếm và 11% cho gallium (Hình 4). Nguồn cung khoáng sản trong thập kỷ tới cần phải tính đến nhu cầu bổ sung từ các trung tâm dữ liệu. Một số ngành, chẳng hạn như quốc phòng, sản xuất công nghệ năng lượng sạch, xây dựng, hàng không và trung tâm dữ liệu, sẽ cạnh tranh để giành nguồn cung cấp các khoáng sản quan trọng này trong tương lai.

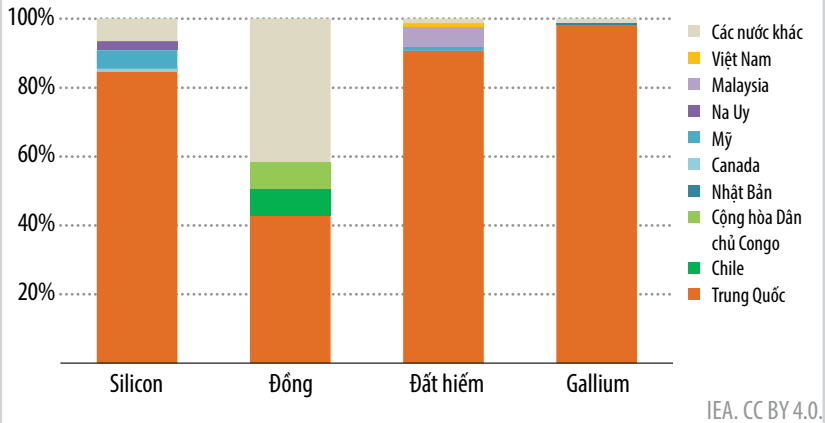
Trong năm 2024, 3 quốc gia chiếm gần 60% nguồn cung tinh chế của đồng, khoảng 90% nhôm và hơn 90% silicon, đất hiếm nam châm và gallium (Hình 5). Sự tập trung thị trường cao này thể hiện điểm yếu đáng kể trước các cú sốc cung ứng nếu, vì bất kỳ lý do gì, nguồn cung từ các nhà sản xuất lớn bị gián đoạn.



Sự phát triển của trung tâm dữ liệu đến năm 2030 sẽ có những tác động khác nhau đến nhu cầu khoáng sản; nhu cầu về kim loại truyền thống chiếm tỷ trọng nhỏ trong tổng nhu cầu, nhưng an ninh nguồn cung khoáng sản vẫn rất quan trọng

Note: The bands for each mineral represent the estimated range of their demand from AI data centres in 2030 as a share of their total demand in 2024.

Hình 4. *Tỷ lệ nhu cầu khoáng sản thiết yếu phục vụ trung tâm dữ liệu năm 2030 so với tổng nhu cầu năm 2024.*



Việc cung cấp nhiều khoáng sản đã qua chế biến cần thiết cho việc xây dựng các trung tâm dữ liệu tập trung cao ở một số ít khu vực, khiến chuỗi cung ứng dễ bị gián đoạn.

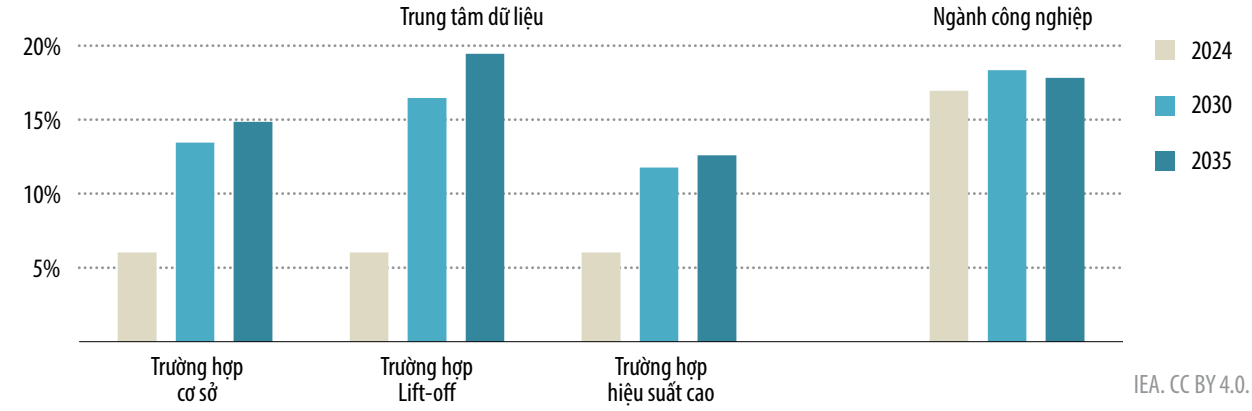
Note: DRC = Democratic Republic of the Congo

Hình 5. *Mức độ tập trung theo khu vực của nguồn cung các khoáng sản thiết yếu đã qua chế biến phục vụ mở rộng trung tâm dữ liệu năm 2024.*

Trung Quốc công bố chính sách hạn chế xuất khẩu gallium, germanium và antimony - các khoáng chất chính cho sản xuất bán dẫn - sang Mỹ. Thống kê cho thấy giá gallium ngoài Trung Quốc đã tăng hơn gấp đôi từ tháng 7/2023 đến tháng 12/2024 (Financial Times, 2024). Đồng thời, Trung Quốc công bố thêm các biện pháp kiểm soát xuất khẩu đối với graphite (thiết yếu cho cực âm pin lithium-ion); tungsten, tellurium, bismuth, indium và molybdenum - các khoáng chất chính chủ yếu được sử dụng trong các ứng dụng công nghệ cao và quốc phòng, bao gồm cả trung tâm dữ liệu (bộ vi xử lý và diode). Gián đoạn nguồn cung khoáng sản quan trọng có thể có tác động lớn đến chi phí công nghệ và thiết bị cho phát triển trung tâm dữ liệu (IEA, 2025).

2.7. Tích hợp các trung tâm dữ liệu để giảm thiểu rủi ro

Tại Mỹ, tổng công suất lắp đặt trung tâm dữ liệu tăng từ 6% lên 13% vào năm 2030 trong Kịch bản cơ sở và 16% trong Kịch bản cất cánh. Các trung tâm dữ liệu đang sẵn sàng chuyển từ các tác nhân ngoại vi sang trung tâm trong hệ thống điện, với tỷ trọng nhu cầu đỉnh có thể so sánh với toàn bộ khu vực công nghiệp của Mỹ trong một số trường hợp (Hình 6).



Công suất lắp đặt của các trung tâm dữ liệu vượt quá 10% nhu cầu điện cao điểm tại Mỹ vào năm 2030 và trong một số trường hợp còn vượt quá toàn bộ ngành công nghiệp

Hình 6. *Tỷ lệ nhu cầu điện cao điểm của các trung tâm dữ liệu và tỷ lệ nhu cầu điện cao điểm của ngành công nghiệp Mỹ.*

Mặc dù quy mô của nguồn cung điện và đầu tư lưới điện cần thiết không phải là vấn đề cấp bách nhất, nhưng tốc độ phát triển mới là vấn đề. Hệ thống điện phải chịu một số "điểm nghẽn" nghiêm trọng có thể khiến việc xây dựng hệ thống và kết nối các trung tâm dữ liệu mới trở thành thách thức. Đó là thời gian thực hiện các dự án phát điện và truyền tải kéo dài, quy trình cấp phép phức tạp và tốn thời gian. IEA dự báo khoảng 1/5 công suất bổ sung trung tâm dữ liệu toàn cầu có thể bị trì hoãn nếu các thách thức này không được giải quyết.

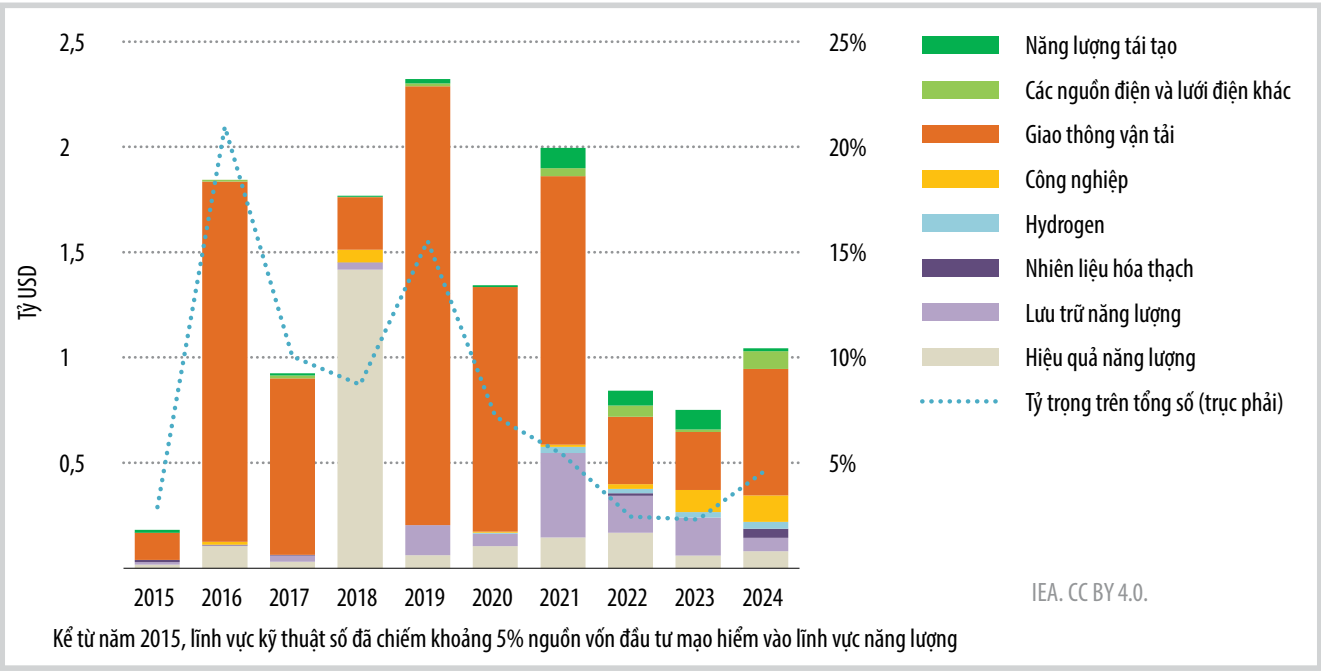
Việc tích hợp trung tâm dữ liệu vào lưới điện được khuyến khích lựa chọn vị trí kết nối với lưới điện linh hoạt hơn, kết hợp nguồn điện dự phòng, lưu trữ năng lượng hoặc nguồn điện riêng; công nghệ mới có thể được tích hợp vào các trung tâm dữ liệu như: lưu trữ năng lượng nhiệt để quản lý tải làm mát, điều chỉnh khối lượng công việc của trung tâm dữ liệu linh hoạt hơn, khi có thể. Tuy nhiên, các trung tâm dữ liệu là tác nhân mới trong lưới điện, cần tăng cường sự hiểu biết giữa các cơ quan quản lý năng lượng và các nhà hoạch định chính sách về các ràng buộc kỹ thuật, đặc điểm hoạt động và độ nhạy cảm với các ưu đãi chính sách.

Ngành năng lượng đang đối mặt với sự không chắc chắn về các triển vọng nhu cầu đối với AI và các trung tâm dữ liệu sẽ phát triển (Hình 7). Đối với năm 2030, kịch bản cao nhất trong các nghiên cứu đã công bố này gần gấp đôi so với Kịch bản cất cánh của IEA và trong các tài liệu kịch bản, dự báo nhu cầu điện cao nhất cho trung tâm dữ liệu đến năm 2030 cao gấp 7 lần mức dự báo thấp nhất.

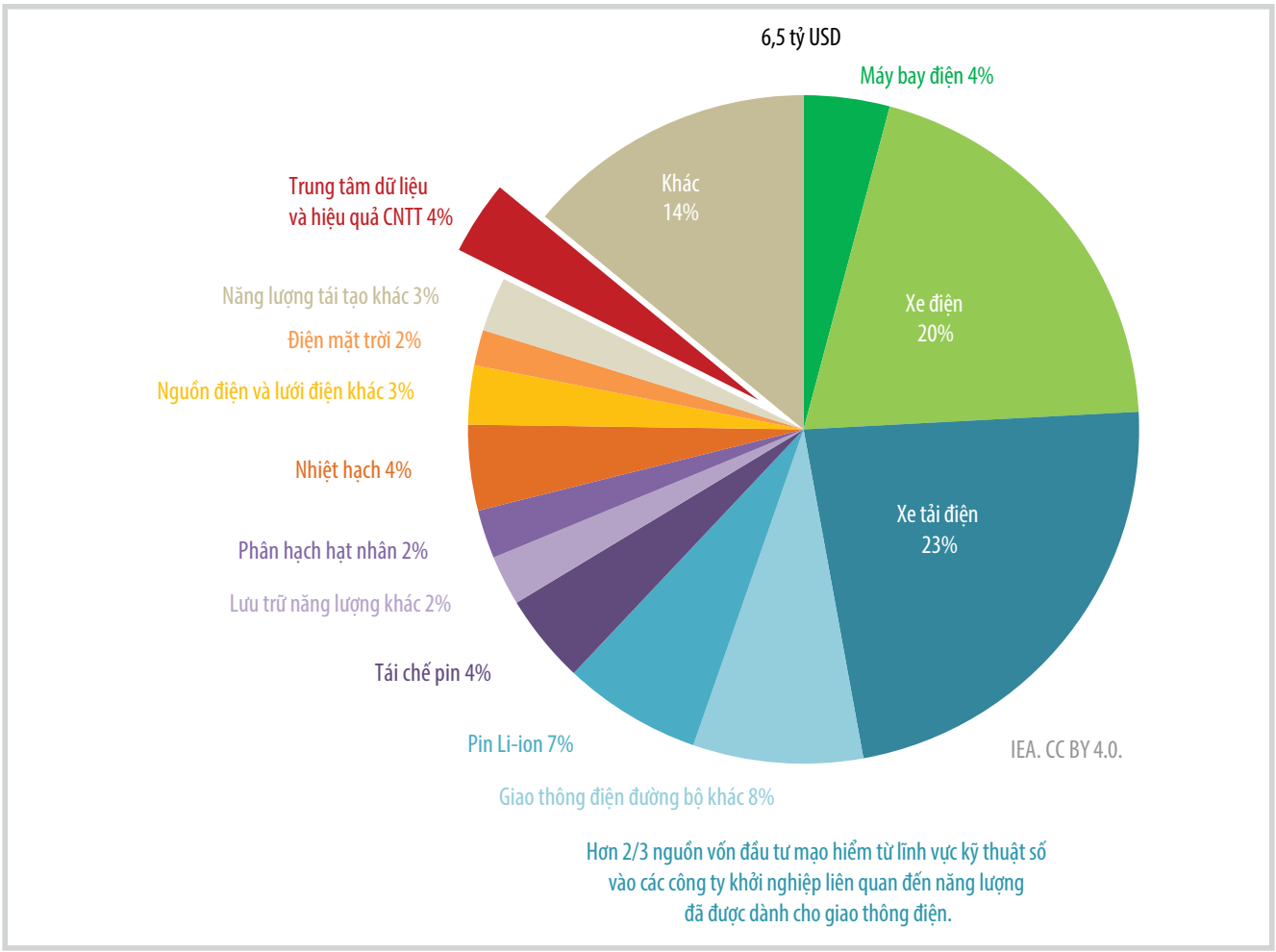
Mức độ không chắc chắn này trong triển vọng gây khó khăn cho việc đầu tư, lập kế hoạch cơ sở hạ tầng và hoạch định chính sách. Điều này càng thách thức hơn do sự khác biệt về thời gian thực hiện giữa cơ sở hạ tầng năng lượng và các trung tâm dữ liệu. Một số điều không chắc chắn trong triển vọng là không thể tránh khỏi, đối với các lĩnh vực như phát điện năng lượng tái tạo cũng gặp khó khăn khi chính sách thay đổi, công nghệ phát triển, các sự kiện kinh tế hoặc địa chính trị bất ngờ...

Để giảm thiểu sự không chắc chắn, cần phân tích rõ hơn mối liên hệ giữa nhu cầu AI và nhu cầu năng lượng. Dữ liệu hạn chế về mức tiêu thụ điện của các loại dịch vụ AI khác nhau, mức độ áp dụng trong thế giới thực và triển vọng tương lai cho nhu cầu dịch vụ AI gây khó khăn trong việc thiết lập mối liên hệ giữa các phát triển trong thế giới thực của AI (ví dụ như việc phát hành DeepSeek-R1) và triển vọng nhu cầu năng lượng. Các dự báo về nhu cầu điện của trung tâm dữ liệu chỉ được kết nối gián tiếp với AI thông qua các biến số cấp hai, như máy chủ hoặc công suất trung tâm dữ liệu đã lắp đặt.

Thiết lập các phương pháp để dự báo tiêu thụ điện từ các trung tâm dữ liệu: Việc dự báo tiêu thụ điện của trung tâm dữ liệu bị cản trở bởi sự thiếu dữ liệu về nhiều điểm. Dữ liệu có sẵn thường chỉ có một phần, còn lại thông qua giấy phép dữ liệu từ một hoặc nhiều nhà cung cấp bên



Hình 7. *Vốn đầu tư mạo hiểm của lĩnh vực kỹ thuật số vào các công ty khởi nghiệp liên quan đến năng lượng giai đoạn 2015 - 2024.*



Hình 8. *Chi tiết nguồn vốn đầu tư mạo hiểm của các công ty công nghệ vào các doanh nghiệp khởi nghiệp liên quan đến năng lượng giai đoạn 2020 - 2024.*

thứ ba, hạn chế quyền truy cập và phổ biến. Dữ liệu hoạt động thương mại, như hiệu quả sử dụng điện, tỷ lệ sử dụng và tỷ lệ điện năng chờ, rất hữu ích, nhưng để mô hình hóa năng lượng, cần thu thập và công khai dữ liệu này cho phép phân tích chính xác hơn.

2.8. Tiềm năng đổi mới của lĩnh vực kỹ thuật số

Lĩnh vực kỹ thuật số là tác nhân quan trọng trong đổi mới ngành năng lượng. Kể từ năm 2015, lĩnh vực này đã chiếm khoảng 5% tổng vốn đầu tư mạo hiểm của các công ty khởi nghiệp liên quan đến năng lượng, có thời điểm lên tới 20% (Hình 8).

Kể từ năm 2020, hơn 2/3 vốn đầu tư mạo hiểm của lĩnh vực kỹ thuật số đã tập trung vào giao thông vận tải điện khí hóa và các lĩnh vực liên quan như pin lithium-ion (Hình 9). Mặc dù vậy, chỉ khoảng 15% vốn đầu tư mạo hiểm của lĩnh vực kỹ thuật số tập trung vào các ứng dụng liên quan đến điện và trong số này, tỷ trọng lớn nhất đã đi vào nhiệt hạch hạt nhân. Khoảng 4% vốn đầu tư mạo hiểm của lĩnh vực kỹ thuật số cho hiệu quả trong các trung tâm dữ liệu và thiết bị công nghệ thông tin và truyền thông (ICT). Các nhà đầu tư vốn mạo hiểm của lĩnh vực kỹ thuật số tập trung vào các lĩnh vực công nghệ có tiềm năng tạo đột phá bằng cách sử dụng các mô hình kinh doanh dựa trên dữ liệu.

6 doanh nghiệp kỹ thuật số có trụ sở tại Mỹ đã đầu tư khoảng 250 tỷ USD cho nghiên cứu và phát triển (R&D) vào năm 2024, tăng từ 50 tỷ USD vào năm 2015; đồng thời tích cực mua lại các doanh nghiệp, một số trong đó liên quan đến năng lượng (ví dụ Google mua lại Waymo và Nest Labs).

Trong trung hạn, lĩnh vực kỹ thuật số sẽ đóng vai trò lớn hơn trong ngành năng lượng khi nhu cầu từ các trung tâm dữ liệu tăng lên. Những thách thức mà hệ thống điện phải đối mặt trong việc đáp ứng sự gia tăng nhanh chóng của các trung tâm dữ liệu nhưng cũng là cơ hội cho đổi mới ở cấp độ hệ thống (ví dụ trung tâm dữ liệu linh hoạt) và ở cấp độ sản phẩm (ví dụ công nghệ lưu trữ nhiệt) để giải quyết thách thức tích hợp này.

3. Kết luận

Nhu cầu điện năng từ các trung tâm dữ liệu AI cần được quản lý bằng các chính sách hiệu suất năng lượng và quy hoạch nguồn cung cẩn trọng. Tuy nhiên, tiềm năng của AI trong việc tối ưu hóa quản lý năng lượng tái tạo, nâng cao hiệu suất năng lượng và thúc đẩy đổi mới công nghệ sạch lớn hơn nhiều.

Để biến AI thành lợi thế chuyển đổi, các nhà hoạch định chính sách cần áp dụng một cách tiếp cận toàn diện: kiểm soát chặt chẽ tiêu thụ năng lượng của hạ tầng AI (Energy for AI), đồng thời tạo điều kiện tối đa cho việc triển khai AI để tối ưu hóa hệ thống năng lượng (AI for Energy). Đối với Việt Nam, việc tận dụng AI để nâng cao khả năng điều độ lưới điện trong bối cảnh bùng nổ năng lượng tái tạo là ưu tiên chiến lược để đảm bảo an ninh năng lượng và đạt được cam kết phát thải ròng bằng 0 vào năm 2050.

Việc cân bằng giữa phát triển hạ tầng AI và chuyển đổi năng lượng đòi hỏi sự can thiệp và định hướng chính sách, đặc biệt quan trọng đối với các quốc gia đang phát triển như Việt Nam. Việt Nam cần đầu tư phát triển hạ tầng số sẵn sàng cho AI trong bối cảnh nhu cầu tiêu thụ năng lượng tăng cao, hạn chế về năng lực lưu trữ và xử lý dữ liệu cũng như các mối quan ngại về an ninh mạng.

Trong đó, Chính phủ cần thiết lập và thực thi các tiêu chuẩn hiệu suất sử dụng điện (PUE) cho các trung tâm dữ liệu mới, đồng thời khuyến khích áp dụng các công nghệ làm mát tiên tiến (liquid cooling) để giảm thiểu điện năng cho quá trình làm mát. Thiết lập quy định yêu cầu các trung tâm dữ liệu quy mô siêu lớn cam kết mua hoặc tự sản xuất năng lượng tái tạo (thông qua DPPA hoặc đầu tư trực tiếp). Chính sách này giúp chuyển đổi nhu cầu điện tăng thêm của AI thành động lực thúc đẩy đầu tư vào năng lượng tái tạo. Ưu tiên đặt trung tâm dữ liệu ở những khu vực có lưới điện mạnh, nguồn cung năng lượng tái tạo dư thừa và điều kiện khí hậu thuận lợi cho việc làm mát tự nhiên.

Hồng Hạnh (dịch)

Tài liệu tham khảo

[1] International Energy Agency (IEA), "Energy and AI", 2025.

DầuKhí
KHOA HỌC, CÔNG NGHỆ VÀ ĐỔI MỚI SÁNG TẠO PETROVIETNAM

AGENTIC AI VÀ TIỀM NĂNG ỨNG DỤNG TRONG NGÀNH CÔNG NGHIỆP NĂNG LƯỢNG

Chỉ trong vài năm, Agentic AI đã chuyển từ khái niệm xa lạ thành một trong những xu hướng công nghệ phát triển nhanh nhất. Với khả năng tự động lập kế hoạch, thực hiện các nhiệm vụ phức tạp gồm nhiều bước, Agentic AI được dự báo sẽ thay đổi hoàn toàn cách làm việc truyền thống.

Hình 1. *So sánh Agentic AI và Generative AI.*

1. Giới thiệu

Agentic AI đang định hình kỷ nguyên mới của các hệ thống thông minh và cách mạng hóa cách thức hoạt động của doanh nghiệp. Bằng cách khai thác sức mạnh của AI tạo sinh (GenAI) và các mô hình ngôn ngữ tiên tiến, Agentic AI đã vượt xa nhiệm vụ đơn thuần là tự động hóa để trở thành chủ thể ra quyết định - có khả năng lập kế hoạch, điều phối và thực thi toàn bộ quy trình làm việc. Deloitte dự báo rằng vào năm 2025, 25% doanh nghiệp sử dụng GenAI sẽ triển khai các dự án thử nghiệm về Agentic AI, con số này sẽ tăng lên 50% vào năm 2027 [1].

Theo Deloitte, tự động hóa quy trình robot (RPA) và Agentic AI đều nhằm mục đích nâng cao hiệu quả, song hoạt động dựa trên nguyên tắc cơ bản khác nhau. RPA tự động hóa các tác vụ rời rạc, dựa trên quy tắc và yêu cầu sự can thiệp của con người đối với các trường hợp ngoại lệ, ghi nhật ký hoặc ra quyết định. RPA hiệu quả đối với các quy trình có cấu trúc, lặp đi lặp lại nhưng thiếu khả năng thích ứng và tự chủ. Trong khi đó, Agentic AI nhận biết bối cảnh và có khả năng đưa ra quyết định, thực thi các tác vụ và xác thực, điều chỉnh cũng như tối ưu hóa các hoạt động theo thời gian thực, chỉ yêu cầu sự tham gia của con người khi thực sự cần thiết. Điều này khiến Agentic AI trở nên linh hoạt, thích ứng và hiệu quả hơn so với tự động hóa truyền thống và GenAI (Hình 1, 2).

Agentic AI giúp các doanh nghiệp hoạt động hiệu quả hơn bằng cách tự động hóa các chức năng, bao gồm các hoạt động tiếp cận và tương tác với khách hàng, đổi

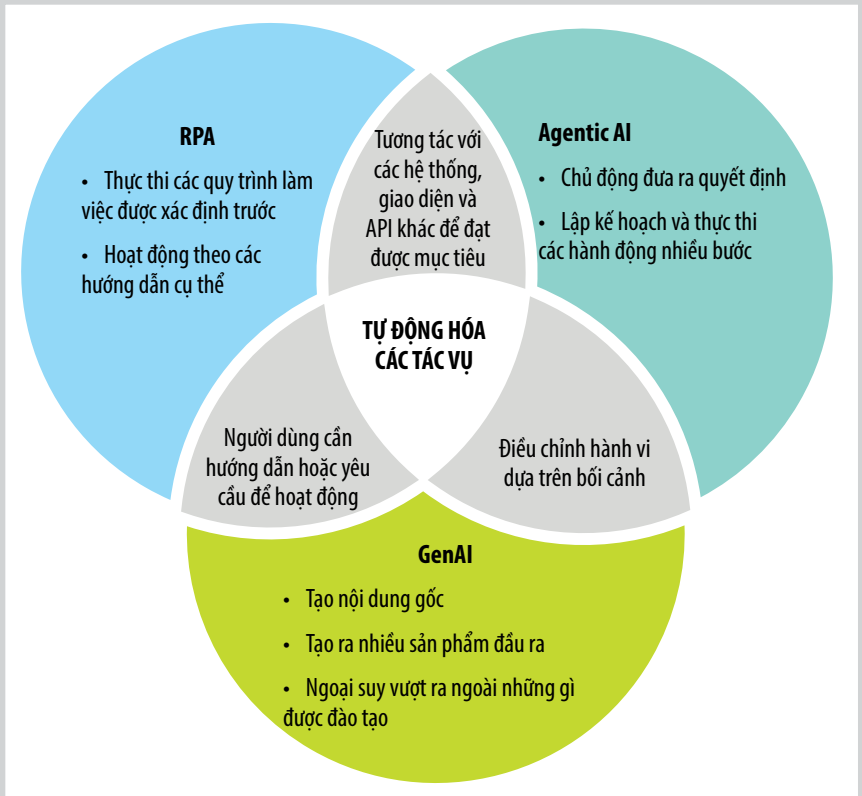
Bảng 1. *Ưu thế vượt trội của Agentic AI so với các hệ thống AI truyền thống [2].*

TT	Ưu thế của Agentic AI so với các hệ thống AI truyền thống	Ghi chú
1	Xử lý chuỗi tác vụ không thể dự đoán	Trước đây, để tạo phần mềm tự động, lập trình viên phải viết từng bước tỉ mỉ theo quy tắc, thậm chí có bước cần phải xử lý thủ công. Hiện nay, các mô hình ngôn ngữ lớn (LLM) trong Agentic AI có thể phản ứng chính xác với những tình huống chưa từng gặp, xử lý linh hoạt các tác vụ phức tạp.
2	Sử dụng công cụ kỹ thuật số như con người	Thay vì cần mã tùy chỉnh để kết nối với từng hệ thống mới, các tác nhân AI có thể sử dụng các công cụ giống như con người như: duyệt web, đọc trang, điền biểu mẫu, thông qua khả năng "đọc hiểu" của LLM.
3	Nhận lệnh bằng ngôn ngữ tự nhiên	Agentic AI có thể giao tiếp với con người như đưa hướng dẫn, góp ý, huấn luyện… bằng ngôn ngữ hàng ngày, không cần biết lập trình.
4	Tạo kế hoạch làm việc có thể hiểu và điều chỉnh	Các tác nhân AI tạo ra kế hoạch công việc bằng ngôn ngữ con người có thể đọc, theo dõi, đánh giá và điều chỉnh hành động một cách minh bạch.

mới sản phẩm và các hoạt động liên quan đến lĩnh vực tài chính. Không chỉ đơn thuần là tự động hóa các tác vụ, Agentic AI được trao quyền tự ra quyết định và thích ứng linh hoạt.

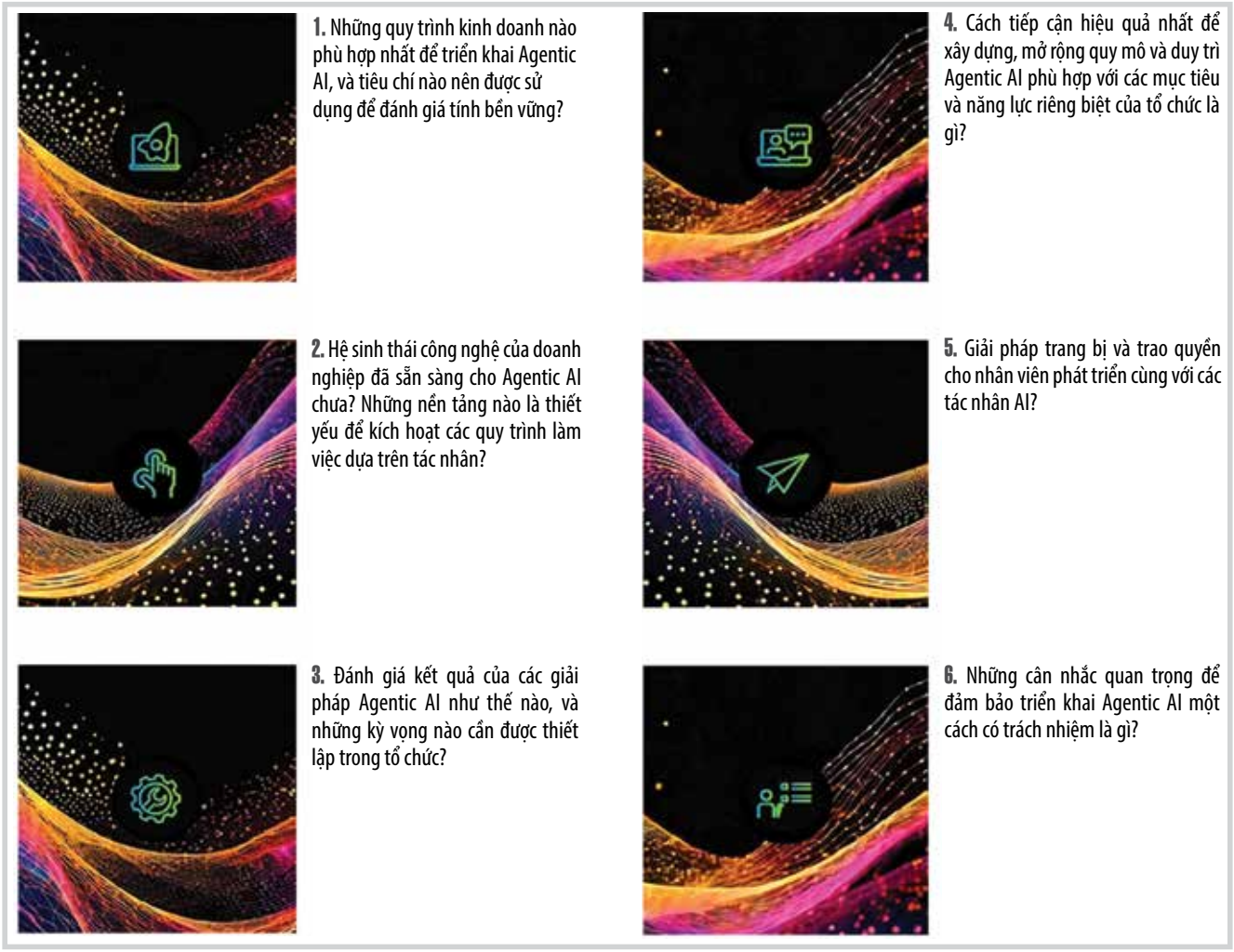
Trong Báo cáo xu hướng công nghệ năm 2025 [2], McKinsey đề cập đến Agentic AI đầu tiên, và cho rằng công nghệ này đang thu hút sự chú ý khi các tổ chức tìm kiếm những cách thức mới để tự động hóa quy trình làm việc và giao nhiệm vụ cho các "đồng nghiệp ảo", thay vì chỉ trò chuyện với chatbot. Điểm khác biệt của Agentic AI là khả năng hành động trong thế giới thực, thường sử dụng các công cụ số, thay vì chỉ đơn thuần cung cấp kết quả đầu ra. Các hệ thống này được xây dựng dựa trên các mô hình nền tảng AI và có thể tự động lập kế hoạch cũng như thực thi các nhiệm vụ gồm nhiều bước. Agentic AI có thể trả lời câu hỏi về sản phẩm, tự động xử lý đơn hàng, quản lý việc trả hàng bằng cách kết nối với hệ thống logistics...

Theo thống kê của McKinsey, số lượng tin tuyển dụng liên quan đến Agentic AI đã tăng 985% trong giai đoạn 2023 - 2024, đầu tư vào lĩnh vực này năm 2024 đạt khoảng 1,1 tỷ USD. Sự quan tâm đối với Agentic AI, được đo lường qua tin tức và lượt tìm kiếm, đang tăng trưởng nhanh hơn bất kỳ xu hướng công nghệ nào khác. Các bằng sáng chế liên quan đến Agentic AI cũng đang tăng nhanh, phản ánh sự phát triển và làn sóng đầu tư ngày càng tăng trong lĩnh vực mới nổi này. Bảng 1 thể hiện các ưu thế vượt trội của Agentic AI so với các hệ thống AI truyền thống [2].



Hình 2. *So sánh RPA, GenAI và Agentic AI [3].*

Một số ví dụ về nền tảng tác nhân đa năng được hỗ trợ bởi AI có thể kể đến OpenAI Operator, (tự động đặt vé máy bay, đặt chỗ nhà hàng và đặt hàng trực tuyến); Manus AI hay việc Google phát hành Gemini 2.5 Flash cho Gemini API trong cả Google AI Studio và Vertex AI. Agentic AI được xây dựng liên quan đến suy luận nhiều bước như



Hình 3. *Sáu vấn đề các doanh nghiệp cần cân nhắc khi triển khai Agentic AI [3]*

McKinsey's QuantumBlack Labs đã triển khai quy trình làm việc tác nhân để tự động hóa việc soạn thảo bản ghi nhớ tín dụng cho ngân hàng, giúp năng suất của các nhà phân tích tín dụng tăng lên đến 60%. Agentic AI được xây dựng cho các giải pháp kinh doanh cụ thể như Darktrace (tự động phát hiện và phản ứng với các mối đe dọa an ninh mạng trong thời gian thực mà không cần sự can thiệp của con người)...

Theo đánh giá của McKinsey [2], các phát triển mới nhất về Agentic AI gồm 6 xu hướng chính: (i) Xây dựng nền tảng Agentic AI đa năng có thể tương tác bằng ngôn ngữ tự nhiên và thực hiện nhiều nhiệm vụ khác nhau; (ii) Cải thiện khả năng suy luận nhiều bước thông qua mô hình đa tác nhân, trong đó tác nhân "quản lý" xây dựng kế hoạch làm việc và giao nhiệm vụ cho các tác nhân chuyên biệt, quản lý sản phẩm đầu ra chính xác hơn và nhận biết ngữ cảnh tốt hơn; (iii) Phát triển các Agentic AI chuyên biệt để giải quyết các vấn đề kinh doanh cụ thể, có giá trị cao, đo lường được trong các chỉ số kinh doanh, đặc biệt là tối ưu hóa bán hàng và tự động hóa hỗ trợ khách hàng; (iv) Phát triển các Agentic AI nghiên cứu chuyên sâu có thể tự động tìm kiếm, đánh giá hàng trăm nguồn và tổng hợp báo cáo toàn diện; (v) Phát triển các Agentic AI có thể giao tiếp với nhau (AI-với-AI) bằng ngôn ngữ riêng; và (vi) Giải quyết các vấn đề về niềm tin, quản trị và trách nhiệm pháp lý khi triển khai Agentic AI, đảm bảo độ tin cậy và trách nhiệm giải trình.

2. Ứng dụng Agentic AI trong ngành công nghiệp năng lượng

Ngành công nghiệp năng lượng đang chứng kiến sự chuyển đổi mạnh mẽ từ AI truyền thống sang Agentic AI với khả năng tự chủ lập kế hoạch, ra quyết định và thực thi mà không cần can thiệp liên tục từ con người. Với đặc thù phức tạp, vận hành 24/7 và nhu cầu ra quyết định theo thời gian thực, lĩnh vực công nghiệp năng lượng đang trở thành thị trường tiềm năng của Agentic AI. Theo Mordor Intelligence, thị trường Agentic AI trong lĩnh vực năng lượng

và tiện ích được định giá ở mức 0,64 tỷ USD năm 2025 và dự báo đạt 3,14 tỷ USD vào năm 2030 [4].

- Tối ưu hóa chuỗi giá trị năng lượng

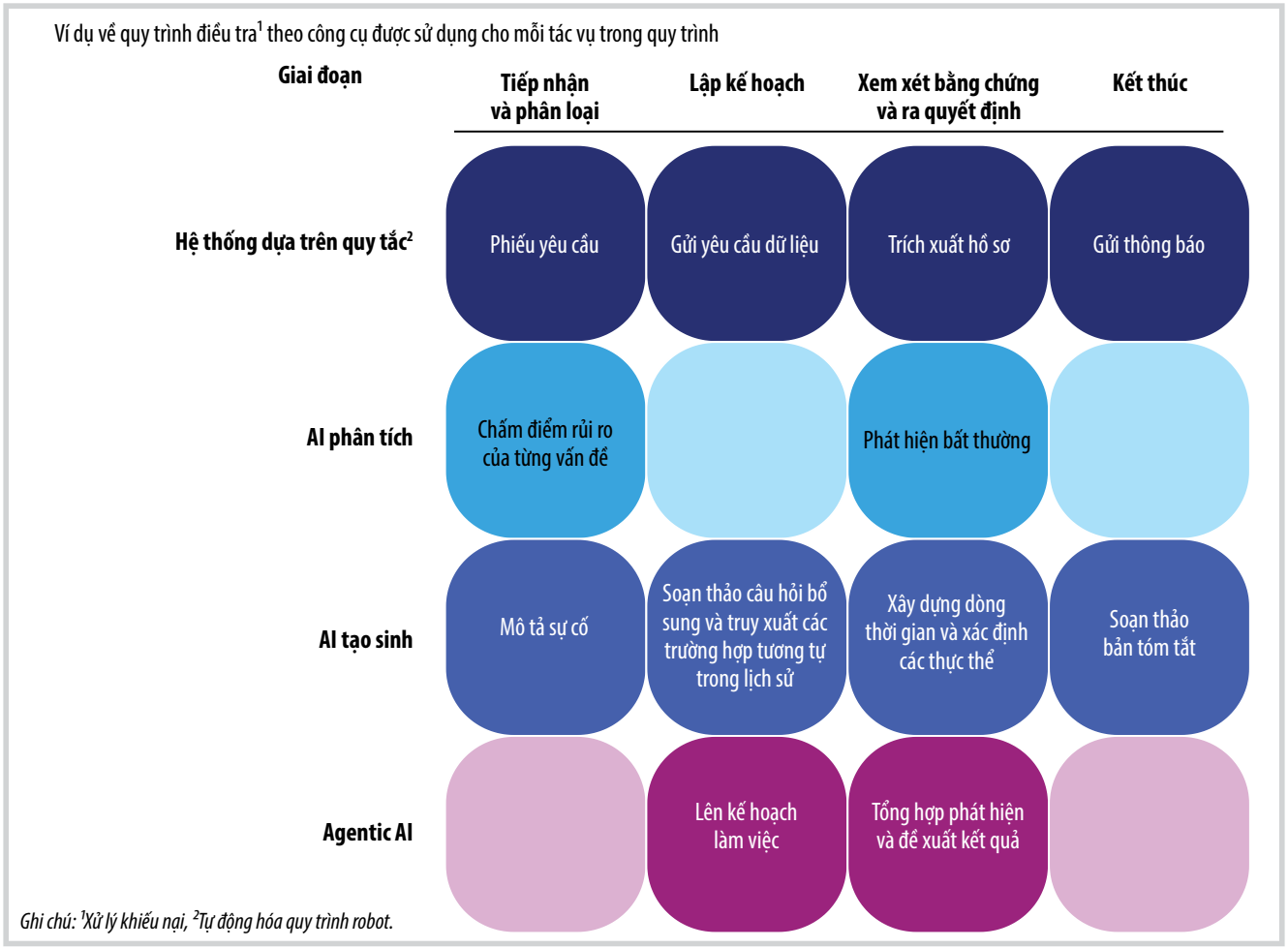
Đầu năm 2025, ADNOC và AIQ công bố thử nghiệm ENERGYai, giải pháp Agentic AI đầu tiên trên thế giới được thiết kế riêng cho lĩnh vực năng lượng. ENERGYai tích hợp mô hình ngôn ngữ lớn (LLM) với 70 tỷ tham số cùng với hơn 50 năm kiến thức và dữ liệu độc quyền của ADNOC. Kết quả cho thấy ENERGYai giúp nâng cao hiệu quả và độ chính xác cho các nhiệm vụ quan trọng từ phân tích địa chấn đến mô hình hóa địa chất và giám sát quy trình thời gian thực, rút ngắn thời gian cho các quy trình kinh doanh thiết yếu từ hàng tháng xuống còn vài ngày, đồng thời giảm thiểu chi phí và lượng khí thải [5, 6].

SLB và AIQ mới đây cho biết sẽ hợp tác phát triển giải pháp Agentic AI ENERGYai, thiết kế và triển khai các quy trình làm việc Agentic AI mới cho các hoạt động của ADNOC, bao gồm địa chất, thăm dò địa chấn và mô hình hóa mỏ, được hỗ trợ bởi nền tảng dữ liệu và AI Lumi™ của SLB, cùng các công nghệ số khác. Được thiết kế để giúp tăng năng suất và hiệu quả, nền tảng AI Lumi™ nâng cao khả năng truy cập dữ liệu, tối ưu hóa quy trình làm việc và mở rộng quy mô triển khai các giải pháp AI [7].

Agentic AI đang chuyển đổi cách sản xuất, quản lý, phân phối và tiêu thụ năng lượng, mặc dù vẫn còn những thách thức về cơ sở hạ tầng, an ninh mạng và khung pháp lý [8].

- Quản lý lưới điện thông minh

Agentic AI được ứng dụng trong điều hành lưới điện thông minh (smart grid) tự động và hiệu quả. Khác với các hệ thống tự động hóa truyền thống. Agentic AI có khả năng liên tục giám sát lưới điện, cân bằng tải điện tự động bằng cách phân tích dữ liệu tiêu thụ theo thời gian thực, dự báo nhu cầu và điều chỉnh tỷ lệ phân phối điện năng giữa các khu vực; phát hiện và xử lý sự cố ngay lập tức khi có biến động bất thường, tự động chuyển sang nguồn điện dự phòng mà không cần con người can thiệp [9].



Hình 4. *Quy trình làm việc phức tạp nên kết hợp công cụ tốt nhất cho từng tác vụ [10].*

Bảng 2. *Mô hình cộng tác bền vững giữa con người và Agentic AI*

Con người	Agentic AI	Kết quả
Phán đoán chiến lược	Xử lý dữ liệu quy mô lớn	Quyết định nhanh và chính xác hơn
Xử lý ngoại lệ	Tự động hóa công việc lặp lại	Tập trung vào giá trị cao
Trách nhiệm pháp lý	Phân tích và đề xuất	Tuân thủ và hiệu quả
Sáng tạo và đổi mới	Tối ưu hóa quy trình	Năng suất tăng vọt

Agentic AI có thể tự động tối ưu hóa năng lượng dựa trên sự thay đổi các yếu tố như thời tiết, giá điện biến động, công suất của các nhà máy và mức độ quá tải của lưới điện...

- Tối ưu hóa nguồn năng lượng tái tạo

Agentic AI có thể phân tích dữ liệu thời tiết, lịch sử sản xuất, tình trạng thiết bị và hàng nghìn biến số khác để dự báo chính xác sản lượng điện năng lượng tái tạo; tự động quyết định thời điểm sử dụng năng lượng mặt trời, khi nào chuyển sang điện gió, hay khi nào cần sử dụng nguồn điện truyền thống; quản lý hệ thống lưu trữ pin, quyết định thời điểm tối ưu để sạc pin và xả pin dựa trên nhu cầu và tình trạng của pin.

- Bảo trì dự đoán cho cơ sở hạ tầng năng lượng

Trong ngành công nghiệp năng lượng, việc bảo trì thiết bị đúng thời điểm có thể tiết kiệm hàng triệu USD và ngăn ngừa nguy cơ xảy ra các thảm họa môi trường. Agentic AI có thể giám sát liên tục hàng nghìn cảm biến, phát hiện các dấu hiệu cảnh báo sớm trước khi xảy ra hỏng hóc, tự động lập kế hoạch bảo trì bằng cách cân nhắc mức độ khẩn cấp, tình trạng tồn kho của thiết bị, lịch sản xuất và sự sẵn sàng của kỹ thuật viên.

- Tối ưu hóa tiêu thụ năng lượng

Các nhà máy sản xuất, khu công nghiệp và tòa nhà có thể sử dụng Agentic AI để giảm chi phí năng lượng: Điều chỉnh tự động hệ thống HVAC (sưởi ấm, thông gió, điều hòa không khí) dựa trên số lượng người, thời tiết, giá điện theo giờ và lịch làm việc; quản lý máy móc sản xuất để vận hành vào khung giờ điện thấp mà vẫn đảm bảo tiến độ; phân tích và loại bỏ lãng phí năng lượng bằng cách xác định thiết bị hoạt động không hiệu quả hoặc tiêu thụ điện bất thường.

3. Bài học kinh nghiệm khi triển khai Agentic AI

Mặc dù có nhiều lợi thế, Agentic AI vẫn đối mặt với nhiều thách thức về trách nhiệm pháp lý. Khi giao cho AI tự động thực hiện giao dịch tài chính hoặc tương tác qua các nền tảng kỹ thuật số, sẽ nảy sinh vấn đề trách nhiệm pháp lý và vấn đề đặt ra là ai sẽ chịu trách nhiệm nếu xảy ra sai sót? Các khung pháp lý hiện tại vẫn chưa theo kịp tốc độ phát triển của công nghệ. Ví dụ khi tác nhân AI quản lý lưới điện đưa ra quyết định sai dẫn đến mất điện diện rộng, ai sẽ chịu trách nhiệm – đơn vị cung cấp AI, công ty điện lực, hay kỹ sư thiết kế hệ thống?

Các tác nhân AI có thể đưa ra quyết định sai, thực hiện hành động không mong muốn, hoặc bị tấn công bởi mã độc. Ngành công nghiệp năng lượng yêu cầu độ tin cậy cực cao (thường 99,99% trở lên), rủi ro này cần được quản lý nghiêm ngặt. Với vai trò quan trọng của cơ sở hạ tầng năng lượng, bất kỳ lỗ hổng bảo mật nào trong hệ thống Agentic AI đều có thể trở thành mục tiêu tấn công nghiêm trọng.

Hiện tại, Agentic AI mới ở giai đoạn thử nghiệm. Từ hơn 50 dự án xây dựng Agentic AI, McKinsey rút ra 6 bài học kinh nghiệm khi triển khai Agentic AI [10].

- Thay đổi quy trình làm việc

Để khai thác giá trị của Agentic AI, cần thay đổi quy trình làm việc. Một số tổ chức thường chỉ tập trung vào tác nhân hoặc công cụ của Agentic AI, nhưng thực tế không cải thiện quy trình làm việc, dẫn đến giá trị không đạt như mong đợi.

Để thiết kế lại quy trình làm việc, bước đầu tiên là lập bản đồ quy trình và xác định các điểm chính của người dùng, giảm công việc không cần thiết và cho phép Agentic AI và con người cộng tác thông qua các vòng lặp học và phản hồi, để hệ thống tự củng cố. Các Agentic AI càng được sử dụng thường xuyên, càng trở nên thông minh hơn và phù hợp hơn.

Tập trung vào quy trình làm việc thay vì Agentic AI cho phép các nhóm triển khai đúng công nghệ vào đúng thời điểm, điều này đặc biệt quan trọng khi tái thiết kế các quy trình làm việc phức tạp, nhiều bước (Hình 4) [10].

Các tổ chức có thể thiết kế lại các loại quy trình làm việc này bằng cách kết hợp có mục tiêu của hệ thống dựa trên quy tắc, AI phân tích, AI tạo sinh và Agentic AI, được hỗ trợ bởi một khung điều phối chung (chẳng hạn như các framework mã nguồn mở như AutoGen, CrewAI và LangGraph).

- Cân nhắc kỹ trước khi lựa chọn phát triển Agentic AI

Trước khi vội vàng áp dụng Agentic AI, cần đánh giá yêu cầu của nhiệm vụ, làm rõ: Quy trình nên được chuẩn hóa ở mức độ nào, Quy trình cần xử lý bao nhiêu biến thể và xác định các phần công việc nào các tác nhân phù hợp nhất để làm.

Các quy trình làm việc có độ biến đổi thấp, chuẩn hóa cao, như đấu thầu, có xu hướng được quản lý chặt chẽ và tuân theo logic có thể dự đoán. Trong trường hợp này, các tác nhân dựa trên LLM không xác định có thể thêm nhiều độ phức tạp và không chắc chắn hơn là giá trị. Ngược lại, các quy trình làm việc có độ biến đổi cao, chuẩn hóa thấp có thể được hưởng lợi đáng kể từ các Agentic AI.

- Đầu tư vào đánh giá và xây dựng niềm tin với người dùng

Các tổ chức nên đầu tư vào phát triển Agentic AI, giống như phát triển nhân viên, cần mô tả công việc rõ ràng, đào tạo ban đầu, phản hồi liên tục và cải thiện thường xuyên, phát triển đánh giá.

- Giám sát và cải thiện hiệu suất liên tục

Khi làm việc với vài Agentic AI, việc xem xét và phát hiện lỗi có thể khá đơn giản. Nhưng khi tổ chức mở rộng quy mô lên hàng trăm Agentic AI, nhiệm vụ thách thức là tìm ra chính xác điều gì đã xảy ra sai sót.

Hiệu suất của Agentic AI cần được xác minh tại mỗi bước của quy trình làm việc. Xây dựng giám sát và đánh giá vào quy trình làm việc có thể cho phép các nhóm phát hiện sai sót sớm, tinh chỉnh logic và liên tục cải thiện hiệu suất, ngay cả sau khi các Agentic AI được triển khai.

- Phát triển Agentic AI có thể dễ dàng được tái sử dụng trên các quy trình làm việc khác nhau

Các tổ chức thường tạo ra Agentic AI duy nhất cho mỗi nhiệm vụ được xác định. Điều này có thể dẫn đến sự dư thừa và lãng phí đáng kể vì cùng 1 Agentic AI thường có thể hoàn thành các nhiệm vụ khác nhau chia sẻ nhiều hành động giống nhau (chẳng hạn như nhập, trích xuất, tìm kiếm và phân tích).

Quyết định đầu tư vào việc xây dựng các tác nhân có thể tái sử dụng (so với 1 tác nhân thực thi 1 nhiệm vụ cụ thể) tương tự như việc xây dựng nhanh nhưng không khóa các lựa chọn hạn chế khả năng trong tương lai. Kinh nghiệm của McKinsey cho thấy điều này giúp giảm từ 30 - 50% công việc không cần thiết.

- Con người giữ vị trí thiết yếu, nhưng vai trò và số lượng sẽ thay đổi

Khi Agentic AI tiếp tục mở rộng, con người cần tập trung giám sát độ chính xác của mô hình, đảm bảo tuân thủ, sử dụng phán đoán và xử lý các trường hợp ngoại lệ. Tuy nhiên, số lượng người làm việc trong một quy trình làm việc cụ thể có khả năng sẽ thay đổi và thường sẽ thấp hơn sau khi quy trình được chuyển đổi bằng cách sử dụng các Agentic AI.

Một bài học quan trọng khác từ kinh nghiệm của McKinsey là các tổ chức cần thiết kế lại công việc để con người và Agentic AI có thể cộng tác cùng nhau (Bảng 2). Nếu không làm được việc này, ngay cả các chương trình Agentic AI tiên tiến nhất cũng có nguy cơ thất bại, tích lũy lỗi và người dùng từ chối sử dụng.

Các điểm can thiệp của con người: (i) Xác minh dữ liệu quan trọng; (ii) Quyết định chiến lược (con người không chỉ xem xét mà còn điều chỉnh khuyến nghị do Agentic AI đề xuất), (iii) Xử lý các trường hợp đặc biệt; (iv) Chịu trách nhiệm pháp lý cuối cùng. Sự tập trung vào trải nghiệm người dùng này không chỉ cải thiện hiệu quả mà còn tạo ra mô hình cộng tác bền vững giữa con người và Agentic AI.

Huy Hoàng (tổng hợp)

Tài liệu tham khảo

[1] Deloitte, "Agentic AI: Autonomous generative AI agents", 2025.

[2] McKinsey, "Technology trends outlook 2025", 2025.

[3] Deloitte, "The business imperative for Agentic AI", 2025.

[4] Mordor Intelligence, "Agentic AI in energy and utilities market size & share analysis - Growth trends and forecast (2025 - 2030)".

[5] ADNOC, "ADNOC and AIQ successfully complete trial phase of Agentic AI solution", 2025.

[6] ADNOC, "ENERGYai brings large-scale agentic AI to the energy sector", 2025.

[7] SLB, "AIQ and SLB to advance development and deployment of Agentic AI solution across ADNOC's subsurface operations", 2025.

[8] Jagreet Kaur, "Agentic AI in energy sector: Pioneering autonomous energy intelligence", 2025.

[9] Aresh Mishra, "Agentic AI in energy and utilities for optimisation with examples", 2025.

[10] Lareina Yee, Michael Chui, and Roger Roberts, "One year of agentic AI: Six lessons from the people doing the work", 2025.