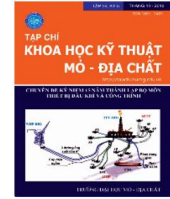




## Tạp chí Khoa học Kỹ thuật Mỏ - Địa chất

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



# Nghiên cứu ứng dụng mạng neural nhân tạo để dự đoán tốc độ ăn mòn carbon dioxide (CO<sub>2</sub>) trong đường ống dẫn dầu khí

Nguyễn Phùng Hưng \*, Lê Đức Vinh, Triệu Hùng Trường

Khoa Dầu khí, Trường Đại học Mỏ - Địa chất, Việt Nam

### THÔNG TIN BÀI BÁO

Quá trình:  
 Nhận bài 08/11/2018  
 Chấp nhận 05/01/2019  
 Đăng online 28/02/2019

Từ khóa:  
 Ăn mòn CO<sub>2</sub>  
 Đường ống  
 Mạng neural nhân tạo  
 Tốc độ ăn mòn  
 Norsok M-506

### TÓM TẮT

Ăn mòn trong đường ống do tác nhân Carbon Dioxide (CO<sub>2</sub>) là vấn đề đáng quan tâm trong công nghiệp dầu khí. Trong nghiên cứu này, nhóm tác giả đề xuất phương pháp tiếp cận mới để dự đoán tốc độ ăn mòn do Carbon Dioxide (CO<sub>2</sub>) trong đường ống, ứng dụng phương pháp máy học, cụ thể là mạng Neural nhân tạo (ANN). Sau khi tiến hành thu thập, chọn lọc đặc trưng, tiền xử lý dữ liệu, tập dữ liệu phục vụ cho nghiên cứu bao gồm 40 điểm dữ liệu với 9 đặc trưng (features). Mô hình mạng ANN đã được xây dựng với 2 lớp ẩn có kích thước lần lượt là 18 node và 9 node; với hàm kích hoạt lần lượt là ReLU và Sigmoid. Nhóm tác giả sử dụng các thuật toán Early Stopping và RMSprop để tăng độ chính xác của mô hình ANN, giảm ảnh hưởng của bộ dữ liệu nhỏ. Mô hình dự đoán theo tiêu chuẩn Norsok M-506 cũng được áp dụng để so sánh hiệu quả với mô hình mạng neural nhân tạo. Các chỉ tiêu bao gồm căn bậc hai sai số bình phương trung bình (RMSE), sai số tuyệt đối trung bình (MAE), hệ số xác định (R<sup>2</sup>). Kết quả nghiên cứu cho thấy dự đoán của mô hình mạng neural nhân tạo đạt các chỉ tiêu R<sup>2</sup> = 0,938, RMSE = 0,014, MAE = 0,011 tốt hơn nhiều so với mô hình thực nghiệm theo tiêu chuẩn Norsok M-506.

© 2019 Trường Đại học Mỏ - Địa chất. Tất cả các quyền được bảo đảm.

## 1. Mở đầu

Ăn mòn kim loại trong và ngoài đường ống dầu khí là một trong những vấn đề lớn cần giải quyết của ngành công nghiệp dầu khí. Ước tính rằng hư hỏng do ăn mòn chiếm tới 30-33% tổng số hư hỏng ghi nhận được trên toàn tuyến ống trên thế giới (Bersani et al., 2010), và là nguyên nhân gây ra rò rỉ hoặc vỡ ống. Trên thế giới cũng như trong nước đã có nhiều nghiên cứu về ăn mòn

trong và ngoài đường ống. Trong điều kiện chất lưu vận chuyển chỉ là các hydro cacbon, đường ống hầu như không bị ăn mòn trong. Sự xuất hiện của các tác nhân hóa học: nước, hơi nước ngưng tụ, các khí hòa tan (H<sub>2</sub>S, CO<sub>2</sub>, O<sub>2</sub> ...), chất rắn hòa tan, phân tán, muối NaCl, MgCl<sub>2</sub>... kết hợp với các tác nhân vật lý: nhiệt độ, áp suất, chế độ dòng chảy ... gây nên các phản ứng điện hóa dẫn đến ăn mòn và phá hủy đường ống (Pierre, 2002). Trong đó ăn mòn trong do carbon dioxide (CO<sub>2</sub>) là dạng ăn mòn thường gặp nhất.

Quá trình ăn mòn trong đường ống được đánh giá thông qua chỉ số tốc độ ăn mòn

\*Tác giả liên hệ

E - mail: [nph5680@gmail.com](mailto:nph5680@gmail.com)

(mm/năm). Việc dự đoán chính xác tốc độ ăn mòn trong đường ống giúp giải quyết hàng loạt các vấn đề về kỹ thuật dầu khí: từ thiết kế, xây dựng tuyến ống; tối ưu chi phí vận hành: chống ăn mòn, lên kế hoạch và thực hiện kiểm tra, bảo dưỡng, sửa chữa, thay thế một phần; ngăn chặn rò rỉ dầu và khí gây tổn hại tới môi trường.

Hiện nay, phương pháp phổ biến nhất để đánh giá ăn mòn trong và các khuyết tật cho toàn bộ đường ống dẫn dầu khí là dùng thiết bị kiểm tra bên trong đường ống (ILI: in-line inspection), dựa trên kỹ thuật từ thông rò rỉ (MFL: Magnetic Flux Leakage) hoặc kỹ thuật sóng siêu âm nén giãn (Ultrasonic Compression Wave) (Kexi Liao et al., 2012; NACE SP0102, 2010). Các tiêu chuẩn (NACE SP0110, 2010; NACE SP0116, 2016; NACE SP0206-SG, 2016) của Hiệp hội các quốc gia về chống ăn mòn (NACE) đề xuất cách tiếp cận để tính toán tốc độ ăn mòn trên toàn bộ tuyến ống dựa trên kết quả khảo sát bên trong của một đoạn đường ống (ILI) căn cứ vào các mô hình dự đoán tốc độ ăn mòn đã được thế giới công nhận như: NORSOK model (Statoil, Saga, Hydro), de Waard model (Shell), Cassandra (BP), HYDROCOR (Shell), CORPLUS (Total), Tulsa model (Đại học Tulsa), ULL model (Đại học Louisiana Lafayette)...

Các mô hình trên thu được nhờ các phương pháp tính toán lý thuyết, thí nghiệm, bán thực nghiệm và dữ liệu thực tế; là kết quả nghiên cứu của các tổ chức, công ty dầu khí, các trung tâm nghiên cứu lớn trên thế giới. Tuy nhiên, kết quả dự đoán của các mô hình đôi khi không phù hợp với giá trị tốc độ ăn mòn thực tế và rất khác nhau, do các mô hình có đối tượng nghiên cứu khác nhau, được tính toán trên bộ tham số khác nhau. Rofth Nyborg nhận định không mô hình thực nghiệm nào có độ tin tưởng  $\pm 50\%$  khi áp dụng trên điều kiện dữ liệu rộng và chứng minh bằng việc so sánh mô hình Norsok M-506 với mô hình de Waard trên đoạn ống 10km, với nhiệt độ đầu vào 90°C và đầu ra 20°C (Rofth Nyborg, 2005, 2009).

Do không thể có một hàm số chính xác hoặc phù hợp để liên kết những thông số đầu vào tới những thông số đầu ra, vì vậy phương pháp học máy xuất hiện như ứng viên phù hợp để nhận được những mối liên hệ hàm số giữa những thông số đó và kết quả mong muốn (Mohaghegh, 1995). Trong những năm gần đây, nhiều phương pháp máy học máy và thống kê đã được phát triển để dự đoán tốc độ ăn mòn trong do CO<sub>2</sub>: mạng neural

nhân tạo, thuật toán tiến hóa, mạng neural mờ, logic mờ, phương pháp Monte Carlo (Mohamed Layouni, 2014; Bassam, 2009). Obaseki M. xây dựng mạng neural nhân tạo cho bộ dữ liệu 60 điểm với 13 đặc trưng đầu vào để dự đoán tốc độ ăn mòn trên đường ống vùng Niger Delta ở Nigeria và đạt được kết quả tốt so với kết quả phần mềm mô phỏng được sử dụng tại 6 công ty dầu khí Niegera (Obaseki et al., 2017). Kexi Liao sử dụng các thuật toán tối ưu để xây dựng 3 mô hình neural nhân tạo 1 lớp ẩn 14 nodes cho tập dữ liệu 116 điểm dữ liệu, với 7 đặc trưng dữ liệu đầu vào có hệ số tương quan cao hơn 0,7 để dự đoán tốc độ ăn mòn trong đường ống dẫn khí có hơi nước. Kết quả chỉ ra rằng mô hình mạng neural networks cho kết quả dự đoán tốt hơn mô hình de Waard, và mô hình Top - of - Line (Kexi Liao et al., 2012). Tuy nhiên cả hai công trình nêu trên đều đánh giá hiệu quả mô hình trên giá trị sai số tuyệt đối và giá trị sai số tương đối, theo nhóm nghiên cứu là 2 chỉ tiêu đó chưa đủ để đánh giá tính hiệu quả của mô hình dự đoán. Mazura Mat Din sử dụng trực tiếp dữ liệu đầu vào là các kết quả của thiết bị kiểm tra bên trong đường ống (In line Inspection): độ sâu, chiều dài, chiều rộng, hướng và hành trình, để xây dựng mạng neural nhân tạo với ba lớp ẩn. Kết quả thu được rất tốt với các chỉ tiêu hệ số xác định  $R^2 = 0,9988$ , sai số tuyệt đối trung bình MAE = 0,0016, căn bậc hai sai số bình phương trung bình RMSE = 0,003 (Mazura Mat Din et al., 2015). Giulia De Masi sử dụng tập dữ liệu 10 đặc trưng đầu vào: 8 đặc trưng cho dòng chảy và thông số vận chuyển, và 2 đặc trưng là kết quả dự đoán tốc độ ăn mòn bởi mô hình Norsok và de Waard. Một mô hình neural nhân tạo khớp (Fitting Neural Network) một lớp ẩn 2 nodes được xây dựng, kết quả thu được có các giá trị chỉ tiêu chưa thực sự cao: hệ số tương quan  $R = 0,66$ , căn bậc hai của sai số phần trăm trung bình RMSPE = 52, trung bình sai số phần trăm tuyệt đối MAPE = 31 (Giulia De Masi et al., 2014). Hiện nay, chưa có một công trình nghiên cứu nào ứng dụng phương pháp học máy để dự đoán tốc độ ăn mòn được công bố tại Việt Nam.

Trong nghiên cứu này, nhóm tác giả trình bày phương pháp phát triển một mô hình mạng neural nhân tạo (Artificial Neural Network) để dự đoán tốc độ ăn mòn trên đường ống dẫn dầu và khí. Phương pháp thực nghiệm theo tiêu chuẩn Norsok M-506 được áp dụng để so sánh hiệu quả

với mô hình mạng neural đã phát triển.

## 2. Phương pháp nghiên cứu

Trong nghiên cứu này, phương pháp máy học với thuật toán hồi quy phi tuyến tính thông qua mạng neural nhân tạo được sử dụng để dự đoán tốc độ ăn mòn. Các bước tiến hành bao gồm: chuẩn bị dữ liệu đầu vào; chia tập dữ liệu ra thành các thành phần với tỷ lệ: bộ dữ liệu huấn luyện, bộ dữ liệu xác thực, bộ dữ liệu kiểm tra; xây dựng mạng neural nhân tạo dự đoán tốc độ ăn mòn; đánh giá hiệu quả của mạng neural nhân tạo. Kết quả dự đoán của mô hình mạng neural được sẽ được so sánh với kết quả tính toán theo tiêu chuẩn Norsok M-506.

Để đánh giá hiệu quả của quá trình huấn luyện và kiểm tra so sánh giữa các mô hình, các tiêu chuẩn đánh giá được sử dụng bao gồm: căn bậc hai của sai số bình phương trung bình RMSE (Root Mean Square Error), sai số tuyệt đối trung bình (Mean Absolute Error), hệ số xác định  $R^2$  (coefficient of determination). Giá trị RMSE biểu thị độ lớn trung bình của sai số. RMSE và MAE nằm trong khoảng  $(0, +\infty)$ , có giá trị lý tưởng khi bằng 0. Hệ số xác định  $R^2$  là thông số độ lệch thống kê đơn giản biểu thị sự phù hợp với dữ liệu của một phép hồi quy.  $R^2$  của một phép hồi quy tuyến tính thường nằm trong khoảng  $(0,1)$ . Lý tưởng nhất khi  $R^2 = 1$  nghĩa là mô hình hồi quy cho giá trị dự đoán hoàn toàn phù hợp với dữ liệu mà không có sai số. Khi  $R^2 = 0$ , mô hình hồi quy không giải thích được bất cứ dữ liệu nào.  $R^2$  cũng có thể đạt giá trị âm, khi mà mô hình hồi quy cho kết quả tổng các độ lệch bình phương phần dư lớn hơn tổng các độ lệch bình phương toàn bộ, nghĩa là mô hình không phù hợp để dự đoán xét trên quan điểm thống kê.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (1)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (3)$$

Trong đó:  $y_i$  - giá trị thực tế thứ  $i$  của  $n$  điểm dữ liệu;  $\hat{y}$  - giá trị dự đoán;  $\bar{y}$  - giá trị trung bình.

### 2.1. Mô hình dự đoán tốc độ ăn mòn theo tiêu chuẩn Norsok M-506

Đây là mô hình thực nghiệm được xây dựng

bởi các công ty dầu khí Na Uy: Statoil, Norsk Hydro và Saga Petroleum và thuộc sở hữu của Norway Petroleum Industry (Norsok M-506, 2017). Mô hình này dựa trên các kết quả từ phòng thí nghiệm. So với các mô hình thực nghiệm khác, mô hình này có ưu thế khi có thể dự đoán cho khoảng nhiệt độ rộng từ  $5 \div 150^\circ\text{C}$ . Ngoài ra mô hình cũng hơn các mô hình khác khi tính đến tác dụng chống ăn mòn của lớp film phủ ở nhiệt độ cao, và độ pH cao. Trong phần mềm tính toán theo phiên bản tiêu chuẩn M- 506 mới nhất 2017, giới hạn để chương trình tính toán được là áp suất thành phần  $\text{H}_2\text{S}$  lớn hơn 0,05 bar, tỷ số áp suất thành phần  $\text{CO}_2$  trên áp suất thành phần  $\text{H}_2\text{S}$  nhỏ hơn 20, một bước tiến so với phiên bản 2005 khi áp suất thành phần  $\text{H}_2\text{S}$  phải lớn hơn 0,5 bar. Tuy nhiên mô hình này tương đối nhạy với sự thay đổi của pH. Khi tổng tỷ lệ của axit hữu cơ vượt qua 100ppm và áp suất thành phần  $\text{CO}_2$  nhỏ hơn 0,5 bar, mô hình dự đoán có thể không chính xác.

Công thức tính toán theo mô hình tiêu chuẩn Norsok M-506 như sau:

$$CR_t = K_t \times (fCO_2)^{0,62} \times \left(\frac{S}{19}\right)^{(0,146+0,0324 \times \log(fCO_2))} \times f(pH)_t \quad (4)$$

$CR_t$  - tốc độ ăn mòn (mm/năm) tại nhiệt độ  $T$  ( $^\circ\text{C}$ );  $K_t$  - hằng số cân bằng tại nhiệt độ  $T$ ,  $fCO_2$  - độ loãng (fugacity) của  $\text{CO}_2$  (bar);  $f(pH)_t$  - hệ số tác nhân pH tại nhiệt độ  $T$ ;  $S$  - ứng suất ma sát lên thành (wall shear stress) đơn vị Pascal (Pa)

### 2.2. Mô hình mạng neural nhân tạo dự đoán tốc độ ăn mòn

Mạng neural nhân tạo là một thuật toán học máy (machine learning) xử lý thông tin được thiết kế mô phỏng chức năng và sự phức tạp của hệ thống neural sinh học (Robert, 1997). Neural sinh học là một khối cơ sở của hệ thống neural bao gồm 3 phần chính: thân neural (soma), sợi nhánh (dendrite) và sợi trục (axon). Thông tin ở dạng tín hiệu điện được nhận bởi các nhánh, được xử lý bởi thân neural sau đó truyền qua sợi trục. Mô hình mạng neural nhân tạo được sử dụng rộng rãi nhất là mô hình mạng neural đa lớp MLP (Multi Layers Perceptron) sử dụng thuật toán lan truyền ngược (back propagation).

Kiến trúc cơ bản một mạng neural đa lớp có cấu trúc bao gồm: 1 lớp đầu vào,  $n$  lớp ẩn ( $n > 1$ ) và 1 lớp đầu ra. Quá trình học của một mạng neural

nhân tạo nói trên thông thường bao gồm các bước: lan truyền tiến (feedforward), lựa chọn và tối ưu hóa hàm mất mát (loss function) thông qua việc tính đạo hàm hàm mất mát nhờ thuật toán lan truyền ngược dựa theo quy tắc chuỗi đạo hàm của hàm hợp (Robert, 1997; Geoffrey, 2014)

Mô phỏng neural sinh học, tiến trình lan truyền tiến diễn ra như sau: qua những đầu vào (input), mỗi đầu vào được nhân bởi một hàm trọng số trước khi đi vào node (unit, neural). Tại các node, diễn ra quá trình xử lý: tất cả dữ liệu đầu vào được nhân với trọng số (weight), cộng với hệ số bias và xử lý tổng qua một hàm kích hoạt (activation function) trước khi truyền thông tin tới những node của lớp tiếp theo và cứ như vậy đến lớp đầu ra. Các hàm kích hoạt thường được sử dụng là: ReLU, Sigmoid, Tanh. Hàm mất mát được lựa chọn cần thể hiện sự chênh lệch giữa hai đại lượng: giá trị được dự đoán và giá trị thực tế sao cho đơn giản hóa việc tính đạo hàm. Trong quá trình đào tạo, trọng số và bias được thay đổi bởi thuật toán tối ưu cho tới khi giá trị của hàm mất mát đủ nhỏ.

2.2.1 Chuẩn bị dữ liệu đầu vào

Để thực hiện nghiên cứu này, nhóm tác giả sử dụng tập dữ liệu được Obaseki M. dùng trong nghiên cứu (Obaseki et al., 2017) đã nêu ở trên và tiến hành tiền xử lý dữ liệu. Việc tiền xử lý dữ liệu là cần thiết nhằm mục đích: lựa chọn các thông số đặc trưng cho mô hình, lọc các mẫu dữ liệu (instants); làm sạch - xóa bỏ các điểm dữ liệu nhiễu, dữ liệu bất thường, chuẩn hóa dữ liệu... Kết quả của bước tiền xử lý dữ liệu, nhóm tác giả sử dụng tập hợp 40 điểm dữ liệu với 9 đặc trưng sau làm dữ liệu đầu vào cho mô hình ANN: đường kính

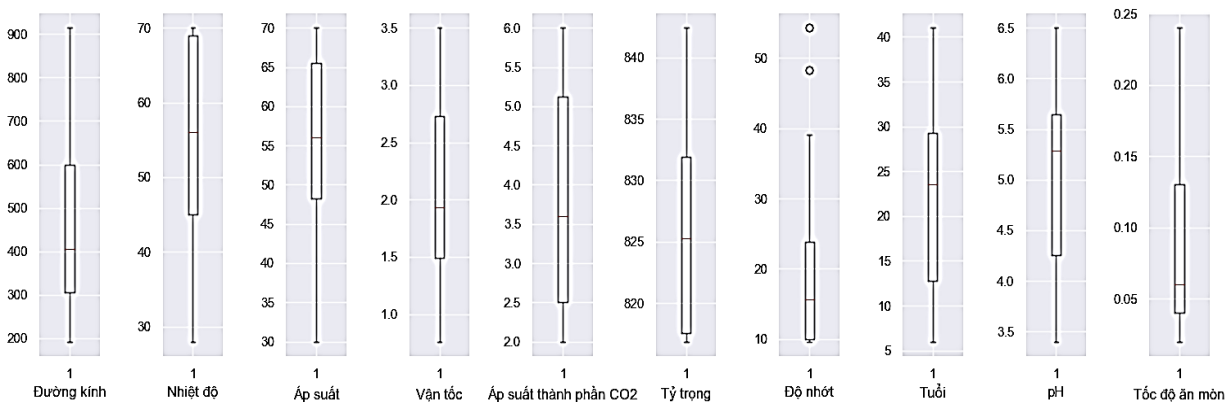
(mm), tuổi (năm), nhiệt độ (°C), áp suất thành phần CO<sub>2</sub> (bar), tốc độ dòng chảy chất lưu (m/s), áp suất chất lưu (bar), tỷ trọng riêng chất lưu (kg/m<sup>3</sup>), độ nhớt động lực học (cP), độ pH (pH). Dữ liệu sử dụng cho nghiên cứu được tóm tắt dưới dạng biểu đồ hộp như trong Hình 1.

Hình 1 tóm tắt các đặc trưng dữ liệu sử dụng cho nghiên cứu này với: đường kính ống dao động từ 192,7 mm đến 914 mm, tuổi ống từ 6 đến 41 năm, nhiệt độ chất lưu từ 28°C đến 70°C, áp suất từ 30 bar đến 70 bar, tốc độ dòng chảy chất lưu từ 0,76 m/s đến 3,5 m/s, áp suất thành phần CO<sub>2</sub> từ 2 bar đến 6 bar, tỷ trọng riêng chất lưu từ 816,88 kg/m<sup>3</sup> đến 842,4 kg/m<sup>3</sup>, độ nhớt động lực học chất lưu từ 9,65 cP đến 54,3 cP; tốc độ ăn mòn từ 0,02 mm/năm đến 0,24 mm/năm.

Hệ số tương quan Pearson (Pearson correlation coefficient) giữa các đặc trưng dữ liệu đầu vào được tính toán và xác định trong bảng 1. Các giá trị trong bảng 2 cho thấy các đặc trưng đầu vào hầu hết có mối tương quan thấp với nhau, ngoại trừ 2 cặp đặc trưng tỷ trọng riêng - độ nhớt động lực học, tỷ trọng riêng - nhiệt độ vốn là các đại lượng vật lý có mối tương quan cao, hoàn toàn tuân theo các phương trình vật lý. Trong tập dữ liệu đang xét, hệ số tương quan giữa tuổi ống và tốc độ ăn mòn là R = 0,55, vì thế có thể xác định tuổi ống là một đặc trưng không thể bỏ qua khi xây dựng mô hình dự đoán tốc độ ăn mòn.

2.2.2. Chuẩn hóa dữ liệu đầu vào cho mạng neural nhân tạo

Do các điểm dữ liệu trên với 9 đặc trưng khác nhau, đơn vị đo khác nhau, giá trị chênh lệch nhau quá lớn, nhóm nghiên cứu chuẩn hóa dữ liệu (feature normalization) sử dụng min-max scaler:



Hình 1. Tóm tắt bộ dữ liệu sử dụng cho nghiên cứu.

Bảng 1. Kết quả phân tích tương quan giữa các đặc trưng đầu vào.

Tương quan	Đường kính	Tuổi ống	Nhiệt độ	Áp suất	Vận tốc	Áp suất thành phần CO <sub>2</sub>	pH	Tỷ trọng	Độ nhớt
Đường kính	1.00	0.35	0.08	0.29	-0.68	0.07	-0.09	-0.08	-0.09
Tuổi	0.35	1.00	-0.02	0.16	-0.17	0.13	-0.09	0.02	0.04
Nhiệt độ	0.08	-0.02	1.00	0.04	-0.28	0.04	-0.15	-1.00	-0.95
Áp suất	0.29	0.16	0.04	1.00	-0.23	-0.05	0.10	-0.04	-0.08
Vận tốc	-0.68	-0.17	-0.28	-0.23	1.00	0.12	0.04	0.28	0.37
Áp suất thành phần CO <sub>2</sub>	0.07	0.13	0.04	-0.05	0.12	1.00	-0.19	-0.04	0.05
pH	-0.09	-0.09	-0.15	0.10	0.04	-0.19	1.00	0.14	0.04
Tỷ trọng	-0.08	0.02	-1.00	-0.04	0.28	-0.04	0.14	1.00	0.95
Độ nhớt	-0.09	0.04	-0.95	-0.08	0.37	0.05	0.04	0.95	1.00

$$x' = (x - \min(x)) / (\max(x) - \min(x)) \quad (5)$$

trong đó  $x$  là giá trị ban đầu,  $x'$  là giá trị sau khi chuẩn hóa. Giá trị cực tiểu  $\min(x)$ , giá trị cực đại  $\max(x)$  được tính trên toàn bộ dữ liệu ở cùng một đặc trưng. Kết quả là các giá trị các đặc trưng nằm trong khoảng  $[-1, 1]$ . Qua đó, giúp việc huấn luyện mô hình dễ dàng hơn và chính xác hơn, ít phụ thuộc vào độ lớn và đơn vị của từng đặc trưng của điểm dữ liệu.

### 2.2.3. Xây dựng mô hình mạng neural nhân tạo.

Mô hình mạng neural nhân tạo dự đoán tốc độ ăn mòn được nghiên cứu thuộc về dạng học có giám sát dạng hồi quy phi tuyến. Để tránh hiện tượng mô hình mạng quá khớp (overfitting) hoặc chưa khớp (underfitting) với tập huấn luyện và nâng cao chất lượng mô hình mạng ANN, dữ liệu ban đầu được phân chia làm 3 phần: đào tạo (training), xác thực (validation), và kiểm tra (testing). Tỷ lệ lần lượt là: phần dữ liệu huấn luyện chiếm 62,5% (25 điểm dữ liệu), bộ dữ liệu xác thực có 12,5% (5 điểm dữ liệu), bộ dữ liệu kiểm tra chiếm 25% (10 điểm dữ liệu). Do lượng dữ liệu thu thập được còn ít, nhóm tác giả sử dụng hàm ReLU, thuật toán Early Stopping, RMSprop tăng chất lượng dự đoán của mô hình, giảm độ phụ thuộc vào số lượng dữ liệu đầu vào.

Số lớp ẩn, số lượng node (unit), các hàm kích hoạt, lựa chọn hàm mất mát (loss function, hay còn gọi là hàm giá trị - cost function) thuật toán tối ưu lan truyền ngược, là các lựa chọn quan trọng trong đào tạo mạng neural nhân tạo, giúp tăng độ chính xác và giảm thời gian tính toán. Số lớp ẩn và số lượng node trong mỗi lớp ẩn cần vừa đủ để đạt được độ chính xác cần thiết. Việc sử dụng nhiều

lớp ẩn hơn, hoặc nhiều node trong mỗi lớp hơn chỉ tăng thời gian tính toán mà không cải thiện độ chính xác. Qua nhiều thử nghiệm và liên tục tối ưu, nhóm tác giả sử dụng 18 node (unit) cho lớp ẩn đầu tiên và 9 node (unit) cho lớp ẩn thứ 2 như Hình 2.

Trong nghiên cứu này, nhóm tác giả sử dụng hàm ReLU (Rectified Linear Unit) (Hahnloser et al., 2000) cho lớp ẩn thứ nhất và hàm Sigmoid (Han et al., 1995) cho lớp ẩn tiếp theo. Hàm ReLU đã được chứng minh là giúp việc huấn luyện học máy và học sâu (deep learning) nhanh hơn rất nhiều, nguyên do đạo hàm của nó bằng 0 với  $x \leq 0$ , và bằng 1 khi  $x > 0$ . Công thức của hàm ReLU và hàm Sigmoid lần lượt như (6), (7).

ReLU:

$$f(x) = \max(0, x) \\ f(x) = 0 \text{ với } x \leq 0, f(x) = x \text{ với } x > 0 \quad (6)$$

Sigmoid:

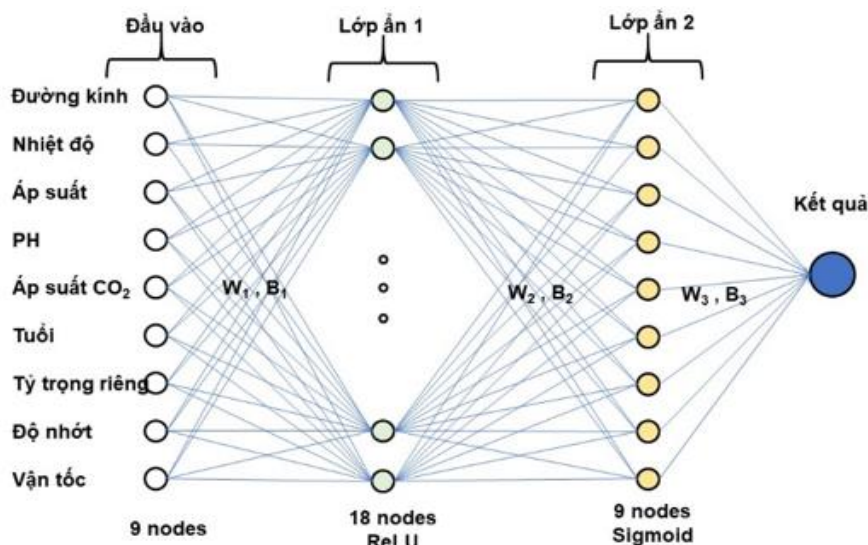
$$f(x) = \frac{1}{1 + e^{-x}} \quad (7)$$

Đồ thị của 2 hàm như Hình 3

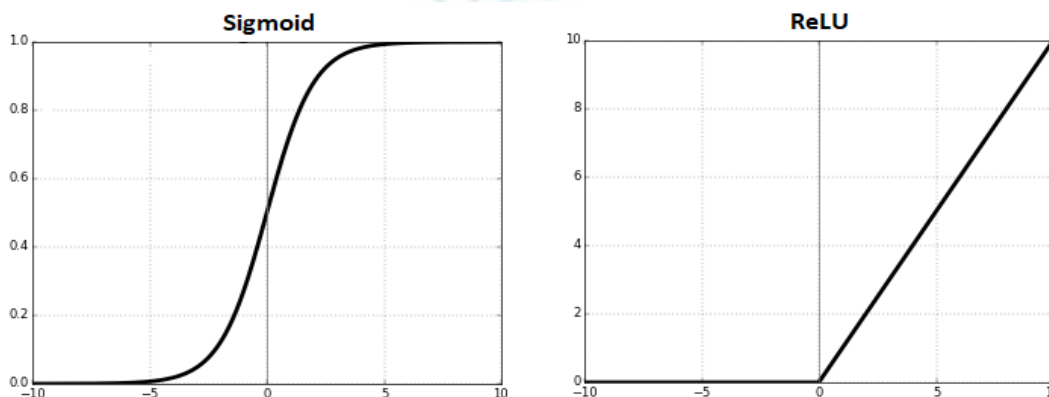
Hàm mất mát được sử dụng là hàm sai số bình phương trung bình, có công thức như (8).

$$mse = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (8)$$

Một vài thuật toán tối ưu đào tạo lan truyền ngược thường được sử dụng Gradient Decent (đạo hàm đi ngược), Stochastic Gradient Descent (SGD), thuật toán Levenberg-Marquardt Algorithm (LMA), Scaled Conjugate Gradient (SCG), Pola-Ribiere conjugate gradient (PCG), Adadelta (Matthew, 2012), Adam, Adagrad và



Hình 2. Mô hình mạng neural nhân tạo dự đoán tốc độ ăn mòn.



Hình 3. Đồ thị hàm số Sigmoid và ReLU.

những thuật toán khác...Nhóm tác giả sử dụng thuật toán RMSprop (Geoffrey, 2014). Thuật toán này có ưu điểm tốc độ học (learning rate) tự động điều chỉnh, thay đổi lớn với các dữ liệu khác biệt nhiều và các thay đổi nhỏ cho các dữ liệu ít khác biệt; và nhanh chóng đạt được giá trị cực tiểu toàn cục.

Kỹ thuật Early stopping - dừng thuật toán trước khi hàm mất mát quá nhỏ, giúp tránh hiện tượng quá khớp (overfitting) cũng được áp dụng trong nghiên cứu này. Các thiết lập Early stopping gồm: chỉ tiêu đánh giá là sai số tuyệt đối trung bình, và số epoch tối đa bằng 50 (epoch - số lần duyệt qua tất cả các điểm dữ liệu) trước khi giá trị chỉ tiêu đánh giá của bộ dữ liệu xác thực có chiều hướng tăng lên.

### 3. Kết quả nghiên cứu và thảo luận

Với mô hình mạng neural nhân tạo được xây

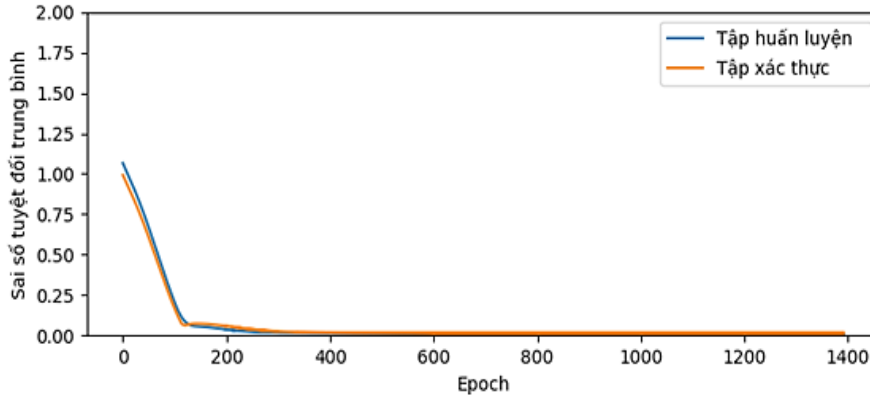
dựng như mục trên, tập dữ liệu huấn luyện và xác thực được sử dụng để đào tạo mạng. Cùng với việc áp dụng kỹ thuật Early Stopping, quá trình đào tạo kết thúc với số lần lặp (epoch) bằng 1293. Sự thay đổi của chỉ tiêu - sai số tuyệt đối trung bình của tập huấn luyện và tập xác thực trong Hình 4.

Bảng 2 cho thấy mô hình mạng neural nhân tạo cho kết quả dự báo khá tốt với giá trị hệ số xác định của tập dữ liệu kiểm tra  $R^2=0,868$ . Chỉ số  $R^2$  của tập dữ liệu kiểm tra thấp hơn tập dữ liệu huấn luyện và xác thực là do số lượng các điểm dữ liệu còn ít. Các chỉ số đánh giá của mô hình mạng neural nhân tạo dự đoán tốc độ ăn mòn trên toàn bộ tập dữ liệu lần lượt có giá trị RMSE = 0,014, MAE = 0,011,  $R^2 = 0,938$ . Ảnh hưởng của từng thông số đầu vào (áp suất CO<sub>2</sub>, độ pH, tuổi ...) lên tốc độ ăn mòn chưa được xét đến và sẽ đề cập đến trong những nghiên cứu sau. Các kết quả dự đoán tốc độ ăn mòn của trên bộ dữ liệu được so sánh

với thực tế và được minh họa trong Hình 5.

Sử dụng mô hình thực nghiệm dự đoán ăn mòn CO<sub>2</sub> theo tiêu chuẩn Norsok M-506 để tính toán bộ dữ liệu, nhóm tác giả thu được các chỉ số đánh giá RMSE = 0,064, MAE = 0,048, R<sup>2</sup> = -0,364. Hệ số xác định âm cho thấy kết quả dự đoán mô hình Norsok M-506 chưa mô phỏng được thực tế

bộ dữ liệu trên. Mạng neural nhân tạo cho thấy khả năng dự đoán vượt trội hơn so với mô hình thực nghiệm theo tiêu chuẩn Norsok M-506. Bảng 3 so sánh hiệu suất của các mô hình dự đoán tốc độ ăn mòn. Hình 6 minh họa giá trị tốc độ ăn mòn dự đoán của các mô hình so với tốc độ ăn mòn thực tế.

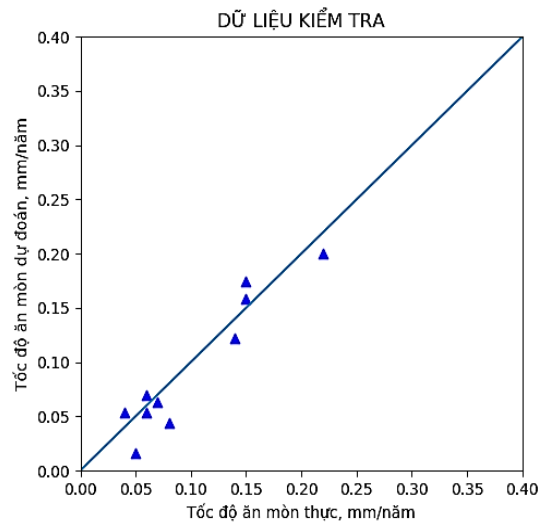
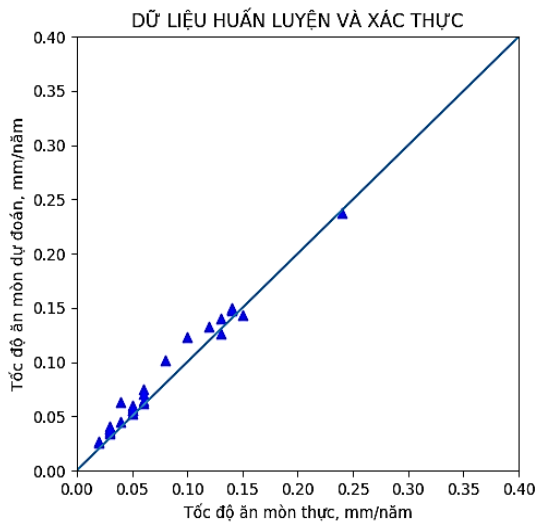


Hình 4. Early stopping; sai số tuyệt đối trung bình của tập huấn luyện và tập xác thực.  
Bảng 2. Các chỉ tiêu của mạng neural nhân tạo dự đoán tốc độ ăn mòn.

Tập dữ liệu	RMSE	MAE	R <sup>2</sup>
Huấn luyện và xác thực	0,010	0,008	0,962
Kiểm tra	0,020	0,017	0,868

Bảng 3. Hiệu suất của các mô hình dự đoán tốc độ ăn mòn.

Mô hình	RMSE	MAE	R <sup>2</sup>
Tiêu chuẩn Norsok M-506	0,064	0,048	-0,364
Mô hình mạng neural nhân tạo	0,014	0,011	0,938



Hình 5. Đồ thị hồi quy của mạng neural nhân tạo dự đoán tốc độ ăn mòn.