

NGHIÊN CỨU XÂY DỰNG CÔNG CỤ HỌC MÁY DỰ BÁO NHIỆT ĐỘ ĐẦU RA CẤP 1 VÀ CẤP 2 CỦA MÁY NÉN KHÍ CAO ÁP

Trần Ngọc Trung, Nguyễn Thành Trung, Nguyễn Duy Minh, Đào Quang Khoa, Trần Vũ Tùng, Hoàng Kỳ Sơn

Công ty Điều hành Dầu khí Biển Đông

Email: trungtn@biendongpoc.vn

<https://doi.org/10.47800/PVSI.2024.01-08>

Tóm tắt

Máy nén khí là thiết bị quan trọng trên giàn xử lý trung tâm PQP-HT. Khí tự nhiên, sau quá trình khử nước và đảm bảo nhiệt độ điểm sương theo điều kiện và thông số kỹ thuật đầu vào đường ống dẫn khí Nam Côn Sơn (NCSP), được chuyển đến hệ thống nén khí bao gồm 2 dây chuyền máy nén khí. Việc tối ưu hóa điều kiện vận hành bằng cách giảm áp suất đầu vào của hệ thống xử lý khí tự nhiên được sử dụng để nâng cao hiệu quả sản xuất của các mỏ khí. Tuy nhiên, việc thay đổi điều kiện vận hành đầu vào ảnh hưởng trực tiếp đến hệ thống máy nén khí khi áp suất đầu vào giảm và nhiệt độ đầu ra của máy nén khí có thể tăng lên vượt quá ngưỡng vận hành an toàn.

Để tính toán ảnh hưởng của việc thay đổi điều kiện vận hành của máy nén khí lên nhiệt độ đầu ra của từng giai đoạn, thường sử dụng phần mềm mô phỏng nhiệt động lực học thương mại (như Hysys, Proll). Từ đó, có thể mô phỏng và chọn điều kiện làm việc tối ưu, đảm bảo an toàn cho hệ thống xử lý khí tự nhiên. Tuy nhiên, chi phí cấp phép và duy trì phần mềm thương mại cao. Ngoài ra, các thuật toán học máy hiện đại được chứng minh có thể dự báo các thông số vận hành dựa trên dữ liệu lịch sử. Nhiều nghiên cứu đã tìm cách dự báo chính xác hiệu suất của máy nén, nhằm cải thiện hiệu quả của các hoạt động vận hành. Thuật toán học máy có ưu điểm là kết quả dự báo có độ chính xác cao, mô hình có thể hoạt động liên tục và tự động tái học khi thay đổi điều kiện vận hành. Do đó, các mô hình học máy có thể được sử dụng làm phương án thay thế phần mềm mô phỏng nhiệt động lực học thương mại.

Từ khóa: Học máy, máy nén khí, dự báo nhiệt độ đầu ra máy nén.

1. Giới thiệu

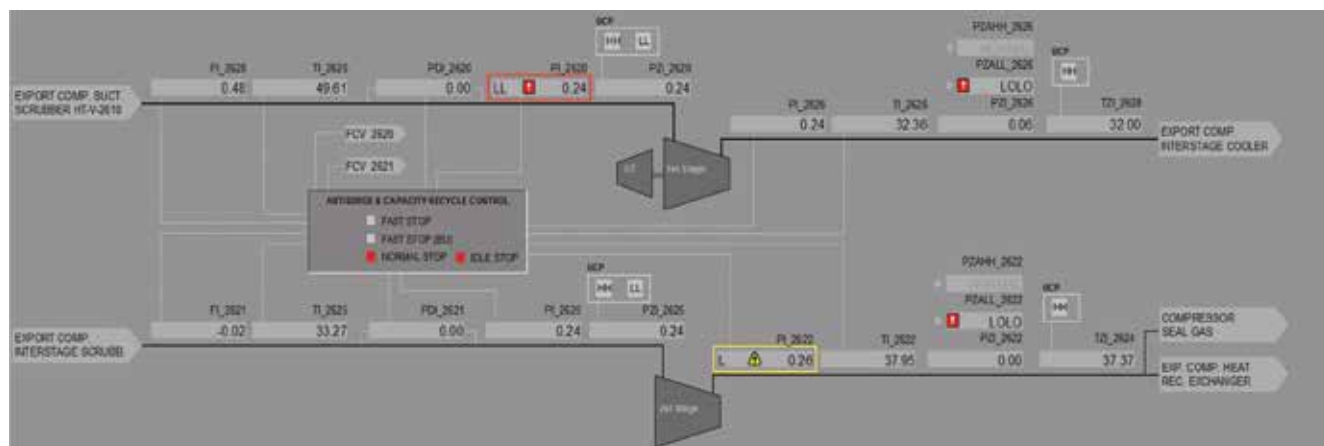
Giàn khai thác dầu khí Hải Thạch - Mộc Tinh bao gồm 1 giàn đầu giếng (WHP-MT1) tại Lô 05-3 (mỏ Mộc Tinh), 1 giàn đầu giếng (WHP-HT1) cùng 1 giàn xử lý trung tâm (PQP-HT) tại Lô 05-2 (mỏ Hải Thạch) do Công ty Điều hành Dầu khí Biển Đông xây dựng và đưa vào vận hành khai thác từ năm 2012. Máy nén khí là một thiết bị công nghệ quan trọng trên hệ thống giàn khoan khai thác khí và công nghệ xử lý khí trên giàn xử lý trung tâm. Trên giàn khoan Hải Thạch - Mộc Tinh, khí sau khi đã tách nước và đảm bảo nhiệt độ điểm sương theo chỉ tiêu chất lượng sản phẩm để ra được đưa đến hệ thống nén khí bao gồm 2 dây chuyền máy nén. Mỗi dây chuyền bao gồm một hệ thống máy nén 2 cấp với thiết bị tách lỏng đầu vào máy nén ở mỗi cấp và thiết bị làm lạnh dòng khí ở đầu ra của mỗi cấp (Hình 1).

Mỗi dây chuyền được thiết kế với công suất 5 triệu m³/ngày. Trong thời gian đầu, áp suất đầu vào mỗi dây chuyền là 61,7 barg và dự kiến vào cuối đời mỏ khai thác, áp suất đầu vào của mỗi dây chuyền sẽ giảm xuống 41,7 barg. Áp suất đầu ra của mỗi dây chuyền đạt mức cực đại là 158 barg trước khi xuất khí qua đường ống Nam Côn Sơn (NCSP) để đưa về bờ. Để tối ưu hóa hiệu quả kinh tế và tăng khả năng thu hồi khí và dầu của dự án, áp suất đầu vào của mỗi dây chuyền có thể được giảm thấp hơn giá trị thiết kế ban đầu là 41,7 barg. Khi giảm điều kiện vận hành áp suất đầu vào của mỗi dây chuyền ra ngoài khoảng vận hành theo thiết kế, nhiệt độ đầu ra tại mỗi cấp của máy nén sẽ tăng cao và có thể ảnh hưởng đến tính toàn vẹn của thiết bị. Khi áp suất đầu vào giảm, nhiệt độ đầu ra của máy nén có thể tăng vượt quá ngưỡng làm việc an toàn dẫn đến hư hỏng các đệm làm kín, cũng như dầu nhớt bôi trơn sử dụng cho máy nén. Các phần mềm thương mại mô phỏng nhiệt động học như Hysys, Proll thường được sử dụng để dự báo ảnh hưởng của việc thay đổi điều kiện vận hành lên hệ thống công nghệ (như điều kiện vận



Ngày nhận bài: 1/8/2023. Ngày phản biện đánh giá và sửa chữa: 2 - 18/10/2023.

Ngày bài báo được duyệt đăng: 27/12/2023.



Hình 1. Sơ đồ công nghệ của hệ thống máy nén khí trên giàn công nghệ trung tâm (PQP-HT).

hành của máy nén lên nhiệt độ đầu ra của mỗi cấp), từ đó người vận hành có thể mô phỏng và lựa chọn các điều kiện làm việc tối ưu, đảm bảo an toàn cho hệ thống công nghệ. Tuy nhiên, chi phí bản quyền phần mềm thương mại cao, hơn 30.000 USD/năm.

Bên cạnh đó, có thể ứng dụng thuật toán học máy để dự báo thông số vận hành dựa trên cơ sở dữ liệu vận hành trong quá khứ. Nhiều nghiên cứu đã tìm cách đưa ra những dự báo chính xác cho hiệu suất hoạt động của máy nén, nhằm nâng cao hiệu quả của công tác tối ưu hóa vận hành và mô phỏng hệ thống công nghệ. Nhiều phương pháp đã được sử dụng như mạng thần kinh nhân tạo (ANN) và các thuật toán dựa trên lý thuyết xác suất thống kê khác.

Máy nén khí lý tâm là thiết bị được sử dụng rộng rãi trong nhiều lĩnh vực kỹ thuật ứng dụng khác nhau. Việc dự báo hiệu năng cũng như điều kiện làm việc của máy nén là một trong những công việc chính trong quá trình thiết kế, tối ưu hóa vận hành và mô phỏng hệ thống. Để dự báo các thông số vận hành của máy nén, người ta thường sử dụng các phương pháp tính toán truyền thống dựa trên phương trình khí lý tưởng [1, 2] $PV^k = \text{hằng số}$, trong đó k là hệ số tỷ lệ giữa nhiệt dung riêng đẳng áp (C_p) và nhiệt dung riêng đẳng tích (C_v). Các công cụ tính toán thủ công thường sử dụng các phương trình nhiệt động dành cho khí lý tưởng bởi tính đơn giản của công thức tính toán. Tuy nhiên, kết quả thu được thường có độ sai lệch lớn so với số liệu thực tế.

Để tăng độ chính xác của các công thức tính toán, John M. Schultz [3] đã đưa ra các phương trình tính toán cho khí thực (real gas). Bên cạnh hệ số nén của khí Z, tác giả đã đưa thêm các hàm số X và Y. Các hàm số bổ sung này có thể được khái quát hóa và vẽ biểu đồ cho các tính chất của chất khí ở các trạng thái tương ứng. Tuy nhiên, các phương trình tính toán của tác giả phức tạp, thường

được đưa vào cơ sở dữ liệu tính toán nhiệt động học cho các phần mềm mô phỏng thương mại hiện nay như Hysys, Pro II – đều có chi phí bản quyền rất lớn. Trong những năm gần đây, nhiều nghiên cứu đã tìm cách đưa ra những dự báo chính xác cho hiệu suất hoạt động của máy nén nhằm nâng cao hiệu quả của công tác thiết kế, tối ưu hóa và mô phỏng hệ thống với chi phí thấp hơn việc sử dụng các phần mềm mô phỏng. Nhiều phương pháp đã được sử dụng như nội suy, điều chỉnh đường cong (curve fitting), mạng thần kinh (NN - neural network) và các thuật toán khác dựa trên lý thuyết xác suất thống kê. Tuy nhiên, các phương pháp này thường cần một lượng lớn dữ liệu và việc thu thập dữ liệu có thể tốn đáng kể tài nguyên máy tính hoặc tài nguyên thử nghiệm.

Wei Chunzhi [4] dựa trên các kết quả nghiên cứu của John M. Schultz [3] để dự báo nhiệt độ đầu ra của máy nén. Kết quả cho thấy, nhiệt độ đầu ra của máy nén tính theo các phương trình trạng thái của khí thực có độ chênh lệch khoảng 0,77%, trong khi giá trị nhiệt độ đầu ra của máy nén tính theo phương trình trạng thái khí lý tưởng có độ chênh lệch hơn 8% so với thông số thử nghiệm thực tế do nhà sản xuất máy nén cung cấp.

Yu Zhang cùng cộng sự [5] đã nghiên cứu dự báo nhiệt độ đầu ra của máy nén bằng cách sử dụng 1 mô hình hướng dữ liệu (data-driven model) và mô hình này được xây dựng dựa trên nền tảng cơ sở dữ liệu thử nghiệm (test bed) động cơ của nhà thầu. Tác giả đề xuất sử dụng mô hình hóa mờ thần kinh Bayes (Bayesian neuro - fuzzy modelling, BNFM), kết hợp hệ thống suy luận mờ thần kinh thích ứng (adaptive neuro - fuzzy inference system, ANFIS) và kỹ thuật mô hình hỗn hợp Bayesian Gaussian (variational Bayesian Gaussian mixture model, VBGM). Mô hình máy nén dựa trên các cơ sở dữ liệu được xây dựng bằng ANFIS và VBGM đã được áp dụng trong giai đoạn xây dựng thiết bị để tự động lựa chọn các dữ liệu

đầu vào trong hệ thống mờ (fuzzy system). Phương pháp tiếp cận BFNFM được thiết lập thông qua các thử nghiệm thực tế đối với turbine khí công suất dưới 15 MW. Kết quả cho thấy giá trị ước tính của nhiệt độ đầu ra máy nén bằng phương pháp sử dụng BFNFM đưa ra số liệu có độ tin cậy cao, do vậy, kết quả dự báo từ mô hình này có thể được sử dụng để thay thế trong trường hợp các cảm biến đo đặc hoạt động thiếu chính xác hay bị hư hỏng không thể sử dụng được. Bên cạnh đó, kết quả tính toán bằng mô hình cũng được sử dụng như một cảm biến ảo để cung cấp thêm bằng chứng cho việc giám sát tình trạng của máy nén bên cạnh các cảm biến vật lý hiện có.

Jingzhou Fei cùng cộng sự [6] đã sử dụng mạng thần kinh nhân tạo mới tích hợp mạng thần kinh truyền ngược nguồn cấp dữ liệu với chức năng chính của thuật toán Gaussian để dự báo hiệu suất hoạt động của máy nén. Kết quả nghiên cứu chỉ ra rằng mạng thần kinh được đề xuất có hiệu suất dự báo vượt trội so với mạng thần kinh truyền ngược nguồn cấp dữ liệu chuyển tiếp và máy vector hỗ trợ hiện có, đặc biệt là đối với phép ngoại suy với các mẫu nhỏ. Hơn nữa, nghiên cứu này được sử dụng trong việc tinh chỉnh mô hình dựa trên hiệu suất hiện có để cải thiện phân tích mô phỏng, giám sát tình trạng và chẩn đoán lỗi của máy nén turbine khí.

Hongsheng Jiang cùng cộng sự [7] tiến hành xây dựng mô hình tỷ lệ áp suất và hiệu suất đẳng entropy của máy nén bằng cách sử dụng một số lượng mẫu dữ liệu hạn chế trong khi vẫn đảm bảo được độ chính xác cao. Đầu tiên, dữ liệu mẫu được tạo ra bằng phương pháp mô phỏng Vista CCD. Tiếp đó, lưu lượng và tốc độ quay đã hiệu chỉnh được sử dụng làm các biến độc lập và các biểu thức hồi quy với giá trị trung bình của tỷ lệ áp suất và hiệu suất đẳng entropy được rút ra và đơn giản hóa thông qua phân tích nhiệt động lực học và phân tích tổn thất của máy nén ly tâm. Các mô hình kriging dựa trên mô hình mạng thần kinh và đa thức bậc 2 cũng đã được xây dựng. Kết quả cho thấy, khi tính toán dự báo bên trong phạm vi giới hạn của dữ liệu, mô hình dựa trên phân tích tổn thất (loss-analysis-based model) và mô hình Kriging tạo ra dự báo có độ chính xác cao hơn ngay cả khi có 1 cơ sở dữ liệu nhỏ và kết quả dự báo ổn định, trong khi mô hình mạng thần kinh chỉ cung cấp kết quả tốt hơn trong 1 tập dữ liệu lớn hơn. Đối với dự báo bên ngoài phạm vi giới hạn của dữ liệu, mô hình dựa trên phân tích tổn thất có thể cung cấp kết quả tương đối chính xác. Bên cạnh đó, để đào tạo và sử dụng mô hình dựa trên phân tích tổn thất cũng mất ít thời gian hơn so với các mô hình khác.

Như vậy, việc áp dụng các công cụ hiện đại như trí

tuệ nhân tạo và học máy đã cho thấy có nhiều lợi thế như thuật toán đơn giản, mô hình có thể vận hành liên tục, không cần hiệu chỉnh khi thay đổi điều kiện vận hành. Trong lĩnh vực xây dựng mô hình tiên báo hiệu suất hoạt động của máy nén ly tâm, cũng như dự báo sự thay đổi của các thông số vận hành một cách chính xác, các mô hình này được xem như là một giải pháp thay thế để so sánh với các phần mềm mô phỏng thương mại như Hysys. Do vậy, mục đích chính của nghiên cứu này là ứng dụng các công cụ trí tuệ nhân tạo và học máy vào việc dự báo nhiệt độ đầu ra cấp 1 và cấp 2 của máy nén khí EGC tại mỏ Hải Thạch - Mộc Tinh nhằm tối ưu hóa hoạt động sản xuất, cũng như tiết kiệm các chi phí sử dụng phần mềm chuyên dụng.

2. Cơ sở lý thuyết

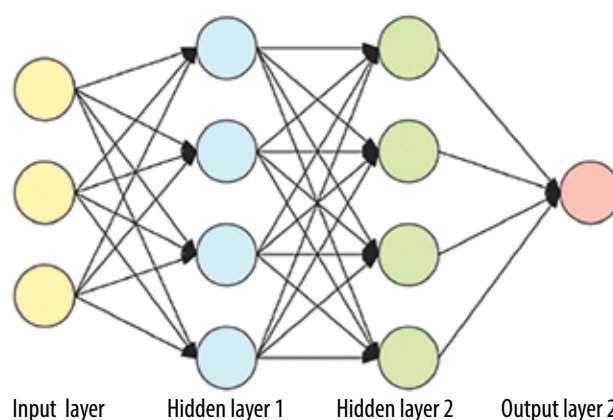
2.1. Tổng quan

Mạng thần kinh nhân tạo (ANN) và mạng bộ nhớ dài hạn, ngắn hạn (LSTM) được nhóm lựa chọn là 2 mô hình để thực hiện dự báo nhiệt độ đầu ra của máy nén. Về cấu trúc, cả 2 mô hình nhóm sử dụng đều bao gồm 4 lớp mạng xếp chồng lên nhau, giữa các lớp mạng có sử dụng lớp dropout để làm chậm quá trình đào tạo và tránh hiện tượng quá khớp (overfitting).

2.2. Mạng thần kinh nhân tạo (ANN)

Artificial neural network (ANN) hay mạng thần kinh nhân tạo là 1 hệ thống tính toán lấy cảm hứng từ sự hoạt động của các tế bào thần kinh trong hệ thần kinh. Chúng ta có thể coi ANN là dữ liệu thống kê phi tuyến. Điều này đồng nghĩa với một mối quan hệ phức tạp, được xác định giữa đầu vào và đầu ra, kết quả là sẽ có nhiều mẫu khác nhau.

ANN gồm 3 thành phần chính: Lớp dữ liệu đầu vào (input layer) và lớp dữ liệu đầu ra (output layer) chỉ gồm



Hình 2. Mạng thần kinh nhân tạo - Artificial neural network.

1 lớp, lớp ẩn (hidden layer) có thể có 1 hay nhiều lớp tùy vào bài toán cụ thể. ANN hoạt động theo hướng mô phỏng lại cách hoạt động của hệ thần kinh với các neuron được kết nối với nhau.

ANN được sử dụng để giải quyết các vấn đề liên quan đến dữ liệu dạng bảng, hình ảnh hay văn bản. ANN có khả năng học bất kỳ hàm phi tuyến nào, do đó, thường được biết đến cái tên “phương pháp xấp xỉ hàm phổ quát” (universal function approximators). Một trong những lý do chính đằng sau tính xấp xỉ phổ quát là hàm kích hoạt. Các hàm kích hoạt thể hiện các thuộc tính phi tuyến cho mạng. Điều này giúp mạng tìm hiểu bất kỳ mối quan hệ phức tạp nào giữa đầu vào và đầu ra.

Tuy nhiên, ANN cũng có một số nhược điểm. ANN dễ xảy ra hiện tượng gradient biến mất (vanishing gradient) và gradient bùng nổ (exploding gradient) như một đặc điểm cố hữu của các mạng thần kinh, nhất là khi số lượng lớp ẩn lớn. Trong quá trình huấn luyện các mạng thần kinh, trong một số trường hợp, thời gian huấn luyện có thể kéo dài hay kết quả trả về có độ chính xác không như mong muốn. Một trong những nguyên nhân gây nên những hiện tượng trên có liên quan đến gradient (tốc độ biến thiên của hàm số), hay cụ thể hơn là vanishing/exploding gradients. Gradient của 1 hàm số là đạo hàm của hàm số đó tương ứng với mỗi biến của hàm. Đối với hàm số đơn biến, chúng ta sẽ sử dụng khái niệm derivative thay cho gradient. Đối với hàm đa biến, các đạo hàm thành phần tương ứng với 1 biến nào đó được gọi là đạo hàm riêng (đạo hàm từng phần hay partial derivative).

Tuy nhiên, trên thực tế gradient thường sẽ có giá trị nhỏ dần khi đi xuống các lớp thấp hơn. Kết quả là các cập nhật thực hiện bởi gradient descent không làm thay đổi nhiều trọng số của các lớp đó, khiến chúng không thể hội tụ và ANN sẽ không thu được kết quả tốt. Hiện tượng này được gọi là gradient biến mất (vanishing gradient). Trong nhiều trường hợp khác, gradient có thể có giá trị lớn hơn trong quá trình lan truyền ngược, khiến một số lớp có giá trị cập nhật cho trọng số quá lớn khiến chúng phân kỳ (phân rã), tất nhiên ANN cũng sẽ không có kết quả như mong muốn. Hiện tượng này được gọi là gradient bùng nổ (exploding gradient), và thường gặp khi sử dụng mạng thần kinh hồi quy (recurrent neural network, RNN). Tóm lại, trong quá trình huấn luyện ANN chúng ta có thể gặp phải các vấn đề liên

quan đến việc gradient không ổn định khiến cho tốc độ học (learning rate) của các lớp khác nhau chênh lệch khá nhiều.

Hai hiện tượng trên là một trong những nguyên nhân khiến mạng thần kinh không nhận được sự quan tâm trong một thời gian khá dài. Tuy nhiên, trong một nghiên cứu được thực hiện bởi Xavier Glorot và Yoshua Bengio năm 2010 [8], các tác giả đã đưa ra một số nguyên nhân dẫn đến hiện tượng trên. Trong đó, việc lựa chọn hàm kích hoạt (activation function) và kỹ thuật khởi tạo trọng số (weight initialization) là hai nguyên nhân chính.

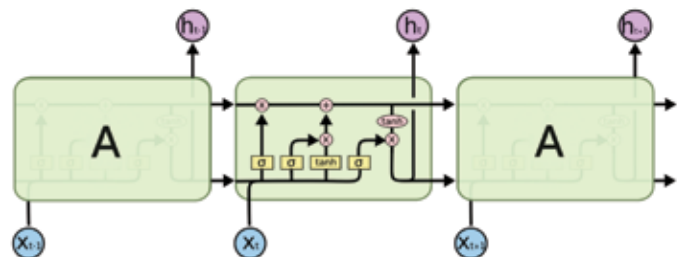
2.3. Mạng bộ nhớ dài - ngắn hạn (LSTM)

Mô hình đề xuất sử dụng mô hình mạng bộ nhớ dài - ngắn hạn (long short-term memory, LSTM) là một dạng đặc biệt của RNN. LSTM được giới thiệu bởi Hochreiter và Schmidhuber nhằm giải quyết các bài toán về phụ thuộc xa (long-term dependency) và lưu giữ các giá trị tính toán sau một khoảng thời gian tùy ý, nhằm loại bỏ phụ thuộc dài hạn. Mạng lưới cải tiến LSTM được biểu diễn ở Hình 3.

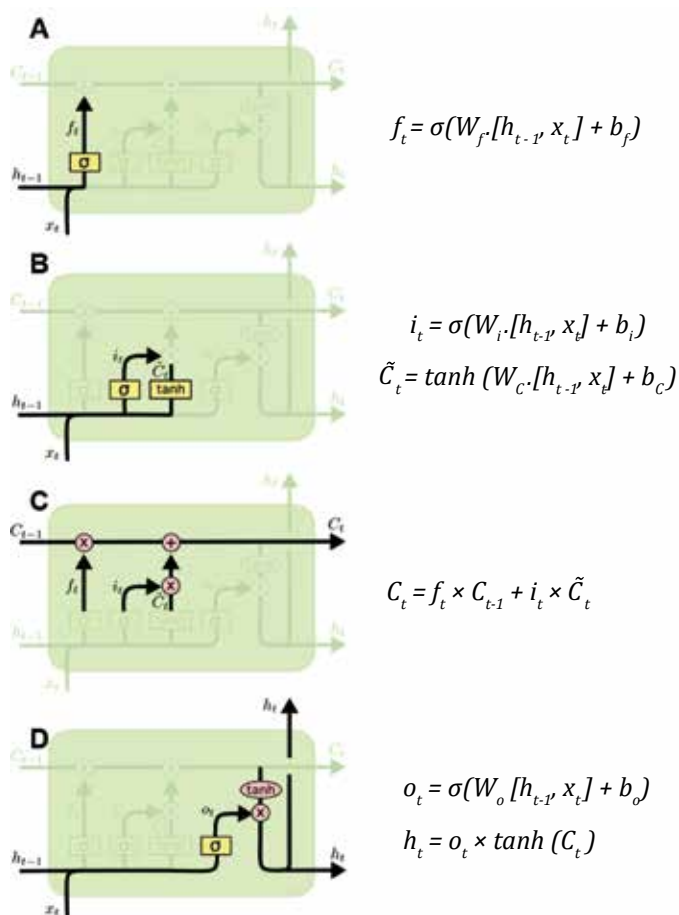
Một mạng lưới LSTM gồm có 4 thành phần chính: 1 đường trạng thái tế bào (cell state) C_t và 3 cổng (gate) với hàm kích hoạt $\sigma: x \mapsto [0, 1]$ (Sigmoid activation function) để kiểm soát mức độ thông quan của dữ liệu. Các cổng kết nối với đường trạng thái bằng một phép toán (hay được gọi là mắt xích).

Đường trạng thái tế bào C_t là đường truyền chạy xuyên suốt các mắt xích giúp thông tin có thể được gán và truyền qua LSTM mà không bị thay đổi.

Cổng thứ nhất trong Hình 4a là tầng cổng quên (forget gate layer) với đầu vào $W_f h_{t-1} + b_f$ (dữ liệu từ đầu ra của mạng LSTM $t - 1$) và $W_f x_t + b_f$ (dữ liệu bổ sung tại t) với tác dụng loại bỏ thông tin không mong muốn. Cổng thứ hai trong Hình 4b bao gồm một tầng cổng vào (input gate layer) với hàm kích hoạt σ để lựa chọn giá trị nào cập nhật vào tầng tanh: $x \mapsto [-1, 1]$. Sau khi thu được vector giá trị tế bào mới \tilde{C}_t , trạng thái tế bào cũ C_{t-1} sẽ được cập nhật thành trạng thái mới C_t trong Hình 4c. Cuối cùng, tại tầng cổng lọc (filter gate layer), hàm kích hoạt σ lựa chọn loại đầu ra cho C_t mà ta mong muốn, đầu ra của bộ lọc là tích của hàm σ và hàm tanh trong Hình 4d.



Hình 3. Đồ thị minh họa mạng lưới và phương trình LSTM [9].



Hình 4. Các bộ phận cấu thành mạng lưới LSTM [9].

Bảng 1. Danh sách thiết bị đo sử dụng cho bộ dữ liệu

Thiết bị	Mô tả	Đơn vị
TT-2620	Nhiệt độ khí vào máy nén cấp 1	°C
PT-2620	Áp suất khí vào máy nén cấp 1	barg
TT-2626	Nhiệt độ khí ra máy nén cấp 1	°C
PT-2626	Áp suất khí ra máy nén cấp 1	barg
TT-2625	Nhiệt độ khí vào máy nén cấp 2	°C
PT-2625	Áp suất khí vào máy nén cấp 2	barg
TT-2622	Nhiệt độ khí ra máy nén cấp 2	°C
PT-2622	Áp suất khí ra máy nén cấp 2	barg

Qua phân tích cơ chế hoạt động của LSTM, có thể thấy mô hình xử lý loại bỏ hoàn toàn cơ chế truyền ngược (backpropagation) trong các mạng lưới RNN phổ biến khác [10]. Điều này giúp mạng LSTM tạo cầu nối giữa các tế bào trạng thái rất cũ với trạng thái hiện tại, tăng hiệu suất xử lý với độ phức tạp (computational complexity) là 0(1).

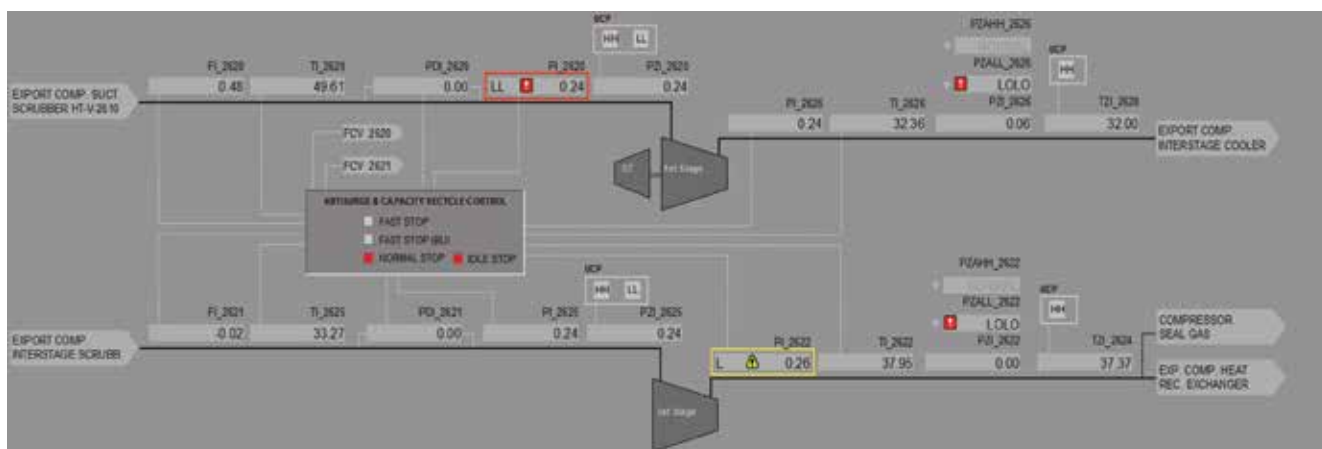
3. Tổng quan và tiền xử lý dữ liệu

3.1. Tổng quan dữ liệu

Hệ thống máy nén khí xuất bán khí qua đường ống Nam Côn Sơn bao gồm 2 dây chuyền máy nén khí hai cấp. Trong nghiên cứu này, nhóm tác giả sử dụng dữ liệu được trích xuất từ hệ thống thu thập số liệu lịch sử của dây chuyền máy nén khí gas EGC số 1. Hệ thống này sử dụng máy nén khí hai cấp sử dụng động cơ turbine model TT130 của hãng Solar để dẫn động máy nén khí của hãng Dresser-rand với công suất thiết kế đạt 5 triệu m³ khí/ngày với áp suất nén từ khoảng 50 bar lên 125 bar. Luồng khí đầu vào sau cấp nén thứ nhất từ 50 bar lên 70 bar, sau đó được làm lạnh qua bộ làm mát trung gian (interstage cooler) và tiếp tục vào máy nén cấp 2 để tăng áp suất từ 70 bar tới 125 bar.

Bộ dữ liệu là thông số vận hành của các cảm biến được lắp trên hệ thống máy nén như Hình 5. Các thiết bị đo được thu thập trong bộ dữ liệu được thể hiện trong Bảng 1. Bộ dữ liệu là file excel bao gồm 17.521 dòng và 8 cột ghi lại dữ liệu vận hành của máy nén khí cao áp trong 1 năm. Mỗi dòng tương ứng với 1 mốc thời gian, các cột là giá trị của những cảm biến trên.

Trong nghiên cứu này, nhóm tác giả chọn ra 3 trường hợp đại diện như Bảng 2 để thực hiện dự báo nhiệt độ đầu ra của mỗi cấp, trong đó trường hợp

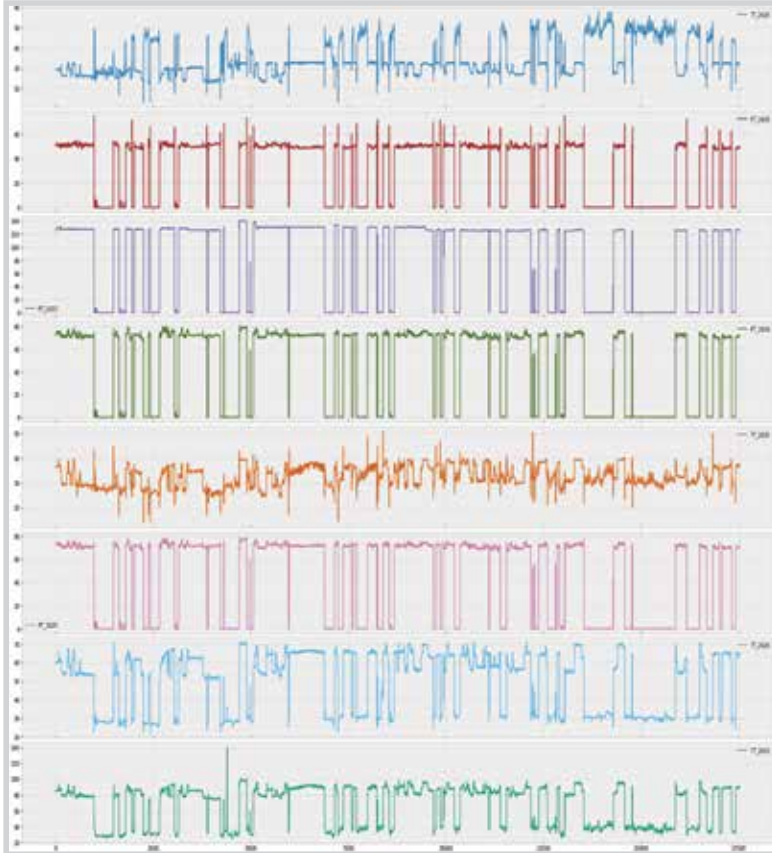


Hình 5. Sơ đồ công nghệ của hệ thống máy nén khí trên giàn PQP-HT.

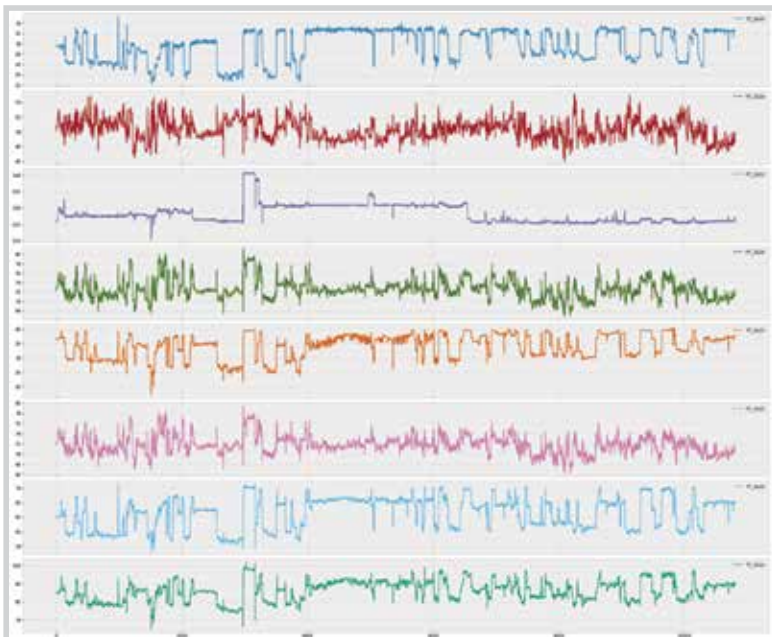
Bảng 2. Bảng tổng hợp các thông số vận hành thực tế của máy nén

Trường hợp	Đầu vào cấp 1		Đầu ra cấp 1		Đầu vào cấp 2		Đầu ra cấp 2	
	TT-2620 (°C)	PT-2620 (barg)	TT-2626 (°C)	PT-2626 (barg)	TT-2625 (°C)	PT-2625 (barg)	TT-2622 (°C)	PT-2622 (barg)
1	27,1	46,4	59,5	67,6	31,2	67,1	87,7	126,2
2	26,7	54,8	53,7	75,6	30,6	74,8	76,8	127,9
3*	41,6	30,0	95,9	54,9	50,0	54,1	151,8	158,0

* Số liệu do nhà thầu máy nén cung cấp



Hình 6. Biểu diễn bộ dữ liệu trên đồ thị với các thuộc tính được liệt kê trong Bảng 2.



Hình 7. Bộ dữ liệu sau khi được làm sạch với các thuộc tính được liệt kê trong Bảng 2.

1 và 2 tương ứng điều kiện vận hành với áp suất đầu vào của cấp 1 đạt giá trị cực tiểu và cực đại, và trường hợp 3 là điểm làm việc mà công ty dự kiến sẽ vận hành trong thời gian tới. Những số liệu của trường hợp 3 là do nhà thầu (vendor) máy nén cung cấp.

3.2. Tiền xử lý dữ liệu

3.2.1. Làm sạch dữ liệu

Quan sát trên đồ thị Hình 6 cho thấy, bộ dữ liệu bao gồm những đoạn đứt quãng có giá trị áp suất khí đầu ra cấp 2 của máy nén PT-2622 giảm từ giá trị vận hành xuống dưới 110 bar, khi đó máy từ chế độ vận hành chuyển về chế độ dừng, khí trong máy đã được xả hết.

Bằng việc loại bỏ những thời điểm máy dừng ra khỏi bộ dữ liệu dựa vào thời điểm áp suất khí đầu ra máy nén cấp 2 dưới 110 bar là áp suất tối thiểu để có thể xuất khí vào đường ống chuyển về bờ, bộ dữ liệu mới (Hình 7) bao gồm 10.808 mẫu - chỉ bao gồm những thời điểm máy hoạt động bình thường.

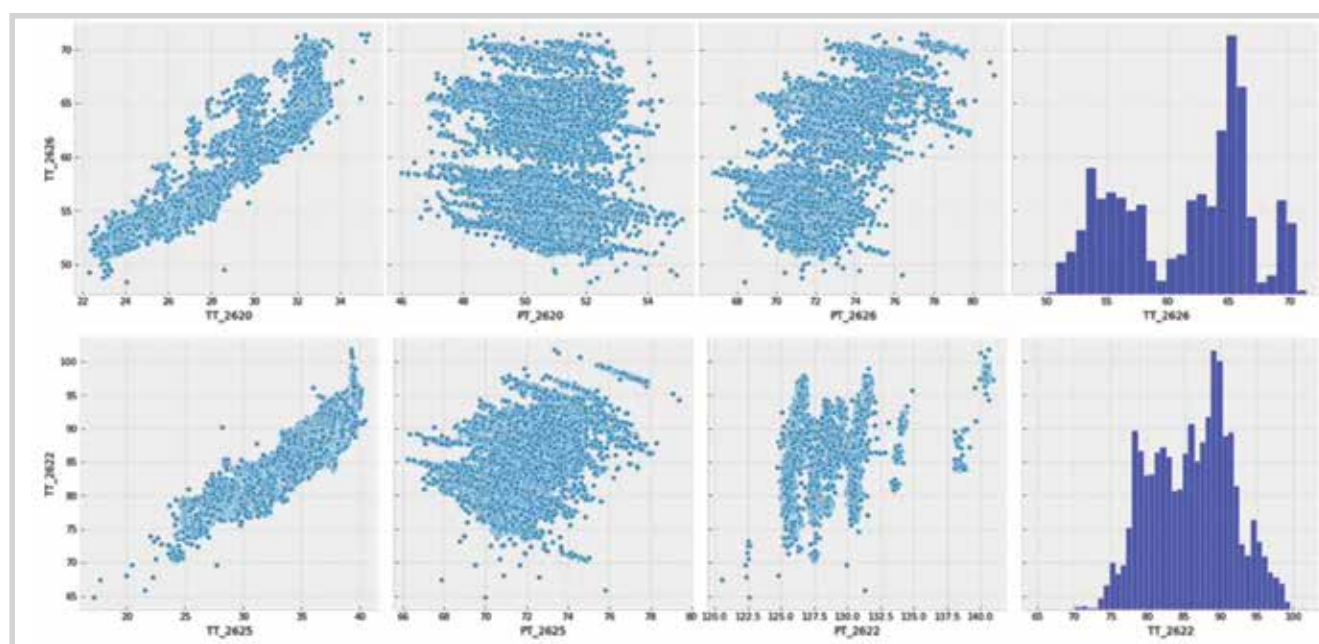
3.2.2. Phân tích dữ liệu

Thống kê cho thấy điều kiện vận hành máy của bộ dữ liệu có áp suất khí đầu vào PT-2620 từ 46 - 55 bar, áp suất khí đầu ra PT-2622 từ 120,5 - 140,9 bar; nhiệt độ khí đầu ra cấp 1 TT-2626 từ 48,3 - 71,4°C, nhiệt độ khí đầu ra cấp 2 TT-2622 từ 64,8 - 101,8°C. Đây là khoảng làm việc mà mô hình sẽ dự báo chính xác nhất.

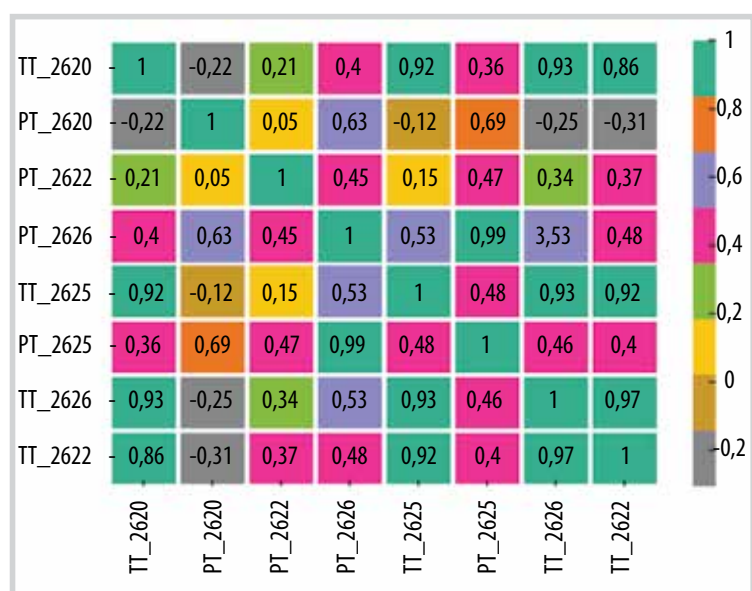
Mối tương quan của nhiệt độ đầu ra tại 2 cấp nén (cấp 1 và cấp 2) được thể hiện trong Hình 9 và 10, cho thấy nhiệt độ đầu vào và đầu ra của máy nén có độ tương quan cao 0,97 là phù hợp vì nhiệt độ này có thể tính toán bằng công thức cũng như bằng phần mềm mô phỏng. Phần sai số của tính toán đến từ áp suất làm việc, thành phần khí, điều kiện làm

	TT_2620	PT_2620	PT_2622	PT_2626	TT_2625	PT_2625	TT_2626	TT_2622
Count	10808.000000	10808.000000	10808.000000	10808.000000	10808.000000	10808.000000	10808.000000	10808.000000
Mean	29.802230	50.262590	128.124816	72.586570	34.044567	71.773163	61.381664	86.012226
Std	3.084856	1.321416	2.867430	2.027851	4.119456	1.746670	5.398998	5.619714
Min	22.308868	46.004211	120.532539	66.825180	17.201925	66.251442	48.371799	64.799156
25%	27.077860	49.284730	125.792988	71.299637	30.650640	70.674622	56.198715	81.474358
50%	30.766187	50.196609	127.366035	72.309273	35.000000	71.626572	62.884621	86.405983
75%	32.610577	51.148556	130.592590	73.466347	37.451923	72.709656	65.576920	90.000008
Max	35.334938	55.094151	140.917603	81.062660	40.432693	79.379936	71.419876	101.790604

Hình 8. Thống kê bộ dữ liệu vận hành.



Hình 9. Phân tích tương quan giữa nhiệt độ đầu ra máy nén và các thông số khác như liệt kê trong Bảng 2.



Hình 10. Phân tích mối tương quan của các thông số bằng bản đồ nhiệt.

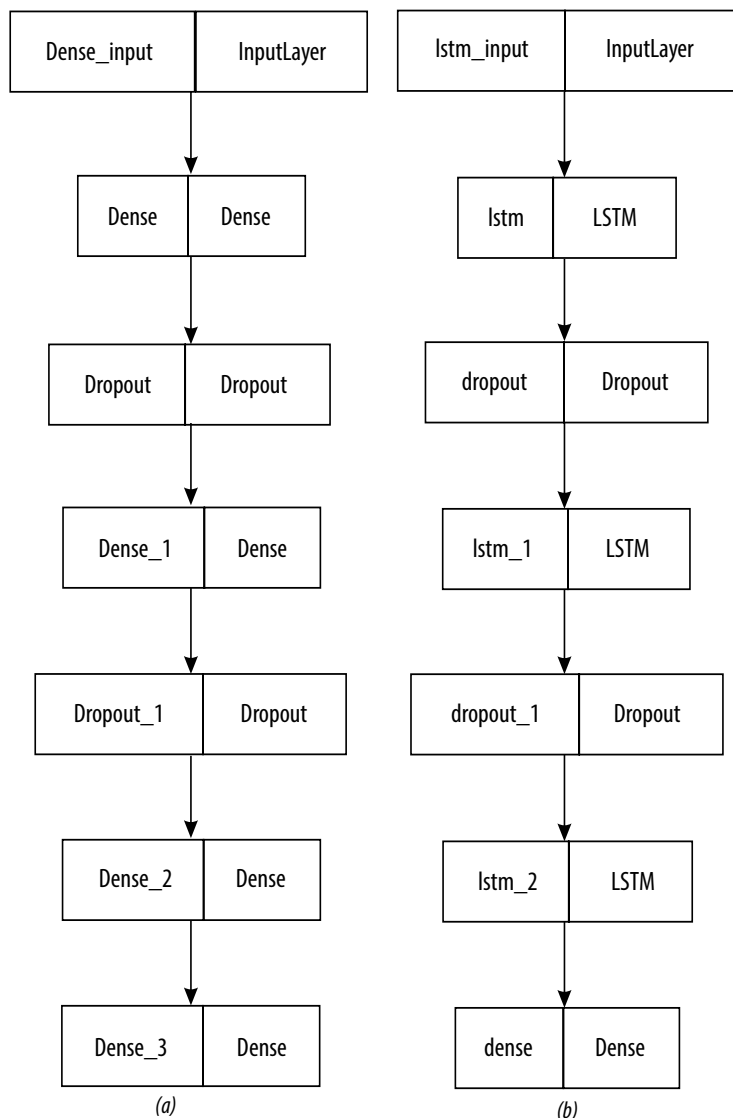
việc của bộ làm mát khí sau mỗi chu trình nén, và có thể áp dụng học máy để tìm ra mối liên hệ đó để dự báo nhiệt độ đầu ra của máy nén dựa trên điều kiện làm việc của hệ thống.

3.2.3. Căn chỉnh dữ liệu

Để giảm sai số cho quá trình huấn luyện và dự báo, dữ liệu được chia tỷ lệ từ 0 - 1 bằng hàm tỷ lệ nhỏ nhất - lớn nhất (min-max scaler).

3.2.4. Chia dữ liệu thành tập đào tạo và tập kiểm tra

Nhóm nghiên cứu đã trộn dữ liệu để tăng độ biến thiên của dữ liệu giúp cho mô hình đào tạo nhanh đạt độ chính xác cao hơn; sau đó, tập dữ liệu được chia thành 1 tập dành cho đào tạo và 1 tập dành cho kiểm tra với tỷ lệ 80/20, nghĩa



Hình 11. Cấu trúc mô hình ANN (a) và LSTM (b).

là 80% các mẫu được sử dụng để đào tạo mô hình, 20% còn lại được sử dụng để kiểm tra.

4. Phương pháp đề xuất

Nhóm nghiên cứu đã sử dụng mô hình ANN và LSTM với cấu trúc được thể hiện trong Hình 11 và các thông số được mô tả như trong Bảng 3 và 4.

5. Kết quả và thảo luận

5.1. Kết quả

Cả 2 mô hình đã dự báo được giá trị nhiệt độ đầu ra của máy nén khí cao áp khi áp dụng tập dữ liệu kiểm tra (Hình 12 - 15).

Hình 12 và 13 là kết quả dự báo của mô hình ANN cho nhiệt độ đầu ra cấp 1 và nhiệt độ đầu ra cấp 2 của máy nén khí cao áp cho 100 điểm đầu tiên trên bộ dữ liệu kiểm tra. Đường dự báo là đường màu xanh trong khi đường màu đỏ là giá trị vận hành. Vẫn có những sai số đáng kể trong quá trình dự báo $RMSE = 0,739$ và nhóm nghiên cứu vẫn đang tìm cách để cải thiện sai số dự báo cho mô hình này.

Hình 14 và 15 thể hiện kết quả dự báo của mô hình LSTM, giá trị dự báo đã khớp hơn với giá trị tập kiểm tra, sai số cũng nhỏ hơn $RMSE = 0,424$.

5.2. Thảo luận

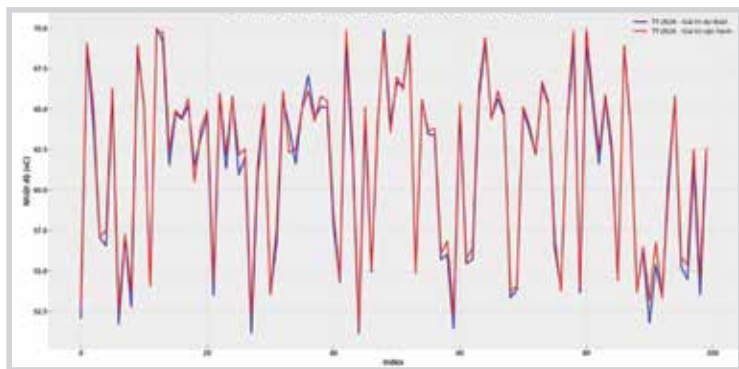
Các tiêu chí đánh giá cho cả 2 mô hình được thể hiện trong Bảng 5 cho thấy mô hình sử dụng

Bảng 3. Số tế bào thần kinh tại mô hình ANN và LSTM

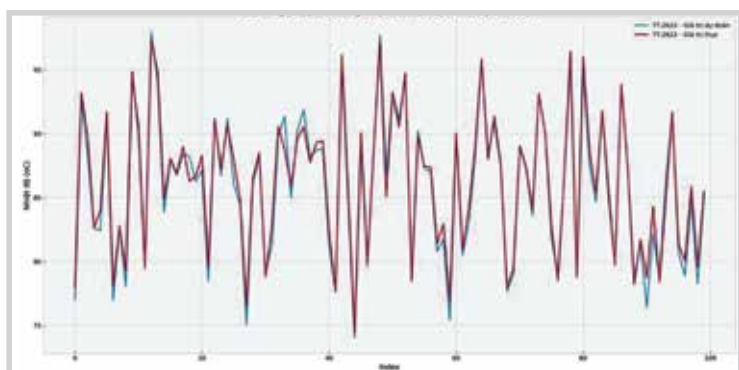
Mô hình	Lớp	Số tế bào thần kinh
ANN	Dense	128
	Dense_1	64
	Dense_2	32
	Dense_3	2
LSTM	LSTM	128
	LSTM_1	64
	LSTM_2	32
	Dense	2

Bảng 4. Các siêu tham số sử dụng để đào tạo mô hình

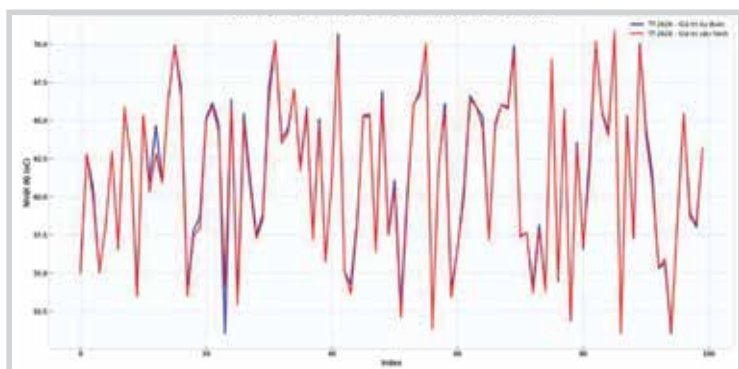
Siêu tham số	Chức năng	Giá trị
Hàm kích hoạt	Chuyển đổi tổng trọng số của đầu vào thành đầu ra từ một hoặc nhiều nút trong lớp mạng	Đơn vị tuyến tính chỉnh lưu (ReLU)
Hàm mất mát	Tìm hiểu các hàm mất mát cung cấp hiệu suất tốt nhất theo chỉ số được xác định trước	MSE
Dropout	Một phần của các đơn vị đầu vào để giảm (hoặc trọng lượng bị đóng băng trong quá trình huấn luyện)	0,2
Kích thước lô và số lượng epoch	Số lượng vòng lặp cho quá trình đào tạo	500
Hàm tối ưu	Tối ưu trong quá trình đào tạo mô hình	Adam



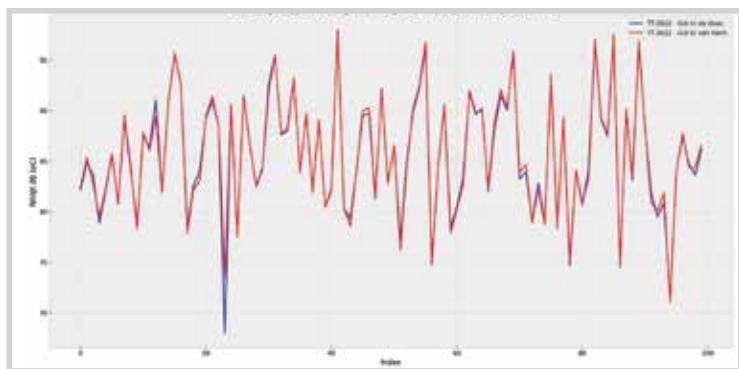
Hình 12. So sánh giá trị dự báo và giá trị thực nhiệt độ đầu ra cấp 1 máy nén bằng mô hình ANN.



Hình 13. So sánh giá trị dự báo và giá trị thực nhiệt độ đầu ra cấp 2 máy nén bằng mô hình ANN.



Hình 14. So sánh giá trị dự báo và giá trị thực nhiệt độ đầu ra cấp 1 máy nén bằng mô hình LSTM.



Hình 15. So sánh giá trị dự báo và giá trị thực nhiệt độ đầu ra cấp 2 máy nén bằng mô hình LSTM.

Bảng 5. So sánh 2 mô hình dự báo

Tiêu chí	ANN	LSTM
Val_loss	0,560	0,140
R2	0,983	0,994
RMSE	0,739	0,4240

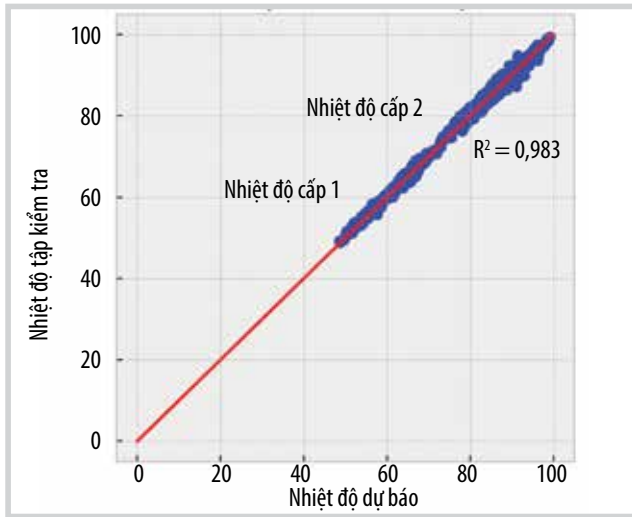
mạng LSTM có kết quả tốt hơn so với mô hình sử dụng mạng ANN. Cụ thể, ở các tiêu chí về độ tương quan R² đạt kết quả cao hơn (0,994 so với 0,983 trên Hình 16 và 17) và sai số RMSE nhỏ hơn (0,424 so với 0,739) thể hiện mô hình LSTM có tính ưu việt hơn mô hình ANN đối với bộ dữ liệu này. Nhóm nghiên cứu sẽ tiếp tục cải thiện mô hình để đạt được kết quả dự báo chính xác hơn.

Bảng 6 so sánh kết quả dự báo nhiệt độ đầu ra cấp 1 và cấp 2 của máy nén bằng thuật toán trí tuệ nhân tạo so với số liệu thực tế (hoặc số liệu do nhà thầu cung cấp) và kết quả tính toán bằng phần mềm mô phỏng Hysys trong cùng một điều kiện vận hành nhiệt độ, áp suất đầu vào và áp suất đầu ra của mỗi cấp.

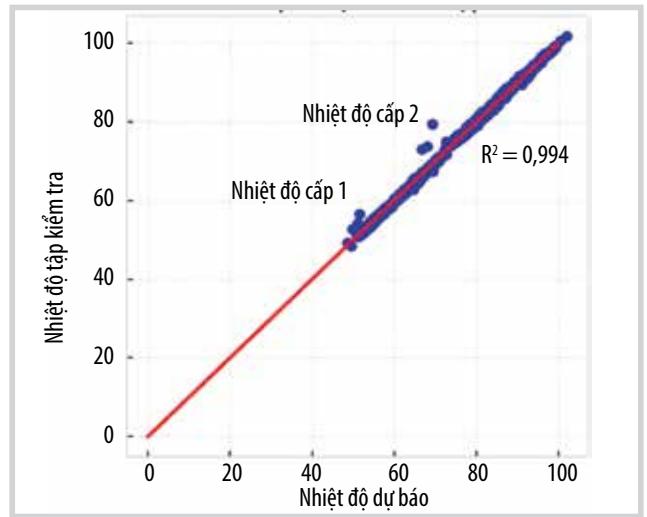
Kết quả cho thấy, đối với những điểm làm việc nằm trong khoảng thông số vận hành của bộ dữ liệu sử dụng để đào tạo cho các thuật toán học máy, cả 2 mô hình sử dụng thuật toán ANN và mô hình sử dụng thuật toán LSTM đều cho kết quả dự báo nhiệt độ đầu ra tại mỗi cấp của máy nén có độ chính xác cao, tương đương với kết quả mô phỏng bằng phần mềm Hysys. Kết quả có độ sai lệch so với số liệu thực tế từ 0,8 - 4,2%.

Đối với trường hợp điểm làm việc nằm ngoài khoảng thông số vận hành của bộ dữ liệu sử dụng để đào tạo cho các học máy (trường hợp 3), cả 2 thuật toán đều cho kết quả có độ sai lệch cao, lên đến 10% so với số liệu thực tế.

Những kết quả trên cho thấy mô hình dựa trên các thuật toán học máy có thể được sử dụng để dự báo một cách chính xác điều kiện vận hành của hệ thống công nghệ với điều kiện điểm làm việc nằm trong khoảng giới hạn của bộ dữ liệu sử dụng cho học máy. Ưu điểm của phương pháp sử dụng các thuật toán trí tuệ nhân tạo là có độ linh động cao, có thể cập nhật cơ sở dữ liệu liên tục theo thời gian thực trong quá trình vận hành; tăng cường độ chính xác cho mô hình dự báo thay vì phải lấy mẫu định kỳ của các mẫu khí để cập nhật sự thay đổi của thành phần và tính chất của các cấu tử; sau đó sử dụng các phần mềm chuyên dụng để tiến hành mô phỏng khi có sự thay đổi điều kiện vận hành với chi phí đắt đỏ. Như vậy, việc sử dụng mô hình dựa trên các thuật toán trí tuệ nhân tạo giúp tiết kiệm được thời gian, chi phí nhân lực, cũng như các chi phí để phân tích các mẫu và thuê phần mềm chuyên dụng.



Hình 16. So sánh kết quả dự báo và kết quả của tập kiểm tra bằng mô hình ANN.

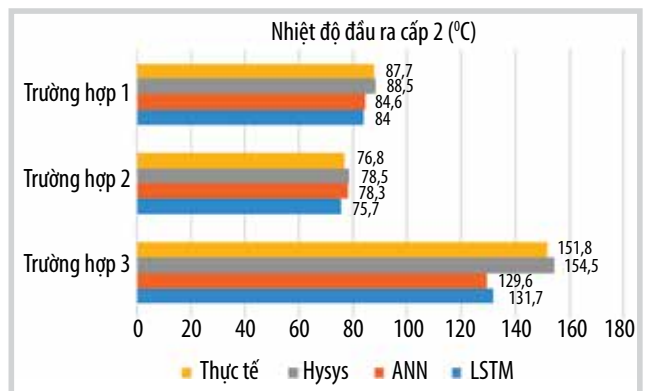
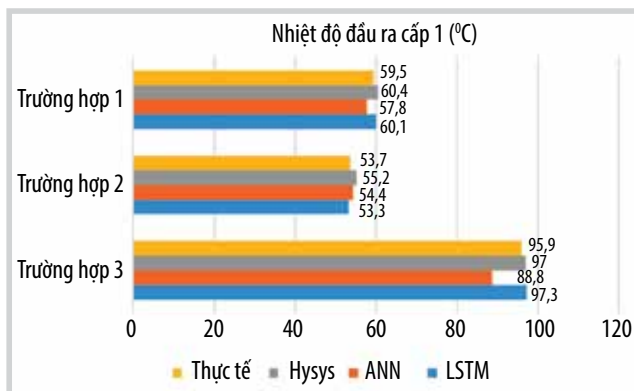


Hình 17. So sánh kết quả dự báo và kết quả của tập kiểm tra bằng mô hình LSTM.

Bảng 6. So sánh kết quả dự báo nhiệt độ đầu ra sử dụng các phương pháp dùng phần mềm mô phỏng và học máy (machine learning, ML)

Trường hợp	Nhiệt độ đầu ra máy nén cấp 1 (°C)				Nhiệt độ đầu ra máy nén cấp 2 (°C)			
	Thực tế	Hysys	ANN	LSTM	Thực tế	Hysys	ANN	LSTM
1	59,5	60,4	57,8	60,1	87,7	88,5	84,6	84,0
2	53,7	55,2	54,4	53,3	76,8	78,5	78,3	75,7
3*	95,9	97,0	88,8	97,3	151,8	154,5	129,6	131,7

*Số liệu do nhà thầu cung cấp và đây là điểm làm việc nằm ngoài bộ dữ liệu cho học máy.



Hình 18. So sánh kết quả dự báo nhiệt độ đầu ra sử dụng các phương pháp dùng phần mềm mô phỏng và ML.

Trong trường hợp điểm làm việc nằm ngoài khoảng thông số vận hành của bộ dữ liệu sử dụng cho học máy, các mô hình cần được tinh chỉnh để kết quả dự báo chính xác hơn.

6. Kết luận

Dự báo nhiệt độ đầu ra tại mỗi cấp của máy nén khí khi điều kiện vận hành thay đổi có ý nghĩa quan trọng trong công tác vận hành và tối ưu hóa hoạt động sản xuất tại giàn xử lý trung tâm PQP-HT. Việc sử dụng các thuật toán học máy cho thấy có nhiều ưu điểm như: xây dựng mô hình đơn giản, tiết kiệm thời gian tính toán, tiết kiệm chi phí thuê phần mềm mô phỏng nhiệt động học thương mại. Kết quả dự báo có độ chính xác cao khi điểm làm việc

nằm trong khoảng giá trị của bộ dữ liệu sử dụng để đào tạo các công cụ học máy.

Với công tác vận hành hàng ngày, việc thay đổi điều kiện vận hành của các thiết bị thường được thực hiện để đáp ứng các nhu cầu khác nhau của sản xuất cũng như tối ưu hóa khả năng thu hồi sản phẩm từ các mỏ. Ứng dụng thuật toán học máy vào xây dựng mô hình dự báo nhiệt độ đầu ra tại mỗi cấp của máy nén giúp dự báo được các sự thay đổi về điều kiện vận hành có thể xảy ra, từ đó có kế hoạch, quy trình vận hành phù hợp nhằm giảm thiểu các sự cố không mong muốn. Với tính linh động cao, thời gian thực hiện dự báo nhanh và chính xác, các mô hình sử dụng thuật toán học máy là lựa chọn phù hợp để hỗ trợ công tác sản xuất hàng ngày.

Tài liệu tham khảo

- [1] John M. Campbell, Larry L. Lilly, and Robert N. Maddox, *Gas conditioning and processing - Volume 2: The equipment modules*, 6th edition. Campbell Petroleum Series, 1984.
- [2] Maurice Stewart Ken Arnold, *Surface production operations: Volume 2: Design of gas-handling systems and facilities*, 2nd edition. Elsevier, 1999.
- [3] John M. Schultz, "The polytropic analysis of centrifugal compressors", *Journal of Engineering for Power*, Volume 84, Issue 1, pp. 69 - 82, 1962. DOI: 10.1115/1.3673381.
- [4] Chunzhi Wei, "Predicting discharge temperature of a centrifugal compressor in a realistic way based on ASME PTC 10", 2017. [Online]. Available: <https://www.semanticscholar.org/paper/Predicting-discharge-temperature-of-a-Centrifugal-a-Wei/ddecfacbc9608bb1fd8ca34203418fd2f11382>.
- [5] Yu Zhang, Miguel Martinez-Garcia, and Anthony Latimer, "Estimating gas turbine compressor discharge temperature using Bayesian neuro-fuzzy modelling", *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 5-8 October 2017. DOI: 10.1109/SMC.2017.8123194.
- [6] Jingzhou Fei, Ningbo Zhao, Yong Shi, Yongming Feng, and Zhongwei Wang, "Compressor performance prediction using a novel feed-forward neural network based on Gaussian kernel function", *Advances in Mechanical Engineering*, Volume 8, Issue 1, 2016. DOI: 10.1177/1687814016628396.
- [7] Hongsheng Jiang, Sujun Dong, Zheng Liu, Yue He, and Fengming Ai, "Performance prediction of the centrifugal compressor based on a limited number of sample data", *Mathematical Problems in Engineering*, 2019. DOI: 10.1155/2019/5954128.
- [8] Xavier Glorot and Yoshua Bengio, "Understanding the difficulty of training deep feedforward neural networks", *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS)*, Sardinia, Italy, 13 - 15 May 2010.
- [9] C. Olah, "Understanding LSTM networks". [Online]. Available: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>.
- [10] Sepp Hochreiter and Jürgen Schmidhuber, "Long short-term memory", *Neural Computation*, Volume 9, Issue 8, pp. 1735 - 1780, 1997. DOI: 10.1162/neco.1997.9.8.1735.

DEVELOPING A MACHINE LEARNING TOOL TO PREDICT DISCHARGE TEMPERATURES OF GAS COMPRESSOR

Tran Ngoc Trung, Nguyen Thanh Trung, Nguyen Duy Minh, Dao Quang Khoa, Tran Vu Tung, Hoang Ky Son

Bien Dong Petroleum Operating Company (Bien Dong POC)

Email: trungtn@biendongpoc.vn

Summary

Gas compressors are important equipment on the central processing platform PQP-HT. After dehydration and ensuring the dew point temperature in accordance with inlet conditions and specifications of the Nam Con Son gas pipeline (NCSP), natural gas is transferred to a gas compression system consisting of 2 compressor lines. Optimizing operating conditions by reducing the inlet pressure of the natural gas processing system is normally used to extend production time of a gas well. However, alterations in inlet operating conditions will directly affect the gas compressor system, potentially causing the discharge temperature to exceed safe operating thresholds.

Commercial thermodynamic simulation software (such as Hysys, Proll) is typically employed to assess the effect of changing gas compressor operating conditions on the outlet temperature of each stage. This allows simulation and selection of optimal working conditions to ensure safety within the natural gas processing system. Nevertheless, the cost of licensing and maintaining commercial software is substantial. Nowadays, machine learning algorithms are proven to be able to predict operating parameters based on historical data. Many studies have been devoted to accurately predicting compressor performance to improve operational efficiency. Machine learning algorithms have the advantage of highly precise prediction results and the ability to operate continuously and re-train automatically upon any operational condition change. Therefore, they can be used as a viable alternative to commercial thermodynamic simulation software.

Key words: Machine learning, gas compressor, discharge temperature prediction.