

XÂY DỰNG MÔ HÌNH TRÍ TUỆ NHÂN TẠO DỰ BÁO PHÁT TÁN PHÓNG XẠ TẠI MỎ ĐỒNG SIN QUYỀN

Vũ Đình Trọng*, Nguyễn Tô Hoài, Phạm Thu Hiền
Trường Đại học Công nghiệp Quảng Ninh

*Email: trongvu@qui.edu.vn

TÓM TẮT

Khí phóng xạ radon là một loại khí phổ biến, là nguyên nhân chủ yếu của hơn 50% các ca nhiễm độc phóng xạ. Mỏ đồng Sin Quyền là một mỏ phát tán phóng xạ trong quá trình khai thác ở Việt Nam. Uranium là một phóng xạ phổ biến trong mỏ, là nguồn chính phát tán khí radon ra ngoài môi trường trong quá trình khai thác. Đánh giá và dự báo lượng radon phát tán là một yêu cầu quan trọng trong quá trình khai thác. Nghiên cứu này phát triển một mạng thần kinh nhân tạo (ANN) dự báo phát tán phóng xạ radon cho mỏ Sin Quyền. Sử dụng hàng triệu điểm dữ liệu để đào tạo, mô hình cho thấy độ chính xác cao cùng với giá trị sai lệch thấp. Các kết quả so sánh với các mô hình học máy phổ biến khác như Random Forest (RF) và Support Vector Machine (SVM) đều cho thấy sự vượt trội của mô hình này. Đồng thời, các phân tích độ nhạy chỉ ra rằng các thông số về khoảng cách và vị trí điểm đánh giá có ảnh hưởng lớn đến kết quả dự báo của mô hình.

Từ khóa: khai thác mỏ, phát tán phóng xạ, radon, uranium, mạng ANN

1. ĐẶT VẤN ĐỀ

Hiện nay, Việt Nam có rất nhiều mỏ khoáng sản chứa các chất phóng xạ đang trong quá trình khai thác, tiềm ẩn nguy cơ cao phát tán các chất phóng xạ ra môi trường xung quanh. Một trong các khí phóng xạ nguy hiểm khi phát tán đó là Radon, là sản phẩm của chuỗi phân rã phóng xạ Urani-238, phát ra tia alpha rất độc hại cho cơ thể con người qua đường hô hấp. Do vậy, việc quan trắc và dự báo phát tán loại khí này ra môi trường xung quanh là một yêu cầu cấp thiết.

Trong những năm gần đây, đã có một số nghiên cứu được tiến hành để quan trắc và dự báo phát tán radon [3, 4, 5, 7] bằng các mô hình toán học hoặc thiết lập các bản đồ phát tán [1, 12, 13]. Tuy nhiên, không có nhiều nghiên cứu sử dụng các thuật toán học máy vốn đã đạt được những tiến bộ quan trọng trong thời gian qua cho công việc này.

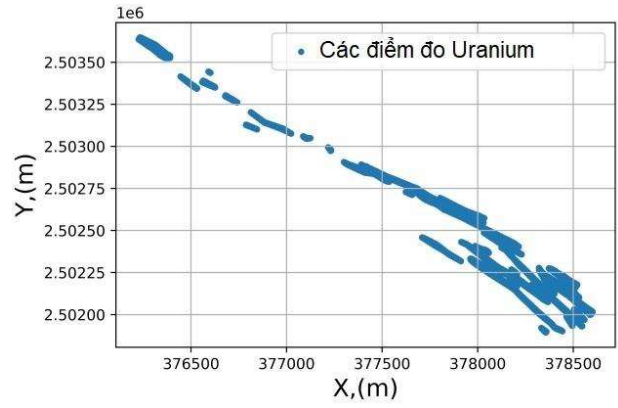
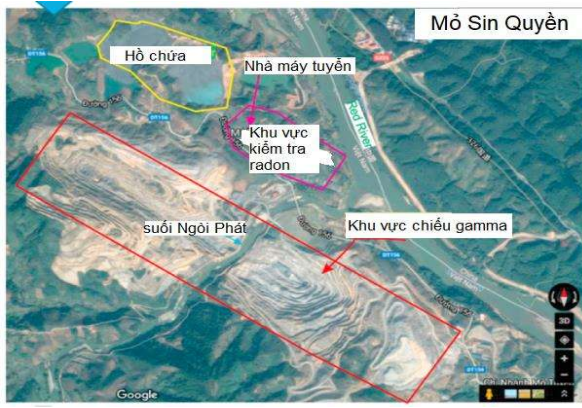
Nghiên cứu này được thực hiện nhằm phát triển một mô hình dự báo phát tán radon dựa trên mạng thần kinh nhân tạo với cấu trúc đơn giản, tiết kiệm chi phí đào tạo và độ sai lệch nhỏ. Mô hình được đào tạo sử dụng dữ liệu lớn thu thập từ mỏ đồng Sin Quyền, bao gồm dữ liệu phổ gamma từ

đồng vị uranium đo được từ mạng lưới quan trắc bên trong mỏ, kết quả đo từ các giếng thăm dò và dữ liệu phát tán radon đo lường thực tế ở các khu vực quanh mỏ. Thêm vào đó, độ nhạy của mô hình cũng được phân tích để đánh giá các yếu tố tác động lên độ chính xác của mô hình

2. DỮ LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Tập dữ liệu

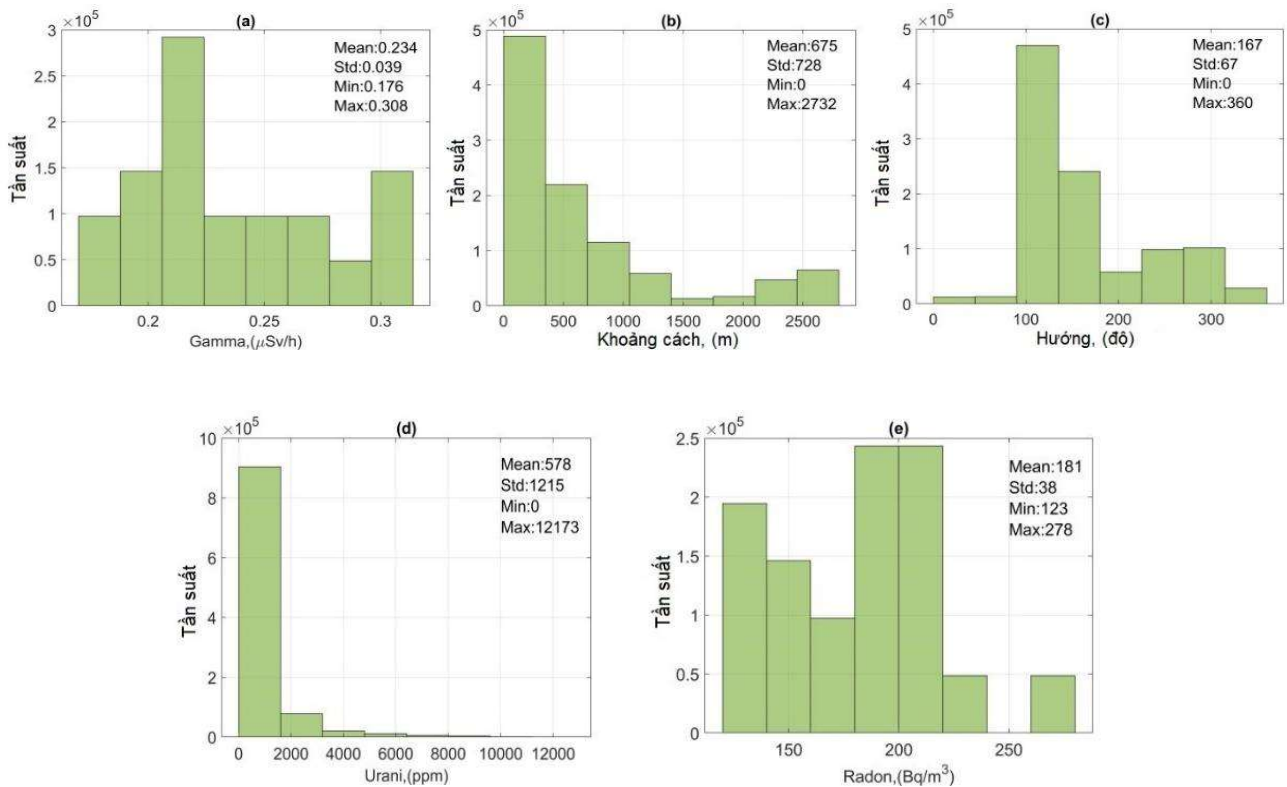
Hơn một triệu dữ liệu sử dụng trong nghiên cứu này bao gồm các giá trị Radon (Bq/m^3) và 6 biến đầu vào của mô hình gồm có tọa độ X, Y (m), liều gamma, ($\mu Sv/h$), khoảng cách (m), hướng (độ) và hàm lượng uranium (ppm). Dữ liệu được thu thập chủ yếu vào 3 năm 2013, 2014 và 2021. Mức độ tập trung uranium ở các khu vực quanh khoáng sàng phản ánh sự phát tán radon từ đất đá và quặng và được đo tại các điểm đo. Dữ liệu uranium được quan trắc bằng máy quang phổ (GammaSurveyor of GF Instruments) trên mắt lưới 3x3m. Khu vực đo giá trị uranium có kích thước 350x1250m trên khu vực khai trường được thể hiện bằng hình chữ nhật màu đỏ như Hình 1. Trong khi đó, lượng radon tích tụ được ghi nhận liên tục trong 3 tháng tại 21 giếng đo quanh khu vực khai thác, sử dụng thiết bị đo CR-39.



Hình 1. (a) Khu vực nghiên cứu and (b) vị trí các điểm đo uranium

Hình 1 biểu diễn khu vực kiểm tra giá trị này bằng đa giác màu hồng. Giá trị liều gamma cũng được quan trắc tại 21 giếng đo radon này. Các

dữ liệu đầu vào khác bao gồm tọa độ (m), khoảng cách (m) và hướng (độ) được ghi nhận bằng thiết bị GPS.



Hình 2. Biểu đồ tần suất (histogram) với các giá trị thống kê cơ bản (mean: trung vị, std: độ lệch chuẩn, min: giá trị nhỏ nhất, max: giá trị lớn nhất) bao gồm các biến gamma (a), khoảng cách (b), hướng (c), giá trị Uranium (d), và Radon (e)

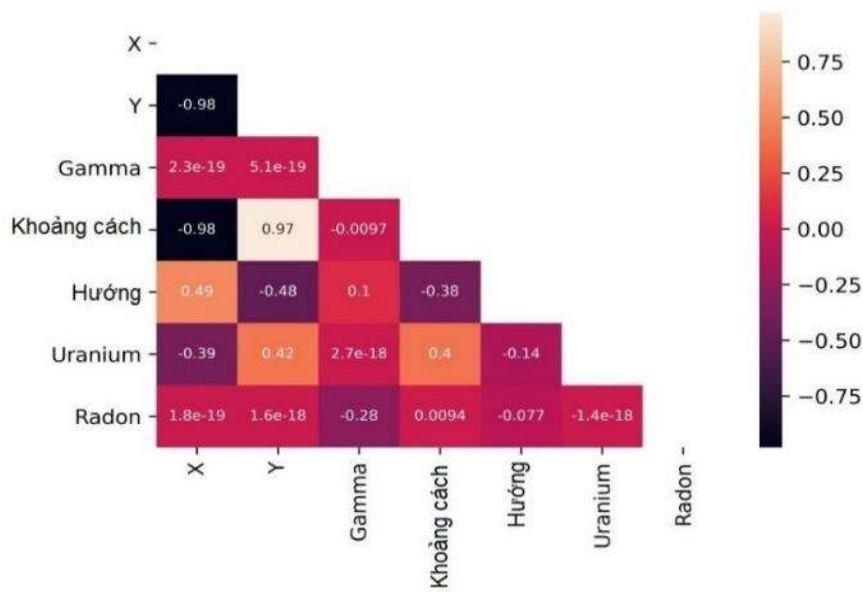
Hình 2 biểu diễn biểu đồ tần suất với các giá trị thống kê của các biến dữ liệu này.

Biểu đồ nhiệt ở Hình 3 được sử dụng để phân tích mối tương quan giữa các biến này. Phần màu đậm cho thấy mức độ tương quan thấp giữa giá trị radon và các biến còn lại.

Do các giá trị này khác nhau về đơn vị cũng như phân bố nên chúng sẽ được chuẩn hóa về khoảng 0 và 1 theo công thức 1.

$$z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (1)$$

Trong đó: $x = (x_1, \dots, x_n)$ là giá trị dữ liệu thu được, z_i là giá trị chuẩn hóa của dữ liệu i_{th} .



Hình 3. Ma trận tương quan giữa các biến dữ liệu

2.2. Mạng thần kinh nhân tạo (ANN)

ANN hoạt động tương tự như não bộ con người, bao gồm hàng tỉ nơ ron sinh học kết nối với nhau theo một cấu trúc phân lớp. Mỗi nơ ron nhận tín hiệu và đồng thời gửi tín hiệu đầu ra cho các nơ ron khác trong mạng. Thông thường, ANN có cấu trúc phức tạp và phân lớp bao gồm một lớp đầu vào và một lớp đầu ra kết hợp với các lớp ẩn nằm ở giữa. Tuy nhiên, ANN với một lớp ẩn duy nhất đã được chứng minh là đủ hiệu quả ở trong rất nhiều các nghiên cứu khác nhau [1]. Trong nghiên cứu này, giá trị radon được coi là một hàm toán học của 6 yếu tố đầu vào khác nhau trong mạng ANN.

Hình 4 minh họa cấu trúc mạng ANN sử dụng trong nghiên cứu này với S là số lượng nơ ron tối ưu trong lớp ẩn. Giá trị S sẽ được xác định trong phần 3.1.

Giả sử rằng véc tơ đầu vào trong mạng ANN ở Hình 4 là P :

$$P = (p_1, p_2, p_3, p_4, p_5, p_6) \quad (2)$$

Mỗi thành phần p_i trong véc tơ P được kết nối với mỗi nơ ron n_j trong lớp ẩn thể hiện qua trọng số w_{ij} . Nơ ron này có giá trị tổng được là các trọng số từ lớp đầu vào cùng với một giá trị thiên vị b_j như trong công thức 3. Tổng giá trị này sẽ được chuyển qua lớp đầu ra với giá trị a_j qua

một hàm chuyển đổi f_h như trong công thức số 4 và 5.

$$s_j = \sum_{i=1}^6 w_{ij} p_i + b_j \quad j=1,2,..S, \quad (3)$$

$$f_h(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}} \quad (4)$$

$$a_j = f_h(s_j) \quad j=1,2,..S, \quad (5)$$

Ở lớp đầu ra, các nơ ron này lại tạo ra giá trị Y_t là kết quả của trọng số w_{jt} và giá trị thiên vị b_t quan hàm chuyển đổi f_o . Bước tính toán này được thực hiện như trong công thức 6, 7 và 8.

$$s_t = \sum_{j=1}^S w_{jt} a_j + b_t \quad t=1, \quad (6)$$

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

$$Y_t = f_o(s_t) \quad t=1, \quad (8)$$

Độ sai lệch giữa giá trị dự đoán (Y_t) và giá trị thật (y_t) của khi đào tạo mô hình được tính toán như trong công thức 9, trong đó T là số lượng dữ liệu đào tạo.

$$MSE = \frac{1}{T} \sum_{t=1}^T (y_t - Y_t)^2 \quad (9)$$

Công thức 10 tính toán độ lệch gradient cho lớp đầu ra.

$$\delta_i = (Y_i - y_i) f'_o(s_i) \quad (10)$$

Thuật toán lan truyền ngược (back-propagation algorithm) được sử dụng để điều chỉnh giá trị trọng số và độ thiên vị trong mạng ANN nhằm tối thiểu hóa giá trị hàm mục tiêu như trong công thức 9. Tương tự, hàm này cũng được sử dụng trong các lớp ẩn và lớp đầu ra như trong công thức 11 và 12.

$$w_{ji}(k+1) = w_{ji}(k) + \alpha (y_i - Y_i) Y_i (1 - Y_i) a_j \quad (11)$$

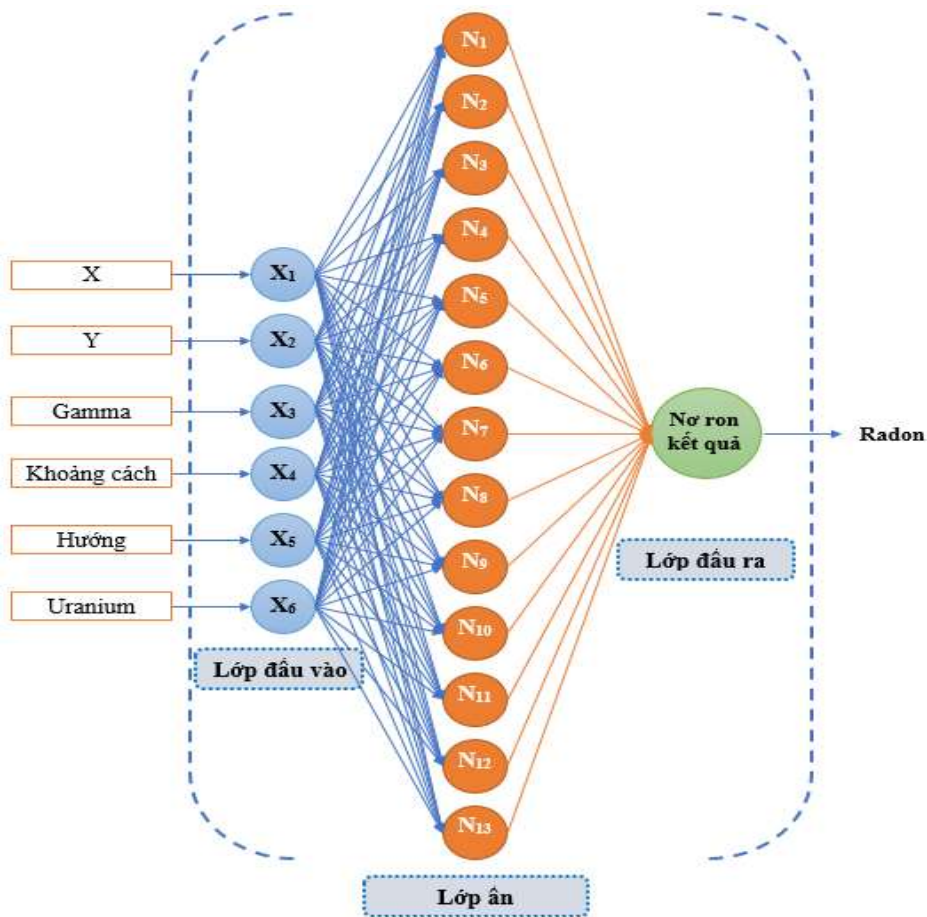
$$b_i(k+1) = b_i(k) + \alpha (y_i - Y_i) Y_i (1 - Y_i) \quad (12)$$

Trong khi đó, công thức 13 và 14 được sử dụng để cập nhật các giá trị trọng số và thiên vị giữa lớp đầu ra và lớp ẩn.

$$w_{ij}(k+1) = w_{ij}(k) + \beta \left[\sum_{l=1}^1 (y_l - Y_l) Y_l (1 - Y_l) w_{jl} \right] a_j (1 - a_j) p_i \quad (13)$$

$$b_j(k+1) = b_j(k) + \beta \left[\sum_{l=1}^1 (y_l - Y_l) Y_l (1 - Y_l) w_{jl} \right] a_j (1 - a_j) \quad (14)$$

trong đó α và β ($0 < \alpha, \beta < 1$) là tốc độ học (learning rate) giữa các lớp và k tương ứng với lần điều chỉnh thứ k . Tốc độ học thể hiện mức độ hội tụ đến kết quả của mạng nơ ron.



Hình 4. Cấu trúc ANN sử dụng trong nghiên cứu

Trong nghiên cứu này, phần mềm Matlab sẽ được sử dụng để đào tạo mạng ANN. Trong vòng lặp đầu tiên (epoch), trọng số (w_{ij}, w_{ji}) và giá trị thiên vị (b_j, b_i) sẽ được cho một giá trị ban đầu ngẫu nhiên. Mạng sẽ được đào tạo sử dụng

phương pháp Levenberg–Marquardt [2] qua nhiều vòng lặp cho đến khi mạng tạo ra được giá trị sai số toàn phương trung bình (MSE) trong công thức 9 đạt được giá trị ổn định.

Bộ chỉ số bao gồm trung bình bình phương sai số $RMSE$ (công thức 15) sai số phần trăm tuyệt đối trung bình $MAPE$ (công thức 16), sai số tuyệt đối trung bình $MABE$ (công thức 17), hệ số tương quan r (công thức 18), hệ số xác định R^2 (công thức 19) được sử dụng để đánh giá và so sánh độ chính xác của mô hình.

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (y_t - Y_t)^2} \quad (15)$$

$$MAPE = \frac{1}{T} \sum_{t=1}^T \frac{|y_t - Y_t|}{y_t} \times 100 \quad (16)$$

$$MABE = \frac{\sum_{t=1}^T |y_t - Y_t|}{T} \quad (17)$$

$$r = \frac{\sum_{t=1}^T (y_t - \bar{y}_t)(Y_t - \bar{Y}_t)}{\sqrt{\sum_{t=1}^T (y_t - \bar{y}_t)^2 \sum_{t=1}^T (Y_t - \bar{Y}_t)^2}} \quad (18)$$

$$R^2 = 1 - \frac{\sum_{t=1}^T (y_t - Y_t)^2}{\sum_{t=1}^T (y_t - \bar{y}_t)^2} \quad (19)$$

Trong đó: y_t và Y_t lần lượt là giá trị radon thật và dự đoán; \bar{y}_t và \bar{Y}_t giá trị radon thật và dự đoán trung bình; và T là số lượng dữ liệu đào tạo.

$RMSE$ cho thấy sự sai khác giữa giá trị radon dự đoán bởi mô hình và giá trị radon thật đo lường được trong thực tế. Một cách tổng quát, giá trị $RMSE$ càng thấp độ chính xác mô hình càng cao. Trong khi đó $MAPE$ là giá trị đo lường thống kê cho thấy mức độ hiệu quả của mô hình. $MAPE$ càng nhỏ cho thấy độ chính xác mô hình càng cao. $MABE$ cho thấy độ sai lệch giữa giá trị radon dự đoán và giá trị thực tế. Hệ số r ($0 \leq r \leq 1$) được sử dụng để đo lường mức độ tương quan giữa giá trị dự báo và thực tế. $r = 1$ cho thấy mối quan hệ tuyến tính giữa giá trị dự báo và giá trị thực tế. Giá trị R^2 cho thấy độ quan trọng của giá trị dự báo sinh ra từ mô hình dự báo.

3. KẾT QUẢ NGHIÊN CỨU VÀ THẢO LUẬN

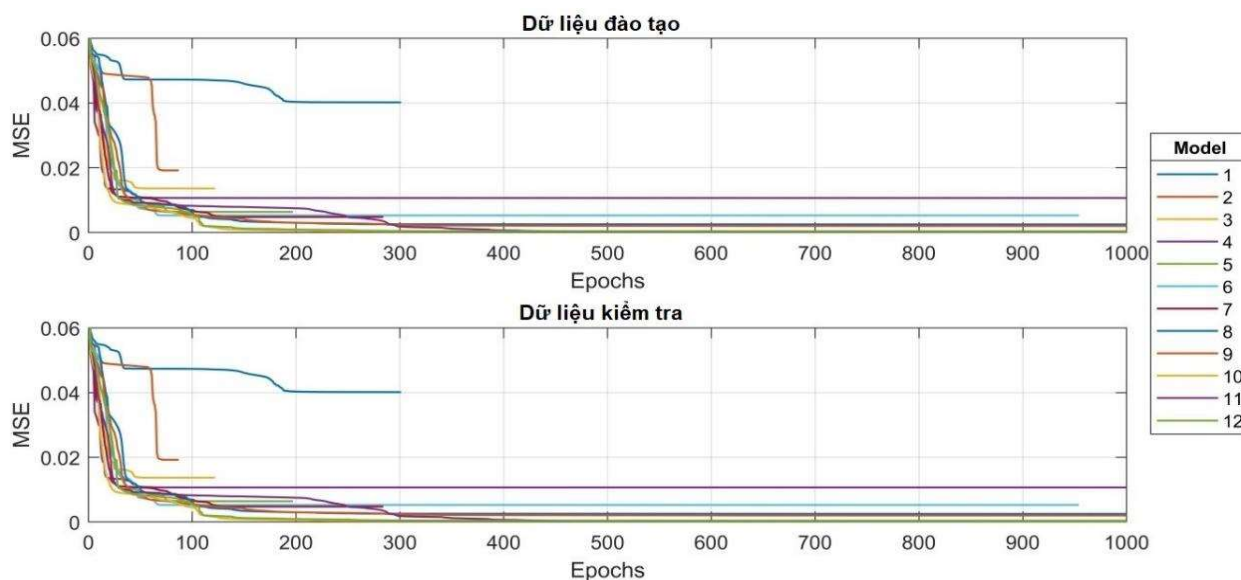
3.1. Cấu hình mạng ANN

Trong nghiên cứu này, mạng ANN chỉ sử dụng một lớp ẩn, như đã thảo luận trong phần 2.2. Số lượng nơ ron S trong lớp ẩn là thông số quan trọng và sẽ được xác định trong phần này. Thực tế, vẫn chưa có công thức hay quy trình nào chính thức xác định giá trị này. Sử dụng định lý Kolmogorov, Hecht-Nielsen [3] cho rằng $2n + 1$ (n là số lượng biến đầu vào dự báo) là số lượng nơ ron lớn nhất mà một lớp ẩn có thể có. Với 6 biến dự báo ($n=6$) sử dụng trong nghiên cứu này thì số lượng nơ ron lớn nhất sẽ là 13 và số lượng nơ ron tối thiểu sẽ được chọn là 2.

Bảng 1. Hiệu quả của các mô hình ANN với số lượng nơ ron khác nhau trong lớp ẩn

Mô hình	Số lượng nơ ron lớp ẩn	$RMSE$ (Bq/m^3)	$MAPE$ (%)	$MABE$ (%)	r	R^2
Đào tạo						
1	2	31.073	12.576	23.553	0.579	0.336
2	3	21.471	9.012	17.023	0.826	0.683
3	4	18.097	7.469	14.084	0.880	0.775
4	5	16.025	6.778	12.792	0.907	0.823
5	6	12.381	4.064	7.435	0.946	0.895
6	7	11.281	3.721	6.921	0.955	0.912
7	8	10.735	3.863	7.200	0.960	0.921
8	9	7.763	3.262	6.137	0.979	0.959
9	10	6.972	2.542	4.971	0.983	0.967

Mô hình	Số lượng nơ ron lớp ẩn	RMSE (Bq/m ³)	MAPE (%)	MABE (%)	r	R ²
10	11	2.959	1.116	2.029	0.997	0.994
11	12	3.027	1.146	2.174	0.997	0.994
12	13	2.793	1.121	2.102	0.997	0.995
Kiểm tra						
1	2	31.048	12.564	23.545	0.580	0.336
2	3	21.487	9.010	17.033	0.826	0.682
3	4	18.139	7.483	14.122	0.880	0.774
4	5	16.006	6.760	12.768	0.908	0.824
5	6	12.395	4.064	7.431	0.946	0.894
6	7	11.254	3.708	6.895	0.955	0.913
7	8	10.689	3.850	7.176	0.960	0.921
8	9	7.742	3.252	6.122	0.979	0.959
9	10	6.938	2.525	4.943	0.983	0.967
10	11	2.971	1.117	2.033	0.997	0.994
11	12	3.020	1.141	2.166	0.997	0.994
12	13	2.791	1.117	2.094	0.997	0.995



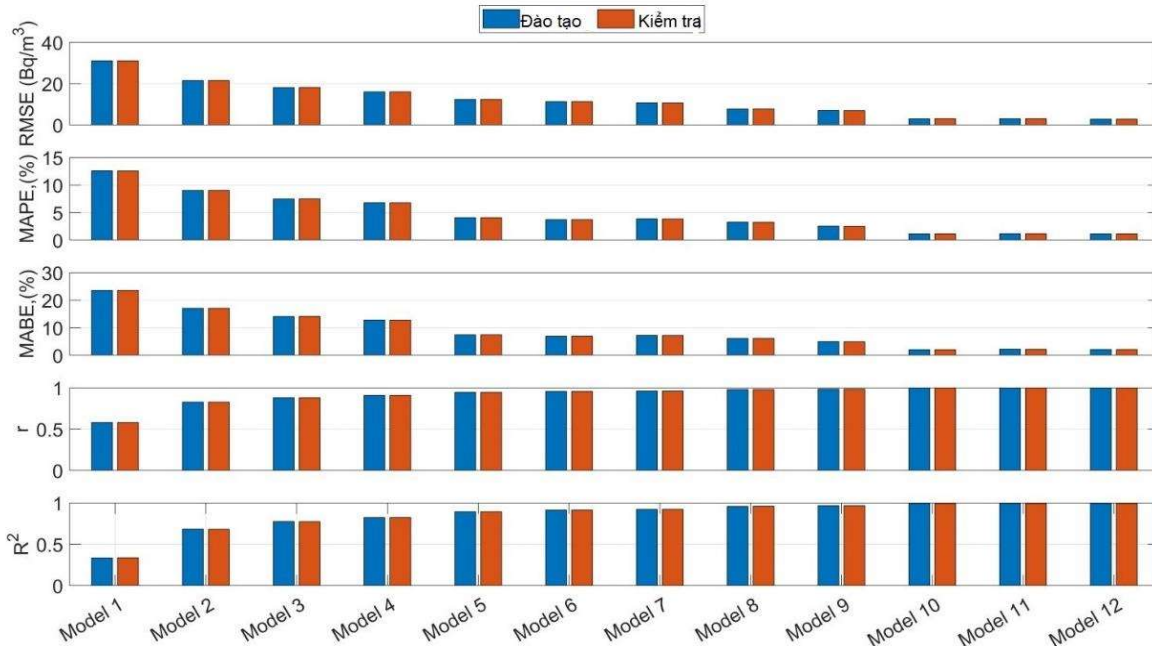
Hình 5. Giá trị MSE của các mô hình ANN khác nhau trên tập dữ liệu đào tạo (training) và kiểm tra (testing) với số epoch

Phần mềm Matlab được sử dụng để khởi tạo các giá trị trọng số và thiên vị ngẫu nhiên cho các cấu hình mạng ANN. Mỗi mô hình ANN sẽ được đào tạo 10 lần để tìm ra số lượng nơ ron tối ưu cho lớp ẩn. Hình 5 biểu diễn giá trị MSE khi đào tạo và đnahs giá với số lượng nơ ron lớp ẩn khác

n nhau. Kết quả cho thấy model 12 và 13 cho giá trị MSE nhỏ nhất ở epoch 300. Nhìn chung, hiệu quả của mạng ANN ở trên tập training và testing là tương đối giống nhau cho thấy mô hình không học thuộc dữ liệu đào tạo và thực sự biểu diễn mối quan hệ thực tế.

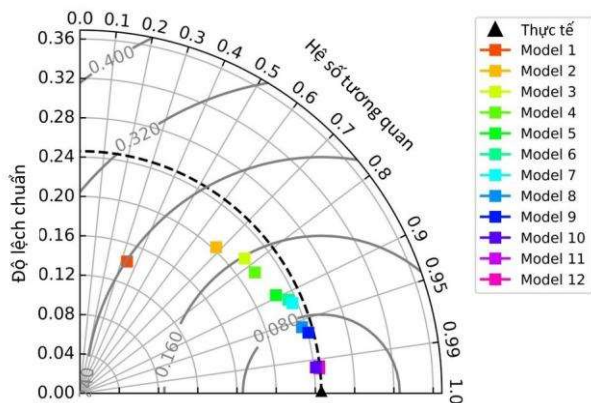
Bảng 1 và Hình 6 cho thấy hiệu quả của mô hình qua các thông số $RMSE$, $MAPE$, $MABE$, r và R^2 . Model 10, 11 và 12 cho thấy hiệu quả cao nhất với giá trị $RMSE < 5$ (Bq/m^3), $MAPE$ và $MABE < 3\%$, và $r, R^2 > 99\%$. Sự khác biệt là

không đáng kể khi so sánh các mô hình này. Tuy nhiên, model 12 cho thấy giá trị R^2 . Các giá trị thống kê $RMSE$, $MAPE$, $MABE$, and r trên dữ liệu kiểm tra lần lượt là 2.791, 1.117, 2.094 and 0.997



Hình 6. Biểu đồ thể hiện hiệu quả của 12 mô hình được đề xuất

Hình 7 so sánh hiệu quả của 12 model sử dụng biểu đồ Taylor [4].



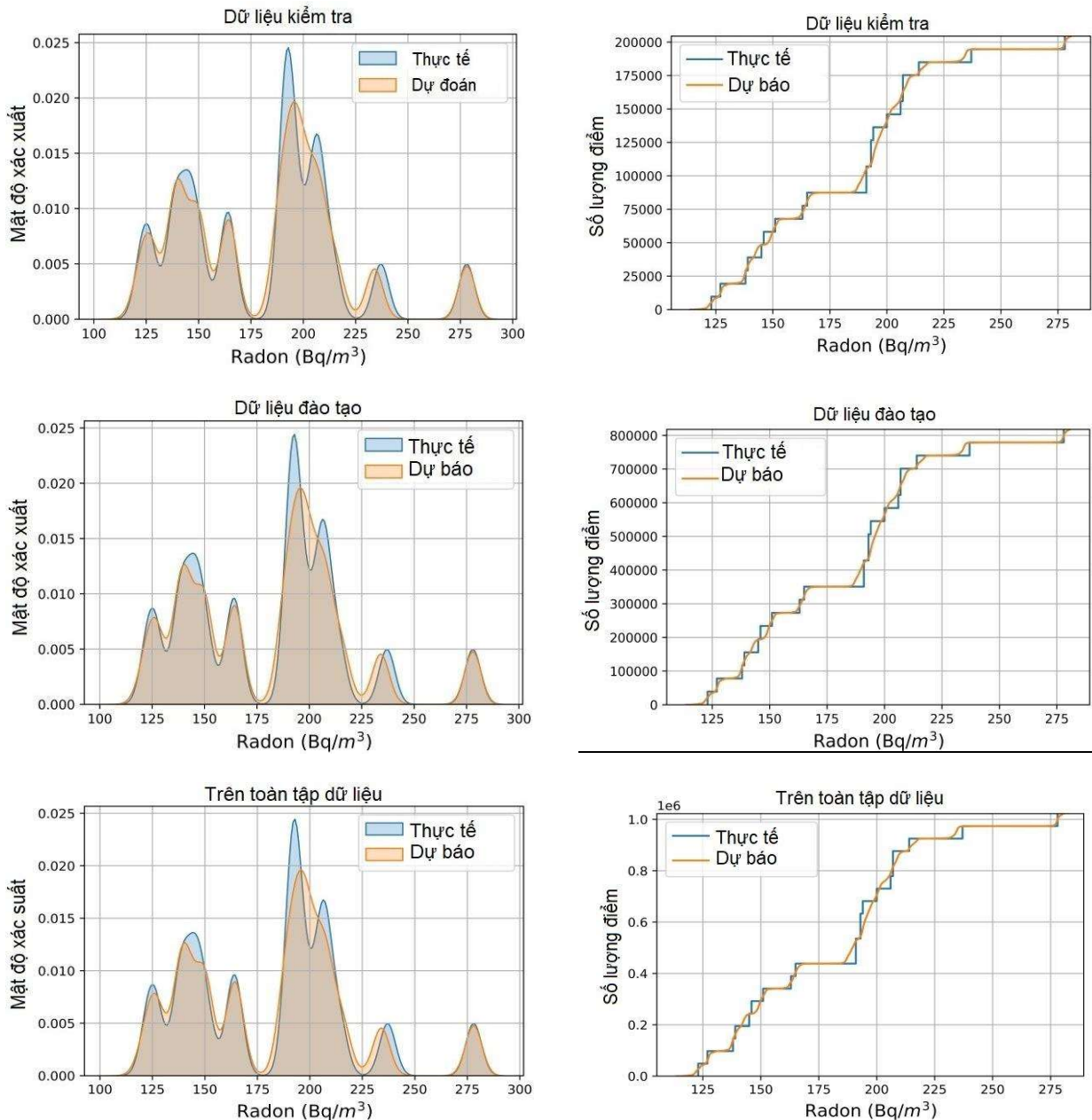
Hình 7. Biểu đồ Taylor so sánh hiệu quả của 12 model

Ưu điểm của loại biểu đồ này là cho thấy được hiệu quả của các model dựa trên các giá trị $RMSE$, độ lệch chuẩn và hệ số tương quan. Khoảng cách giữa model tính từ điểm quan sát (hình tam giác màu đen) được vẽ trên trục hoành cho thấy các model nội suy các giá trị radon gần với thực tế như thế nào. Có thể thấy rằng các model 1-9 có hiệu quả tương đối thấp do nằm xa

điểm quan sát. Trong khi đó, các model 10, 11, và 12 cho thấy hiệu quả tốt hơn với cùng giá trị $RMSE$, độ lệch chuẩn và giá trị tương quan (được minh họa bằng đường nét đứt ở bán kính 0,246). Tuy vậy, model 12 vẫn có giá trị $RMSE$ và độ tương quan cao hơn đôi chút. Chính vì vậy, các kết quả về sau sẽ lựa chọn model 12 để tiếp tục nghiên cứu.

3.2. Hiệu quả của mô hình ANN đề xuất

Hình 8 so sánh biểu đồ mật độ xác suất (hình bên trái) các giá trị radon đo lường được trong thực tế và được ANN dự đoán từ mô hình được lựa chọn trong phần 3.2 và biểu đồ phân bố tích lũy (hình bên phải) cho 3 tập dữ liệu kiểm tra (testing), đào tạo (training) và toàn bộ tập dữ liệu. Kết quả cho thấy mức độ tương quan rất cao giữa giá trị thực tế và giá trị dự báo. Độ lệch chỉ xuất hiện chủ yếu khi model dự báo các giá trị radon dao động ở mức 200 (Bq/m^3). Bên cạnh đó, hình dạng phân bố của kết quả dự báo trên 3 tập dữ liệu là tương đồng nhau cho thấy khả năng dự báo của model với tập dữ liệu mới.



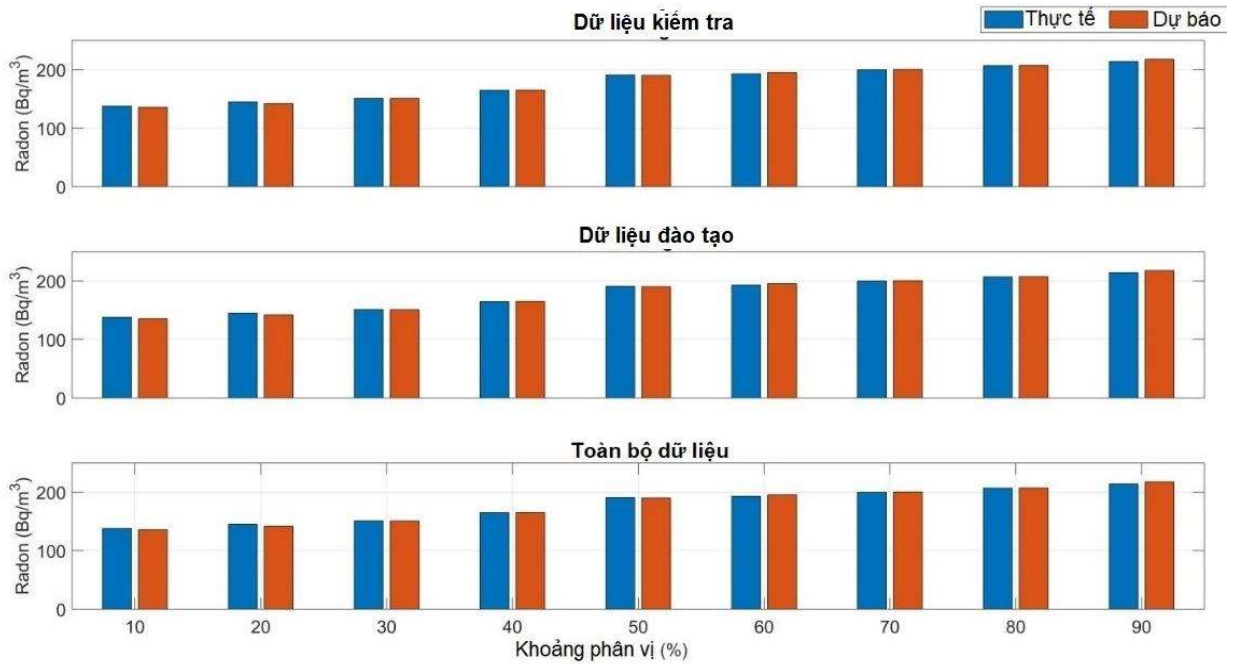
Hình 8. Biểu đồ so sánh phân bố xác suất và tích lũy của dữ liệu thực tế và dự báo trên 3 tập dữ liệu kiểm tra (testing), đào tạo (training) và toàn bộ tập dữ liệu

Độ lệch mô hình trên phân vị (quantile) là một công cụ khác được sử dụng để đánh giá hiệu quả của mô hình trong nghiên cứu này.

Hình 9, các phân vị được chia thành các khoảng bằng nhau từ 10 đến 90%. Nhìn chung, độ lệch trong các khoảng phân vị khi mô hình dự

báo là tương tự nhau cho thấy khả năng dự đoán một cách cục bộ của mô hình.

Bảng 2 đánh giá độ lệch chi tiết trên mỗi phân vị. Độ lệch có giá trị cao nhất nằm ở các phân vị 10, 20, 60 và 90. Bên cạnh đó, độ lệch trung bình trên tập testing có khác biệt rất nhỏ so với trên tập training và toàn bộ tập dữ liệu.



Hình 9. Biểu đồ phân vị so sánh độ lệch giữa giá trị radon dự báo và thực tế trên 3 tập kiểm tra, đào tạo và toàn bộ dữ liệu

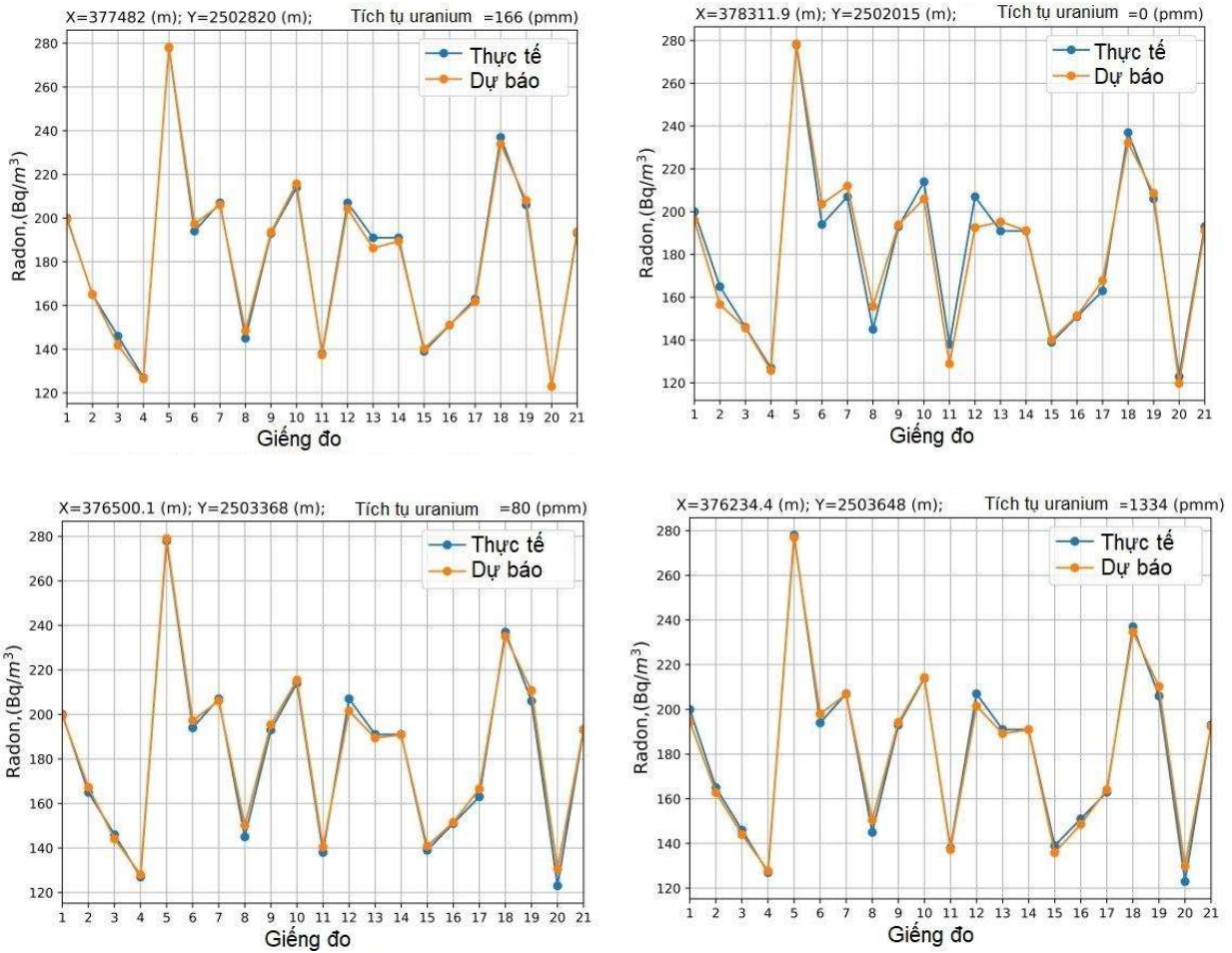
Bảng 2. So sánh độ lệch giữa giá trị radon dự báo và thực tế trên 3 tập kiểm tra, đào tạo và n

Khoảng phân vị (%)	Tập kiểm tra			Tập đào tạo			Toàn bộ dữ liệu		
	M	P	E	M	P	E	M	P	E
10	138	135.930	-2.070	138	135.590	-2.410	138	135.657	-2.343
20	145	142.071	-2.929	145	142.075	-2.925	145	142.074	-2.926
30	151	150.804	-0.196	151	150.675	-0.325	151	150.701	-0.299
40	165	165.290	0.290	165	165.262	0.262	165	165.268	0.268
50	191	190.188	-0.812	191	190.131	-0.869	191	190.143	-0.857
60	193	195.305	2.305	193	195.284	2.284	193	195.288	2.288
70	200	200.513	0.513	200	200.500	0.500	200	200.503	0.503
80	207	207.253	0.253	207	207.221	0.221	207	207.227	0.227
90	214	217.664	3.664	214	217.649	3.649	214	217.652	3.652
	Trung bình		0.113	Trung bình		0.043	Trung bình		0.057

*Ghi chú: M=Thực tế (Bq/m^3); P= Dự báo (Bq/m^3); E= Độ lệch (Bq/m^3).

Hình 10 minh họa khả năng dự báo của mô hình khi so sánh với giá trị đo được thực tế ở 21 giếng. Một cách trực quan, giá trị radon dự báo

khác rất nhỏ so với giá trị đo được ở 21 giếng này. Điều đó cho thấy khả năng dự báo chính xác cao của mô hình.



Hình 10. Dự báo phát tán radon tại 21 giếng

3.3. So sánh với các mô hình dự báo khác

Để chứng minh cho hiệu quả của mô hình ANN trong nghiên cứu này, nhóm nghiên cứu tiến hành so sánh mô hình đề xuất với 2 mô hình phổ biến Support Vector Machine (SVM) và Random Forest (RF) đó trên cùng tập dữ liệu. Thêm vào đó, mô hình cũng được so sánh với mô hình ANN tương tự nhưng với số lớp ẩn tăng lên là 2. Với mô hình SVM, hàm độ đo Gaussian (radial basis kernel function) được sử dụng để đào tạo mô hình cùng với 2 thông số hàm tổn thất (C) và hằng số sigma (δ) cũng được tinh chỉnh để đảm bảo cho độ chính xác của mô hình. Với mô hình RF, kích thước nhỏ nhất của lá (minimum leaf size) (m) và số lượng cây quyết định (nt) được sử dụng

để đánh giá độ chính xác của mô hình. Giá trị tối ưu của các thông số C , δ , m , nt được xác định bằng phương pháp thử và sai (trial and error). Kết quả cho thấy, $C = 51.623$, $\delta = 0.024$ với mô hình SVM và $nt = 800$, $nt = 4$ với mô hình RF.

Bảng 3 cho thấy mô hình đề xuất ANN hiệu quả hơn hẳn các mô hình này với giá trị RMSE, MAPE thấp hơn và r , R^2 cao hơn. Trong khi đó, mô hình SVM đạt độ chính xác kém nhất trên cả hai tập đào tạo và kiểm tra. Có thể thấy rằng, khi số lượng lớp ẩn tăng lên, chẳng hạn như 2 lớp ẩn, độ chính xác của mô hình giảm xuống rõ rệt, đồng thời mô hình bắt đầu học thuộc lòng dữ liệu hơn là suy luận.

Bảng 3. So sánh giữa mô hình đề xuất và các mô hình phổ biến khác

Hiệu quả	Mô hình ANN đề xuất		Mô hình ANN 2 lớp ẩn		SVM		RF	
	Đào tạo	Kiểm tra	Đào tạo	Kiểm tra	Đào tạo	Kiểm tra	Đào tạo	Kiểm tra
RMSE (Bq/m ³)	2.793	2.791	8.696	9.130	12.021	13.112	6.650	9.610
MAPE (%)	1.121	1.117	2.020	2.030	5.286	7.025	2.210	3.220
r	0.997	0.997	0.974	0.970	0.830	0.824	0.987	0.970
R ²	0.995	0.995	0.949	0.942	0.812	0.805	0.974	0.941

3.4. Giải thích mô hình

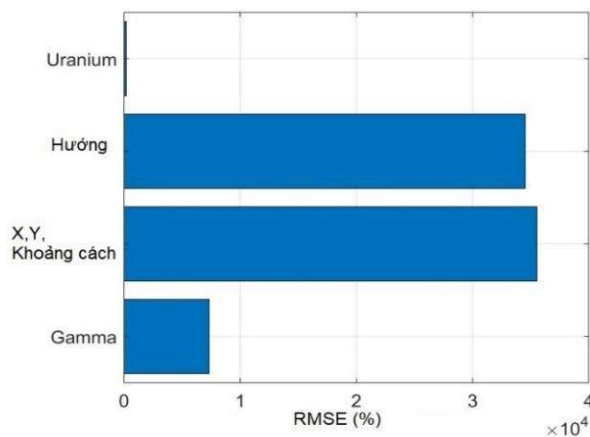
Mô hình ANN đề xuất dự báo giá trị radon dựa trên 5 yếu tố tọa độ X, Y, khoảng cách, hướng, tia gamma, và tích tụ uranium. Tuy nhiên, một câu hỏi cần phải trả lời đó là yếu tố nào ảnh hưởng đến độ chính xác khi dự báo. Để trả lời câu hỏi này, nhóm nghiên cứu sử dụng phương pháp đơn giản và phổ biến đó là Permutation Importance [5]. Phương pháp được thực hiện theo trình tự như sau:

(1) Sử dụng mô hình ANN;

(2) Trộn các giá trị ở trong cùng một yếu tố và tiến hành dự báo trên tập dữ liệu được tạo ra. Giá trị dự đoán và thực tế sẽ được sử dụng để tính toán giá trị tổn thất *RMSE* tạo ra sau khi trộn và đánh giá tầm quan trọng của các biến trộn lẫn.

(3) Chuyển dữ liệu về vị trí ban đầu và lặp lại các bước 2 và 3.

Do sự trộn lẫn là ngẫu nhiên, bước 2 được lặp lại 5 lần để cho ra kết quả trung bình.



Hình 11. Phân tích tầm quan trọng của các yếu tố khi dự báo phát tán radon

Kết quả như trên hình 11 cho thấy các giá trị tọa độ X, Y và khoảng cách có mức độ tương quan cao. Do đó, biến khoảng cách sẽ được sử dụng để làm giá trị đại diện khi giảm tính thiên vị khi xóa trộn dữ liệu.

4. KẾT LUẬN

Radon là một nguyên tố phóng xạ nguy hiểm cho cả con người và môi trường xung quanh. Nghiên cứu này xây dựng một mô hình dự báo sự phát tán radon dựa trên dữ liệu thu thập được từ mỏ đồng Sin Quyền, Lào Cai.

Mô hình sử dụng mạng nơ ron ANN với cấu trúc một lớp ẩn, giúp giảm bớt chi phí đào tạo và tham chiếu mô hình, cho phép dự báo trong thời gian thực. Mô hình đề xuất cũng tránh được tình trạng học thuộc lòng dữ liệu khi các giá trị sai lệch khi đào tạo ($RMSE = 2.793$) và kiểm tra ($RMSE = 2.791$) là không quá khác nhau. Thêm vào đó, mô hình cũng cho thấy hiệu quả cao khi các giá trị độ lệch thấp. $MAPE = 1.121$ (%), $MABE = 2.102$ (%), $r = 0.997$, $R^2 = 0.995$ cho dữ liệu đào tạo and $MAPE = 1.117$ (%), $MABE = 2.094$ (%), $r = 0.997$, $R^2 = 0.995$ cho dữ liệu kiểm tra). Các giá trị tương quan cao và độ lệch thấp khi so sánh giá trị dự báo và giá trị thực tế ở trên các tập dữ liệu kiểm tra, đào tạo và toàn bộ tập dữ liệu đều cho thấy khả năng của mô hình khi dự đoán các dữ liệu mới. Đặc biệt, mô hình còn cho thấy độ chính xác cao hơn khi so sánh với các mô hình truyền thống đã được thể hiện trước đây. Phân tích tầm quan trọng cũng cho thấy yếu tố khoảng cách và tọa độ đóng vai trò quan trọng khi dự báo mức độ phát tán radon.

TÀI LIỆU THAM KHẢO

1. Doering C, McMaster SA, Johansen MP (2018) *Modelling the dispersion of radon-222 from a landform covered by low uranium grade waste rock*. J Environ Radioact 192:498–504.
2. G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Math. Control. signals Syst.*, vol. 2, no. 4, pp. 303–314, 1989.
3. Grant C, Lalor G, Balcázar M (2012) *Radon monitoring in sites of economical importance in Jamaica*. Appl Radiat Isot 71:96–10
4. Hadad K, Doulatdar R, Mehdizadeh S (2007) *Indoor radon monitoring in Northern Iran using passive and active measurements*. J Environ Radioact 95:39–52.
5. Heidary S, Setayeshi S, Ghannadi-Maragheh M, Negarestani A (2011) *Monitoring and measurement of radon activity in a new design of radon calibration chamber*. Radiat Meas 46:694–700.
6. J. J. Moré, "The Levenberg-Marquardt algorithm: implementation and theory," in *Numerical analysis*, Springer, 1978, pp. 105–116.
7. Ramola R, Negi M, Choubey V (2005) Radon and thoron monitoring in the environment of Kumaun Himalayas: survey and outcomes. J Environ Radioact 79:85–92
8. R. Hecht-Nielsen, "Kolmogorov's mapping neural network existence theorem," in *Proceedings of the international conference on Neural Networks*, 1987, vol. 3, pp. 11–14.
9. K. E. Taylor, "Summarizing multiple aspects of model performance in a single diagram," *J. Geophys. Res. Atmos.*, 2001.
11. G. Hooker and L. Mentch, "Please stop permuting features an explanation and alternatives," *arXiv*. 2019.
12. Wu HX, Wei QL, Yang B, Liu QC (2014) Fast prediction method of radon concentration in environment air. Appl Mech Mater 539:819–822.
13. Xie D, Wang H, Kearfott KJ (2012) Modeling and experimental validation of the dispersion of ²²²Rn released from a uranium mine ventilation shaft. Atmos Environ 60:453–459.

Thông tin của tác giả:**TS. Vũ Đình Trọng**

Giảng viên Khoa Mỏ và Công trình, Trường Đại học Công nghiệp Quảng Ninh
Điện thoại: +(84).869.437.970 Email: trongvu@qui.edu.vn

TS. Nguyễn Tô Hoài

Trưởng Bộ môn Kỹ thuật khai thác khoáng sản, Khoa Mỏ và Công trình, Trường Đại học Công nghiệp Quảng Ninh
Điện thoại: +(84).912.298.997 Email: nguyentohoi@qui.edu.vn

ThS. Phạm Thu Hiền

Giảng viên Khoa Mỏ và Công trình, Trường Đại học Công nghiệp Quảng Ninh
Điện thoại: +(84).015.469.068 Email: phamthuhien@qui.edu.vn

DEVELOPMENT OF RADON DISPERSION PREDICTION MODEL BASED ON ARTIFICIAL NEURAL NETWORK FOR SIN QUYEN COPPER MINE

Information about authors:

Vu Dinh Trong, Ph.D., Quang Ninh University of Industry. Email: trongvu@qui.edu.vn

Nguyen To Hoai, Ph.D, Quang Ninh University of Industry.

Pham Thu Hien, M.Eng., Quang Ninh University of Industry.

ABSTRACT:

Radon is one of the most toxic natural radionuclides which occupy more than 50% of natural radiation exposure. The Sinquyen is a natural radioactive bearing mine in the North of Vietnam. The uranium is the dominant radionuclide in this mine, which is the source emitting radon to surrounding when the mine is exploited. The machine learning has been applied in radon prediction still rare and limited. The assessment as well as understanding the radon dispersion released from this mine are important targets for radiation hazard assessment. In this paper, we designed a simple one-hidden layer artificial neural network (ANN) that requires low computation cost to train and reference. Our comparison results suggested that the proposed model outperformed other benchmark methods such as two-hidden-layer ANN, Random Forest (RF), and Support Vector Machine (SVM). The results also revealed that distance and coordinates factors had a greater effect on the prediction of Radon prediction.

Keywords: mining, dispersion, radon, uranium, ANN

REFERENCES

1. Doering C, McMaster SA, Johansen MP (2018) *Modelling the dispersion of radon-222 from a landform covered by low uranium grade waste rock*. J Environ Radioact 192:498–504.
2. G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Math. Control. signals Syst.*, vol. 2, no. 4, pp. 303–314, 1989.
3. Grant C, Lalor G, Balcázar M (2012) *Radon monitoring in sites of economical importance in Jamaica*. Appl Radiat Isot 71:96–10
4. Hadad K, Doulatdar R, Mehdizadeh S (2007) *Indoor radon monitoring in Northern Iran using passive and active measurements*. J Environ Radioact 95:39–52.
5. Heidary S, Setayeshi S, Ghannadi-Maragheh M, Negarestani A (2011) *Monitoring and measurement of radon activity in a new design of radon calibration chamber*. Radiat Meas 46:694–700.
6. J. J. Moré, "The Levenberg-Marquardt algorithm: implementation and theory," in *Numerical analysis*, Springer, 1978, pp. 105–116.
7. Ramola R, Negi M, Choubey V (2005) *Radon and thoron monitoring in the environment of Kumaun Himalayas: survey and outcomes*. J Environ Radioact 79:85–92
8. R. Hecht-Nielsen, "Kolmogorov's mapping neural network existence theorem," in *Proceedings of the international conference on Neural Networks*, 1987, vol. 3, pp. 11–14.
9. K. E. Taylor, "Summarizing multiple aspects of model performance in a single diagram," *J. Geophys. Res. Atmos.*, 2001.
11. G. Hooker and L. Mentch, "Please stop permuting features an explanation and alternatives," *arXiv*. 2019.

12. Wu HX, Wei QL, Yang B, Liu QC (2014) Fast prediction method of radon concentration in environment air. Appl Mech Mater 539:819–822.
13. Xie D, Wang H, Kearfott KJ (2012) Modeling and experimental validation of the dispersion of ^{222}Rn released from a uranium mine ventilation shaft. Atmos Environ 60:453–459.

Ngày nhận bài: 07/4/2024;

Ngày gửi phản biện: 08/4/2024;

Ngày nhận phản biện: 11/6/2024;

Ngày chấp nhận đăng: 12/6/2024.