



(12) BẢN MÔ TẢ SÁNG CHẾ THUỘC BẰNG ĐỘC QUYỀN SÁNG CHẾ

(19) Cộng hòa xã hội chủ nghĩa Việt Nam (VN)
CỤC SỞ HỮU TRÍ TUỆ

(11)



1-0047519

(51)^{2022.01} G06N 3/04; G01T 1/16

(13) B

(21) 1-2023-02444

(22) 13/04/2023

(45) 25/06/2025 447

(43) 25/07/2023 424A

(73) Khoa Các khoa học liên ngành, Đại học Quốc gia Hà Nội (VN)

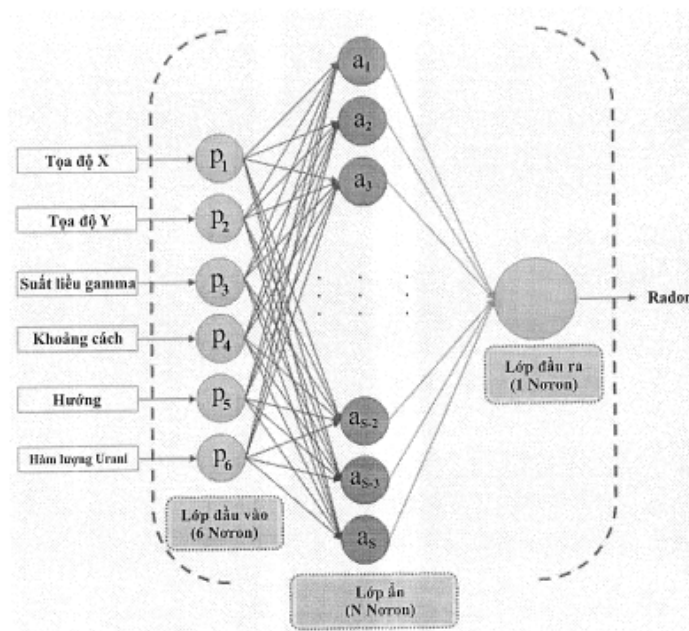
Nhà G7, 144 Xuân Thủy, phường Dịch Vọng Hậu, quận Cầu Giấy, thành phố Hà Nội

(72) Dương Văn Hào (VN); Hoàng Văn Hiệp (VN); Nguyễn Tài Tuệ (VN); Nguyễn Ngọc Trục (VN).

(54) PHƯƠNG PHÁP DỰ BÁO PHÁT TÁN KHÍ PHÓNG XẠ RADON TẠI KHU VỰC
HOẶC MỎ CHỨA CHẤT PHÓNG XẠ

(21) 1-2023-02444

(57) Sáng chế đề cập đến phương pháp dự báo phát tán khí phóng xạ radon (^{222}Rn) tại khu vực hoặc mỏ chứa chất phóng xạ sử dụng mô hình dự báo theo kiến trúc ANN (Artificial Neural Network) gồm có lớp đầu vào, lớp đầu ra, và chỉ có một lớp ẩn, trong đó lớp đầu vào của mô hình dự báo có số lượng nút mạng đầu vào là sáu tương ứng với toàn bộ tập dữ liệu đầu vào bao gồm tọa độ theo trục x, tọa độ theo trục y, suất liều gamma, khoảng cách, hướng, và hàm lượng Urani, lớp đầu ra có một nút mạng tương ứng với kết quả đầu ra là dự báo hoạt độ phóng xạ của khí radon tương ứng với tập dữ liệu đầu vào đã nêu. Việc huấn luyện mô hình dự báo đã nêu sử dụng bộ dữ liệu chuẩn hóa được phân tích và xử lý từ các dữ liệu thực tế thu thập từ các khu vực có mỏ chứa chất phóng xạ, được phân chia thành hai nhóm gồm có các tập dữ liệu huấn luyện và các tập dữ liệu kiểm tra theo tỉ lệ 80:20, để đưa ra bộ các trọng số được tối ưu hoá cho mô hình dự báo. Vị trí cần dự báo phát tán khí phóng xạ radon có thể được xác định nhờ sử dụng hệ thống định vị GPS.



Hình 2

Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến phương pháp dự báo phát tán khí phóng xạ radon (^{222}Rn) tại khu vực hoặc mỏ chứa chất phóng xạ, và cụ thể hơn là phương pháp dự báo phát tán khí phóng xạ radon tại khu vực hoặc mỏ chứa chất phóng xạ bằng mạng nơ ron nhân tạo (ANN - Artificial Neural Network) có một lớp ẩn.

Đối tượng chính được xác định theo sáng chế là ^{222}Rn - khí radon phóng xạ tự nhiên với chu kỳ bán rã 3,82 ngày. Đây là đồng vị ở trạng thái khí phóng xạ tự nhiên có chu kỳ bán rã lớn nhất có khả năng phát tán xa ra môi trường xung quanh khi so sánh với thoron (^{220}Rn) 55 giây, actinon (^{219}Rn) 4 giây. Nó là sản phẩm phân rã của radi (^{226}Ra chu kỳ bán rã 1600 năm), là một thành viên của chuỗi phân rã tự nhiên urani (^{238}U). Khi các khu vực/mỏ/điểm mỏ chứa chất phóng xạ hàm lượng cao được khai thác lộ thiên hay chế biến sẽ gây ra phát tán chất phóng xạ, đặc biệt là khí radon trong quá trình khai thác ra ngoài môi trường xung quanh.

Tình trạng kỹ thuật của sáng chế

Phóng xạ tự nhiên luôn tồn tại ở khắp mọi nơi trong các dạng vật chất khác nhau với hàm lượng/hoạt độ khác nhau. Nhiều khu vực trên thế giới đã và đang chịu tác động phơi nhiễm hoạt độ phóng xạ cao như ở khu vực gần hoặc có các mỏ urani, mỏ đất hiếm, mỏ sa khoáng, mỏ cộng sinh chất phóng xạ, thành hệ granit, khu vực chứa, chế biến, chôn thải chất phóng xạ. Sự phơi nhiễm lâu dài có thể gây ra các tác động đến môi trường, sinh thái, rủi ro tới sức khỏe, thậm chí là tính mạng của con người sống ở khu vực đó. Để có thể xác định hoạt độ, liều lượng bức xạ và sự phân bố nồng độ của nguồn tạo ra bức xạ, có thể sử dụng nhiều kiểu đầu dò và cho đầu dò tương tác với bức xạ. Mỗi trong số nhiều vật liệu của đầu dò lưu trữ một tín hiệu khi nó tiếp xúc với bức xạ. Các tín hiệu là đại diện của bức xạ. Các tín hiệu thu được từ đầu dò sẽ được nhận dạng bằng các công cụ

được thiết kế trong máy dò, các tín hiệu dưới dạng tín hiệu đo/đếm được. Các tín hiệu đo được này được sử dụng để xác định loại và hoạt độ bức xạ cũng như sự phân bố năng lượng của chúng. Hệ thống và phương pháp đo bức xạ tích hợp và thời gian thực này được đề cập trong tài liệu sáng chế Mỹ số US8563947B2 của tác giả Kimberlee J. Kearfott (2013).

Vì các hạt nhân phóng xạ tự nhiên là thành phần của vỏ trái đất nên người ta có thể tìm thấy đồng vị radon ở mọi nơi trong môi trường. Những khu vực có tập trung nồng độ radon cao là trong không khí bề mặt ở các khu vực chứa nồng độ cao các chất phóng xạ như mỏ urani, mỏ đất hiếm, mỏ sa khoáng, mỏ cộng sinh chất phóng xạ, đá granit, mỏ photphat và khu vực chứa, chế biến, chôn thải chất phóng xạ. Nồng độ radon có thể tích tụ trong bầu khí quyển, đặc biệt là trong các không gian kín như hang động, trong nhà; trong đó các mỏ chứa phóng xạ có đóng góp đáng kể. Với đặc tính độc hại tới sức khỏe và chiếm >50% tổng liều phơi chiếu xạ tự nhiên thì khí radon được đặc biệt quan tâm. Để có thể quan trắc và xác định nồng độ radon một số tác giả cùng các nhóm cộng sự như Hadad Doulatdar (2007), Grant Lalor (2012), Marco Laiolo (2012), Ramola (2005), Heidary (2011), Zeeshan Jilani (2017) đã phát triển mô hình phát tán radon và dự đoán dựa trên các điều kiện được coi là dạng mô hình chuẩn hoặc mô hình giả thuyết, lý thuyết hay xây dựng bản đồ tiềm năng radon bằng học máy trí tuệ nhân tạo. Các mô hình này được đánh giá là các mô hình nội suy truyền thống hoặc phát triển theo phương pháp nội suy bằng ron nhân tạo (ANN) với 3 lớp ẩn khác biệt. Trí tuệ nhân tạo với mô hình mạng nơ ron nhân tạo (ANN) được áp dụng trong rất nhiều lĩnh vực và nó cũng không ngoại lệ trong lĩnh vực phóng xạ môi trường. Tuy nhiên các mô hình ANN có nhiều kiểu kiến trúc khác nhau như hai hay nhiều lớp ẩn sẽ phù hợp tối ưu cho từng kiểu bộ dữ liệu khác nhau cũng như từng đối tượng áp dụng sẽ đòi hỏi kiến trúc ANN khác nhau về số lớp ẩn. Không chỉ vậy số nơ ron trong mỗi lớp ẩn cũng có cấu trúc khác nhau đối với từng loại kiểu dữ liệu và đối tượng áp dụng. Chính vì vậy các mô hình truyền thống hay ANN cần đánh giá, thử nghiệm bằng các kỹ thuật chuyên môn để khẳng định tính hiệu quả và mức độ tối ưu để chọn ra mô hình ANN là tốt nhất phù hợp với phương pháp kỹ thuật của đối tượng công việc đặt ra. Có thể liệt kê những nghiên cứu gần đây nhất sử dụng ANN để lập

bản đồ tiềm năng radon liên quan đến đối tượng địa chất (Panahi và cộng sự, 2021; Petermann và cộng sự, 2021; Rezaie và cộng sự, 2021, 2022). Hay một số nghiên cứu đã sử dụng mô hình ANN, máy học, học sâu và cây quyết định chỉ để xác định dị thường radon ở khu vực cần xác định dựa trên tài liệu quan trắc radon theo chuỗi thời gian (Külahcı và cộng sự, 2009; Zmazek và cộng sự, 2003; Mir và cộng sự, 2022).

Một phương pháp khác chỉ sử dụng biến đầu vào tọa độ, giá trị nồng độ radon trung bình có sẵn ở 1261 khu vực hành chính, trên tổng số 1492 khu vực trên khắp Ohio chỉ để dự báo cho những điểm trong khu vực bị thiếu dữ liệu đo đạc radon phong nền và kiểm chứng bằng tài liệu hàm lượng urani (Arjun Akkala và cộng sự 2011, 2012; Pavan Yerrabolu, 2012; Jayaram Gummadi, 2013). Có nghĩa là các phương pháp này chỉ nội suy (dự báo cho các điểm còn thiếu dữ liệu giữa các điểm có dữ liệu chứ không phải dự báo phát tán ra môi trường xung quanh) và sử dụng mô hình mạng nơ ron nhiều lớp ẩn chỉ phù hợp cho trường hợp của mục tiêu và quy mô các nghiên cứu liệt kê ở trên. Về mặt bản chất các phương pháp này không dùng để dự báo phát tán mà chỉ dự báo cho điểm thiếu dữ liệu giữa các điểm không có dữ liệu dựa trên sự phân bố bộ dữ liệu đã có.

Gần đây, phương pháp dự báo phát tán radon từ khu vực chứa chất phóng xạ ra môi trường xung quanh sử dụng mạng nơ ron nhân tạo (ANN) đã được đề xuất và phát triển dựa trên kiến trúc mạng ANN có nhiều lớp ẩn. Mặc dù, việc sử dụng kiến trúc mạng ANN có nhiều lớp ẩn được cho là có thể dễ dàng hơn trong việc tùy biến kiến trúc và tham số mạng để thích ứng với tập dữ liệu huấn luyện, tuy nhiên yêu cầu về thời gian xử lý và sử dụng tài nguyên tính toán của hệ thống máy tính sẽ tương đối lớn. Do đó bộ dữ liệu đưa vào cần phải giảm bớt kích thước và thu nhỏ bộ dữ liệu đại diện dẫn tới bản chất của quy trình phương pháp ANN sẽ đi theo một kết quả có thể hoàn toàn khác hoặc không đạt độ chính xác phù hợp với thực tiễn yêu cầu (Van-Hao và cộng sự 2021).

Nói chung, các phương pháp nêu trên được áp dụng bằng những mô hình lý thuyết truyền thống hoặc áp dụng mô hình ANN cho việc xác định dị thường hay dự báo tiềm năng radon trong các thành hệ địa chất khác nhau, từ đó xây dựng bản đồ tiềm năng radon địa chất, hoặc dự báo các điểm nội suy còn thiếu

kết quả đo giữa các điểm đã có dữ liệu radon bằng kiến trúc ANN nhiều lớp ẩn, hoặc dự báo phát tán radon bằng ANN kiến trúc hai hay nhiều lớp ẩn nhưng với kỹ thuật thu nhỏ kích thước bộ dữ liệu do phải đối mặt với tốc độ/thời gian xử lý khi sử dụng nhiều lớp ẩn.

Việt Nam được đánh giá là một nước có nguồn tài nguyên thiên nhiên phong phú. Việt Nam có hơn 150 mỏ/điểm mỏ giàu nguyên tố phóng xạ như urani, thori; các mỏ này chủ yếu nằm chủ yếu ở khu vực Tây Bắc (điển hình là các mỏ Nậm Xe, Đông Pao, Mường Hum, Sin Quyền, Núi Pháo, Phú Yên...), trung Trung Bộ (các mỏ, urani tại khu vực Pà Lừa - Pà Ròng, Khe Hoa - Khe Cao, Quảng Nam), Tây Nguyên và ven biển từ Quảng Ninh đến Vũng Tàu (có 92 điểm mỏ sa khoáng titan). Ngoài ra, còn nhiều điểm quặng và các thành hệ địa chất chứa hoặc cộng sinh với hàm lượng các nguyên tố phóng xạ cao như: các thể đá granit; các điểm mỏ kim loại như đồng, chì, apatit, bauxit; các mỏ than; các khu vực khai thác chế biến, điểm chứa chất phóng xạ, chất thải phóng xạ. Việc áp dụng phương pháp học máy trí tuệ nhân tạo không những giúp các chủ mỏ khai thác/các công ty chế biến quặng, đơn vị quản lý môi trường và chất thải phóng xạ tại Việt Nam xây dựng bộ cơ sở dữ liệu phóng xạ toàn diện, xây dựng các mô hình phát tán khí phóng xạ có độ xác thực cao, tăng tỷ lệ chính xác và mở rộng phạm vi dự báo đồng thời giảm chi phí thời gian và tiền bạc tiến hành quan trắc, đánh giá. Mà kết quả đạt được còn cung cấp thông tin, dữ liệu đánh giá chính xác khả năng dự báo phát tán khí radon phóng xạ ở khu vực xung quanh các mỏ/điểm mỏ/khu chế biến quặng chứa chất phóng xạ cho các nhà quản lý, hoạch định chính sách bảo vệ môi trường và phát triển kinh tế bền vững của địa phương. Tuy nhiên, hiện nay, theo tìm hiểu của các tác giả, các công trình và các đề án nghiên cứu về môi trường phóng xạ hiện nay mới chỉ sử dụng những phương pháp truyền thống để đo đạc, xác định hay quan trắc các tham số về môi trường phóng xạ. Việc áp dụng ANN trong nghiên cứu phát tán khí phóng xạ ở Việt Nam còn rất hạn chế, đặc biệt là chưa có các nghiên cứu ứng dụng ANN-OHL với kỹ thuật sử dụng toàn bộ dữ liệu không thu nhỏ đại diện trong lĩnh vực giám sát, dự báo phát tán và cảnh báo nguy cơ ô nhiễm phóng xạ khí radon nhằm phát hiện nhanh, xác định nhanh chính xác thông số hoạt độ của chúng.

Do vậy, có nhu cầu về phương pháp được lựa chọn và nghiên cứu tối ưu để áp dụng trong phạm vi các mỏ, điểm mỏ, khu vực khai thác, khu chế biến quặng, địa điểm sử dụng chất phóng xạ, xả thải để dự báo phát tán khí phóng xạ radon tại khu vực hoặc mỏ chứa chất phóng xạ bằng mạng nơ ron nhân tạo một lớp ẩn, đặc biệt là ở các khu vực có mỏ chứa chất phóng xạ ở Việt Nam.

Bản chất kỹ thuật của sáng chế

Mục đích của sáng chế là đề xuất phương pháp dự báo phát tán khí phóng xạ radon tại khu vực hoặc mỏ chứa chất phóng xạ, khắc phục ít nhất là một hoặc một số các vấn đề nêu trên.

Một mục đích khác của sáng chế là đề xuất phương pháp dự báo phát tán khí phóng xạ radon tại khu vực hoặc mỏ chứa chất phóng xạ, sử dụng tập dữ liệu đầu vào có tính đến yếu tố Radon phong nền và đo Radon tích lũy tại vị trí cần dự báo để nâng cao độ chính xác dự báo.

Một mục đích khác của sáng chế là đề xuất phương pháp dự báo phát tán khí phóng xạ radon tại khu vực hoặc mỏ chứa chất phóng xạ, có kiến trúc mạng nơ ron nhân tạo được tối giản hóa một lớp ẩn và sử dụng toàn bộ dữ liệu nhưng vẫn có thể đáp ứng tốt các chỉ tiêu về độ chính xác cao, hiệu quả kinh tế và thời gian.

Cần hiểu rằng sáng chế không bị giới hạn ở các mục đích được nêu ra trên đây, mà có thể còn bao gồm các mục đích khác nữa, trong đó có các mục đích mà có thể được nhận biết bởi người có hiểu biết trung bình về lĩnh vực kỹ thuật tương ứng thông qua các mô tả dưới đây.

Để đạt được ít nhất là một hoặc một số mục đích trên, sáng chế đề xuất phương pháp dự báo phát tán khí phóng xạ radon tại khu vực hoặc mỏ chứa chất phóng xạ, bao gồm các bước:

sử dụng mô hình dự báo theo kiến trúc ANN (Artificial Neural Network) gồm có lớp đầu vào, lớp đầu ra, và chỉ có một lớp ẩn,

trong đó lớp đầu vào của mô hình dự báo toàn bộ dữ liệu lớn, không bị

thu nhỏ, hay lựa chọn đại diện có số lượng nút mạng đầu vào là sáu tương ứng với tập dữ liệu đầu vào bao gồm tọa độ theo trục x, tọa độ theo trục y, suất liều gamma, khoảng cách, hướng, và hàm lượng Urani,

lớp đầu ra có một nút mạng tương ứng với kết quả đầu ra là dự báo hoạt độ phóng xạ của khí radon tương ứng với tập dữ liệu đầu vào đã nêu;

dự báo, bởi mô hình dự báo nêu trên nhờ nhập vào lớp đầu vào của mô hình dự báo tập dữ liệu đầu vào tương ứng với vị trí cần dự báo, để đưa ra kết quả đầu ra tại lớp đầu ra của mô hình dự báo, kết quả đầu ra này được xác định là dự báo hoạt độ phóng xạ của khí radon tại vị trí cần dự báo;

trong đó:

việc huấn luyện mô hình dự báo đã nêu sử dụng bộ dữ liệu chuẩn hóa được phân tích và xử lý từ các dữ liệu thực tế thu thập từ các khu vực có mỏ chứa chất phóng xạ, được phân chia thành hai nhóm gồm có các tập dữ liệu huấn luyện và các tập dữ liệu kiểm tra theo tỉ lệ 80:20, để đưa ra bộ các trọng số được tối ưu hoá cho mô hình dự báo.

Theo một phương án, phương pháp nêu trên sử dụng hệ thống định vị GPS tại vị trí cần dự báo để tọa độ theo trục x, tọa độ theo trục y.

Theo một phương án, lớp ẩn của mô hình dự báo bao gồm từ mười đến mười ba nơ ron.

Tốt hơn là, ít nhất là các tập dữ liệu kiểm tra được sử dụng để đánh giá hiệu suất và độ chính xác của mô hình dự báo được huấn luyện thông qua ít nhất là một trong số các chỉ số đánh giá gồm có sai số gốc bình quân phương (RMSE), sai số phần trăm tuyệt đối trung bình (MAPE), sai số sai lệch tuyệt đối trung bình (MABE), hệ số tương quan (r), và hệ số xác định (R^2), hoặc sự kết hợp của chúng.

Các chỉ số đánh giá đã nêu có thể được xác định thông qua các biểu thức sau đây:

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

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

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

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

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

trong đó: y_t và Y_t lần lượt là báo các giá trị hoạt độ phóng xạ của khí radon tại vị trí cần dự báo đo được và dự đoán; $\bar{y}_t$ và $\bar{Y}_t$ là các giá trị hoạt độ phóng xạ của khí radon trung bình đo được và dự đoán; và T là số lượng các tập dữ liệu huấn luyện.

Tốt hơn là, các dữ liệu thực tế thu thập từ các khu vực có mỏ chứa chất phóng xạ được chuẩn hóa để có giá trị nằm trong khoảng từ 0 đến 1 thông qua biểu thức sau đây:

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

trong đó: $x(x_1, \dots, x_n)$ là dữ liệu thực tế, z_i là dữ liệu được chuẩn hóa thứ i .

Theo một phương án, suất liều gamma có thể được thực hiện thông qua phép đo tích lũy trong một khoảng thời gian xác định trước sao cho lược bỏ được sự ảnh hưởng từ các tham số khí tượng gồm có lượng mưa, nhiệt độ, gió, độ ẩm, và yếu tố địa hình địa mạo.

Mô tả vắn tắt các hình vẽ

Hình 1 là sơ đồ hình minh họa ma trận tương quan của các biến đầu vào và đầu ra theo một phương án ưu tiên của sáng chế;

Hình 2 là sơ đồ khối minh họa kiến trúc mô hình dự báo sử dụng mạng ANN một lớp ẩn có sáu đầu vào và một đầu ra (6-N-1) theo một phương án ưu tiên của sáng chế;

Hình 3 là lưu đồ thể hiện các bước huấn luyện mạng ANN để thực hiện việc dự báo phát tán khí phóng xạ radon tại khu vực hoặc mỏ chứa chất phóng xạ theo một phương án ưu tiên của sáng chế.

Mô tả chi tiết sáng chế

Dưới đây, các ưu điểm, hiệu quả và bản chất của sáng chế có thể được hiểu rõ hơn thông qua việc mô tả chi tiết các phương án ưu tiên thực hiện có dựa vào các hình vẽ kèm theo. Tuy nhiên, cần hiểu rằng các phương án này chỉ được mô tả với mục đích làm ví dụ giúp cho việc hiểu rõ hơn về bản chất và các ưu điểm của sáng chế, mà không giới hạn phạm vi của sáng chế theo các phương án được mô tả này.

Nói chung, phương pháp dự báo phát tán khí phóng xạ radon tại khu vực hoặc mỏ chứa chất phóng xạ theo sáng chế hướng tới việc sử dụng mô hình mạng nơ ron nhân tạo một lớp ẩn (gọi tắt là ANN-OHL) để tối giản hóa kiến trúc của mạng. Mặc dù có kiến trúc mạng tối giản, tuy nhiên mô hình ANN-OHL được đề xuất bởi sáng chế có thể đạt được độ chính xác dự báo tốt khi so sánh với một số mô hình khác, ví dụ mô hình ANN nhiều hơn một lớp ẩn, SVM (Support Vector Machine), RF (Random Forest), chẳng hạn. Điều này sẽ được thể hiện rõ hơn trong bảng kết quả các ví dụ và ví dụ so sánh sẽ được trình bày chi tiết hơn trong phần sau.

Hơn nữa, một phương án ưu tiên của sáng chế đưa ra tập dữ liệu đầu vào bao gồm tọa độ theo trục x, tọa độ theo trục y, suất liều gamma, khoảng cách, hướng, và hàm lượng Urani, tương ứng với lớp đầu vào của mô hình dự báo có số lượng nút mạng đầu vào hay số nơ ron là sáu. Tọa độ theo trục x, tọa độ theo trục y, suất liều gamma, và hàm lượng urani là các giá trị được xác định tại vị trí cần dự báo. Khoảng cách là khoảng cách từ vị trí cần dự báo tới mỏ chứa chất phóng xạ, và hướng là hướng được xác lập giữa vị trí cần dự báo và mỏ chứa

chất phóng xạ.

Dễ thấy là, các dữ liệu đầu vào như tọa độ theo trục x, tọa độ theo trục y, khoảng cách, hướng, và hàm lượng Urani thường được sử dụng cho việc dự báo phát tán khí phóng xạ radon. Ngoài việc sử dụng thêm yếu tố dữ liệu đầu vào là suất liều gamma, việc xử lý chuẩn hoá dữ liệu cũng được cho là góp phần làm tăng độ chính xác dự báo trong khi mô hình dự báo theo sáng chế sử dụng kiến trúc mạng được tối giản hoá chỉ còn một lớp ẩn.

Suất liều tương đương bức xạ có đơn vị là $\mu\text{Sv/h}$, trong đó μSv là giá trị đo suất liều bức xạ. Như vậy suất liều bức xạ thể hiện giá trị liều bức xạ (hay liều tích lũy) nhận được trong một đơn vị thời gian, có thể được đo nhờ sử dụng các thiết bị hoặc phương tiện chuyên dụng, ví dụ các thiết bị dò tìm vết DKS-96 chẳng hạn, phản ánh phần góp liên quan Radon từ phong địa phương. Tức là, khi có sự phát tán khí phóng xạ radon mới từ một nguồn phóng xạ hay từ mỏ chứa chất phóng xạ nào đó, thì giá trị phát tán khí radon được tính là giá trị phát tán khí radon từ nguồn phóng xạ cộng với giá trị radon từ phong địa phương tại nơi đo/khảo sát/dự báo. Hơn nữa, giá trị radon từ phong địa phương cũng được cho là yếu tố phản ảnh kết quả trung bình thực tế của các vụ phát tán trước đó tại khu vực cần dự báo còn sót lại trong môi trường, vì vậy cũng được cho là có thể làm tăng độ chính xác dự báo.

Trong trường hợp phương pháp dự báo phát tán khí phóng xạ radon tại khu vực hoặc mỏ chứa chất phóng xạ theo sáng chế được thực hiện thông qua hệ thống hoặc thiết bị thích hợp, có thể bao gồm các phương tiện đo, phương tiện phân tích, xử lý, tính toán, và các thành phần để xác định được các dữ liệu đầu vào của mô hình dự báo và biến đổi thành dữ liệu đầu ra của mô hình dự báo.

Theo một phương án ưu tiên, phương pháp dự báo phát tán khí phóng xạ radon tại khu vực hoặc mỏ chứa chất phóng xạ theo sáng chế được thực hiện thông qua hệ thống hoặc thiết bị di động được, hệ thống hoặc thiết bị này có thể được trang bị/tích hợp hệ thống định vị GPS để xác định tọa độ theo trục x và tọa độ theo trục y. Điều này có thể làm tăng số điểm đo một cách linh hoạt, giảm

thời gian và thuận tiện trong việc thiết lập phép đo, do chỉ cần bố trí hệ thống hoặc thiết bị di động được đã nêu tại vị trí bất kỳ thích hợp về mặt địa hình mà không cần quan tâm đến vị trí chính xác của nó, do vị trí chính xác có thể được xác định sau đó thông qua hệ thống định vị GPS.

Tiếp theo, việc xử lý/chuẩn hóa các dữ liệu thực tế thu thập từ các khu vực có mỏ chứa chất phóng xạ có thể được thực hiện theo mô tả dưới đây.

Lựa chọn tập dữ liệu

Sự phát tán radon phụ thuộc vào hàm lượng urani, thori có trong quặng và đất đá, đặc điểm nứt nẻ, đập vỡ, thời tiết (mưa, nhiệt độ, hướng gió), đặc điểm địa hình, địa mạo, hướng tới điểm dự báo và cường độ gamma tại vị trí đo. Từ đó, tập dữ liệu được xác định là tọa độ theo trục x, tọa độ theo trục y, suất liều gamma, khoảng cách, hướng, và hàm lượng Urani.

Ưu tiên là, nồng độ radon có thể được thực hiện thông qua phép đo tích lũy trong một khoảng thời gian xác định trước, ví dụ ba tháng, sao cho lược bỏ được sự ảnh hưởng từ các tham số khí tượng gồm có lượng mưa, nhiệt độ, gió, độ ẩm, và yếu tố địa hình địa mạo. Ngoài ra tham số suất liều gamma được lựa chọn là đặc tính kỹ thuật quan trọng của biến đầu vào có tính đại diện cho yếu tố phong địa phương hay cụ thể là phần góp của radon địa phương khi dự báo. Về tổng thể, các dữ liệu có thể được thu thập theo nhiều cách thức khác nhau, ví dụ được cung cấp từ mạng lưới các thiết bị/phương tiện thu thập kèm theo tọa độ và hàm lượng urani.

Theo một ví dụ cụ thể, tọa độ vị trí cần dự báo có thể được xác định thông qua GPS, dữ liệu đo về suất liều gamma (phổ gamma-GF-512) hoặc DKS-96, khí radon và thoron có thể được xác định bằng các thiết bị dò tìm vết CR-39.

Khám phá tập dữ liệu

Các dữ liệu thu thập được nêu trên sẽ được tìm hiểu chi tiết, thống kê và phân tích số lượng đối với từng thuộc tính bằng các kỹ thuật địa thống kê, xác suất, các hàm phân bố để có thể nhận định và đánh giá được kiểu dữ liệu, loại

dữ liệu, chất lượng dữ liệu. Tính chất của dữ liệu radon cần quan tâm đó là xu hướng tập trung, phân tán, xác suất phân bố chuẩn, đặc tính dữ liệu, tính đầy đủ, tính đại diện, đánh giá biến độc lập, mối liên hệ các biến của từng thuộc tính. Tập dữ liệu liên quan đến sự phát tán khí phóng xạ radon được khám phá theo sáng chế là kết quả nghiên cứu, xây dựng các đồ thị, biểu đồ thống kê và ma trận tương quan giữa hoạt độ radon phân tán đối với hàm lượng urani, suất liều gamma, khoảng cách, hướng và hoạt độ radon. Tương tự các tham số đầu vào khác được khám phá trước khi sử dụng.

Tiền xử lý tập dữ liệu

Theo một ví dụ cụ thể việc tiền xử lý tập dữ liệu được thực hiện như sau:

- + tại mỗi cột có một số thuộc tính khác nhau khi chuẩn bị và xử lý tập số liệu, tại đây tiến hành kiểm tra và xử lý dữ liệu thiếu bằng cách loại bỏ chúng theo hàng;

- + chuẩn hoá các dạng khác nhau: chuẩn hoá dữ liệu cho từng thuộc tính, sau đó thống kê lại số lượng dữ liệu của từng thuộc tính; và

- + chuyển đổi về số Label-Encoder: chuyển đổi các nhãn của mô hình học máy về dạng số, ở đây thì dạng cao-thấp, tốt-kém tương đương số lớn-số bé được lựa chọn.

Tiếp theo, sau khi các dữ liệu thực tế thu thập từ các khu vực có mỏ chứa chất phóng xạ được chuẩn hóa thành dạng bộ dữ liệu chuẩn hóa, bộ dữ liệu chuẩn hóa này sẽ được phân chia thành hai nhóm gồm có các tập dữ liệu huấn luyện và các tập dữ liệu kiểm tra theo tỉ lệ 80:20.

Dựa trên hai nhóm gồm có các tập dữ liệu huấn luyện và các tập dữ liệu kiểm tra, việc xây dựng và huấn luyện mô hình mạng nơ ron nhân tạo một lớp ẩn (ANN- OHL) với toàn bộ bộ dữ liệu theo sáng chế sẽ được tiến hành theo các phương pháp đã biết và được sử dụng rộng rãi trong lĩnh vực kỹ thuật học máy. Theo đó mô hình ANN- OHL với nhiều tham số và số lượng nút mạng khác nhau của các lớp ẩn có thể được đưa ra và được tối ưu hóa các tham số thông qua quá trình huấn luyện sử dụng các tập dữ liệu huấn luyện, và sau đó các tập

dữ liệu kiểm tra sẽ được sử dụng để kiểm tra mô hình được huấn luyện theo các tiêu chuẩn được xác định trước, nhằm lựa chọn và đưa ra mô hình được tối ưu hóa mà đáp ứng tốt các tiêu chuẩn được xác định trước.

Để kiểm tra tính hiệu quả của mô hình ANN-OHL, các ví dụ được đưa ra để so sánh với các mô hình khác, ví dụ mô hình ANN hai lớp ẩn, SVM, RF cũng được thực hiện.

Ngoài ra, các kỹ thuật Permutation Importance cũng có thể được sử dụng để đánh giá độ nhạy của các tham số biến đầu vào ảnh hưởng đến kết quả mô hình dự báo. Tuy nhiên sáng chế không bị giới hạn ở đó.

Ví dụ thực hiện sáng chế

Ví dụ thứ nhất

Ví dụ này được thực hiện và trình bày để lựa chọn và đưa ra mô hình ANN-OHL được tối ưu hóa mà đáp ứng tốt các tiêu chuẩn được xác định trước.

Ở đây, việc dự báo phát tán khí phóng xạ radon được thực hiện tại khi vực mỏ Sin Quyền. Mỏ Đồng Sin Quyền là một trong số các điểm mỏ quặng đồng tự nhiên chứa phóng xạ lớn nhất miền Bắc Việt Nam. Urani là hạt nhân phóng xạ chiếm ưu thế trong mỏ này, là nguồn phát ra khí radon ra xung quanh khi mỏ được khai thác. Khu vực này được khai thác từ năm 2006 với hàng triệu tấn quặng đồng mỗi năm.

Bước 1: Lựa chọn dữ liệu

Việc lựa chọn dữ liệu được thực hiện để thu được bộ dữ liệu bao gồm hơn một triệu giá trị được cấu thành từ các dữ liệu bao gồm hoạt độ phóng xạ của hoạt độ phóng xạ của khí radon (Bq/m^3) tương ứng với đầu ra cần dự báo của mô hình và sáu dữ liệu đầu vào là tọa độ theo trục x (m), tọa độ theo trục y (m), suất liều gamma ($\mu\text{Sv/h}$), khoảng cách (m), hướng (độ) và hàm lượng urani (ppm). Vị trí nghiên cứu dữ liệu urani có diện tích 350×1.250 m bên trong điểm khai thác của mỏ Sin Quyền. Phép đo tích lũy CR-39 trong khoảng ba tháng về dữ liệu hoạt động radon ở 21 ngôi nhà xung quanh mỏ đã được thu thập, liệu

gamma cũng được thu thập ở khu vực này, điều này liên quan đến sự phát xạ radon cục bộ tại điểm mục tiêu radon sự dự đoán. Quy mô và phân phối dữ liệu đầu vào có sự khác biệt lẫn nhau. Do đó dữ liệu được chia tỉ lệ trong phạm vi từ 0 đến 1 bằng cách sử dụng phương trình:

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

Trong đó: x ($x_1, \dots, x_n$) là tập giá trị đo được, z_i là dữ liệu chuẩn hóa thứ i .

Tiến hành xây dựng biểu đồ của các giá trị thống kê của các biến đầu vào.

Như được minh họa trên Hình 1, mối quan hệ tuyến tính giữa từng cặp biến đầu vào và đầu ra được thực hiện bằng bản đồ. Màu càng đậm chứng tỏ mối tương quan nghịch giữa các biến càng lớn và ngược lại.

Bước 2: Khám phá tập dữ liệu

Tìm hiểu chi tiết, thống kê số lượng của toàn bộ tập dữ liệu của từng thuộc tính bằng các kỹ thuật địa thống kê, xác xuất, các hàm phân bố để có thể nhận định và đánh giá được kiểu dữ liệu, loại dữ liệu, chất lượng dữ liệu. Tính chất của dữ liệu radon cần quan tâm đó là xu hướng tập trung, phân tán, xác suất phân bố chuẩn, đặc tính dữ liệu, tính đầy đủ, tính đại diện, đánh giá biến độc lập, mối liên hệ các biến của từng thuộc tính. Tập dữ liệu của radon được khám phá ở đây là kết quả nghiên cứu, xây dựng các đồ thị, biểu đồ thống kê và ma trận tương quan giữa hoạt độ radon phân tán đối với hàm lượng urani, suất liều gamma, khoảng cách, hướng và hoạt độ radon cho toàn bộ tập dữ liệu đã được lựa chọn và tương tự với các tham số đầu vào khác.

Bước 3: Tiền xử lý tập dữ liệu dựa theo các công việc:

+ tại mỗi cột có một số thuộc tính khác nhau khi chuẩn bị và xử lý tập số liệu, tại đây tiến hành kiểm tra và xử lý dữ liệu thiếu bằng cách loại bỏ chúng theo hàng;

+ chuẩn hoá các dạng khác nhau: chuẩn hoá dữ liệu cho từng thuộc tính, sau đó thống kê lại số lượng dữ liệu của từng thuộc tính; và

+ chuyển đổi về số Label-Encoder: chuyển đổi các nhãn của mô hình học

máy về dạng số, ở đây thì dạng cao-thấp, tốt-kém tương đương số lớn-số bé được lựa chọn.

Bước 4: Chuẩn bị dữ liệu huấn luyện và kiểm thử mô hình dự báo (train-test theo tỉ lệ 80:20) thông qua phân chia các tập dữ liệu huấn luyện/các tập dữ liệu kiểm tra và xác định rõ các dữ liệu đầu vào, dữ liệu đầu ra và theo nhãn của các dữ liệu được chuẩn hóa.

Kiến trúc mạng ANN được đề xuất với giả định số lượng nơ ron tối ưu trong lớp ẩn là S. Véc tơ đầu vào của mạng là P. Sáng chế sử dụng sáu đầu vào, do đó P được xác định theo phương trình sau đây:

$$P = (p_1, p_2, p_3, p_4, p_5, p_6) \quad (\text{phương trình 2}).$$

Như được minh họa trên Hình 2, mỗi phần tử p_i của véc tơ đầu vào P được kết nối với mỗi nơ ron n_j trong lớp ẩn thông qua trọng số w_{ij} . Tế bào thần kinh này có một tổng tập hợp các trọng số của nó từ lớp đầu vào và độ lệch b_j như thể hiện trong phương trình 3.

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

Sau đó, tổng được truyền qua một hàm chuyển f_h để tạo ra đầu ra a_j tương ứng được đưa ra trong phương trình 4 và phương trình 5.

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

$$a_j = f_h(s_j) \quad (\text{phương trình 5})$$

Trong lớp đầu ra, nơ ron đầu ra lại tạo ra một sản phẩm Y_t là đầu ra dựa trên trọng số được kết nối với trọng số w_{jt} và độ lệch b_t đi qua hàm truyền f_o . Phương trình 6,7 và 8 trình bày biểu thị điều này.

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

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

$$Y_t = f_o(s_t) \quad t=1, \quad (\text{phương trình 8})$$

Việc huấn luyện lỗi giữa đầu ra dự đoán (Y_t) và dữ liệu đo được (y_t) tuân theo phương trình 9, với T là số lượng dữ liệu huấn luyện.

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

Độ dốc lỗi cho lớp đầu ra được đo bằng phương trình:

$$\delta_t = (Y_t - y_t) f'_o(s_t) \quad (\text{phương trình 10})$$

Thuật toán lan truyền ngược được sử dụng để điều chỉnh các trọng số và độ lệch của ANN giúp giảm thiểu hàm mục tiêu trong phương trình 9. Các chức năng điều chỉnh trọng số và độ lệch giữa lớp ẩn và lớp đầu ra được đưa ra bởi phương trình 11 và 12 tương ứng.

$$w_{jt}(k+1) = w_{jt}(k) + \alpha (y_t - Y_t) Y_t (1 - Y_t) a_j \quad (\text{phương trình 11})$$

$$b_t(k+1) = b_t(k) + \alpha (y_t - Y_t) Y_t (1 - Y_t) \quad (\text{phương trình 12})$$

Trong khi đó, phương trình 13 và 14 được sử dụng để cập nhật trọng số và độ chệch tương ứng giữa lớp đầu vào và lớp ẩn.

$$w_{ij}(k+1) = w_{ij}(k) + \beta \left[\sum_{t=1}^T (y_t - Y_t) Y_t (1 - Y_t) w_{jt} \right] a_j (1 - a_j) p_i \quad (\text{phương trình 13})$$

$$b_j(k+1) = b_j(k) + \beta \left[\sum_{t=1}^T (y_t - Y_t) Y_t (1 - Y_t) w_{jt} \right] a_j (1 - a_j) \quad (\text{phương trình 14})$$

α và β ($0 < \alpha, \beta < 1$) là tốc độ học giữa các lớp và k hằng số đối với lần điều chỉnh thứ k . Tỷ lệ học đại diện cho tốc độ hội tụ của mạng.

Dựa trên định lý Kolmogorov, Hecht-Nielsen, sáng chế đề xuất rằng $2n + 1$ (n là số lượng tham số dự đoán) phải là số lượng nơ ron tối đa trong mạng nơ ron một lớp ẩn. Theo gợi ý này, số lượng nơ ron ẩn tối đa theo sáng chế có thể là 13 ($n = 6$), và số lượng nơ ron ẩn tối thiểu được đặt là 2 (Bảng 1). Tuy nhiên sáng chế không bị giới hạn ở đó.

Như được minh họa trên Hình 3, dựa trên cơ sở bộ dữ liệu đã được xây dựng ở bước trước và cơ sở của phương pháp tiến hành chia các tập dữ liệu huấn luyện/các tập dữ liệu kiểm tra theo tỉ lệ 80:20.

Bước 5: Xây dựng mô hình mạng nơ ron nhân tạo một lớp ẩn.

Phần mềm Matlab đã được triển khai để thiết lập mô hình ANN đề xuất. Trong giai đoạn đầu tiên của quá trình huấn luyện, các trọng số (w_{ij} , w_{jt}) và độ chệch (b_j , b_t) được khởi tạo ngẫu nhiên. Các mạng đã được huấn luyện bằng phương pháp Levenberg - Marquardt trong nhiều chu kỳ cho đến khi mạng đạt đến giá trị MSE ổn định (phương trình 9).

Để định lượng và so sánh độ chính xác của các mô hình được đề xuất, sáng chế đã sử dụng 5 chỉ số phổ biến, bao gồm RMSE (phương trình 15), sai số phần trăm tuyệt đối trung bình MAPE (phương trình 16), sai số sai lệch tuyệt đối trung bình MABE (phương trình 17), hệ số tương quan r (phương trình 18), hệ số xác định R^2 (phương trình 19).

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

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

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

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

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

trong đó: y_t và Y_t lần lượt là các giá trị radon đo được và dự đoán; và $\bar{y}_t$ and $\bar{Y}_t$ là các giá trị radon trung bình đo được và dự đoán; và T là số dữ liệu huấn luyện.

Kết quả thu được trong Bảng 1 dưới đây.

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

Mô hình	Số nơ ron ẩn	RMSE (Bq/m ³)	MAPE (%)	MABE (%)	r	R2
Huấn luyện						
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
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

RMSE thể hiện sự khác biệt giữa các giá trị radon được dự đoán bởi các mô hình và các giá trị radon đo được thực. Nói chung, RMSE thấp tốt hơn RMSE cao. Trong khi đó, MAPE là thước đo thống kê về mức độ chính xác của các mô hình ANN được đề xuất. Giá trị tối thiểu cho thấy mô hình có độ chính xác cao. MABE là một số liệu khác để ước tính mức độ gần đúng của các giá trị radon dự đoán với các giá trị radon đo được. MABE thấp có nghĩa là độ chính xác của mô hình cao. Hệ số tương quan r ($0 \leq r \leq 1$) dùng để lượng hóa mối tương quan giữa mô hình và các quan sát, $r = 1$ biểu thị mối quan hệ tuyến tính chính xác giữa các giá trị được dự đoán và đo được. Cuối cùng, R2 hiển thị thông tin về sự thay đổi ở các giá trị radon mà mô hình dự đoán giải thích. R2 gần bằng 1 cũng cho thấy tầm quan trọng của dự đoán.

Mô hình 12 đã sử dụng 13 nơ ron ẩn và cho giá trị MSE tốt nhất ở epoch 300 (Bảng 1). Nhìn chung, hiệu suất của các mô hình ANN trên dữ liệu huấn luyện và kiểm tra là tương tự nhau, điều này chứng tỏ rằng các mô hình không ghi nhớ dữ liệu huấn luyện mà học được mối quan hệ thực tế. Mô hình 10, 11 và 12 thể hiện hiệu suất cao nhất, với RMSE nhỏ hơn 5 (Bq/m³), sai số MAPE và MABE nhỏ hơn 3%, và r , R^2 xuất sắc (cao hơn 99%). Có một sự khác biệt nhỏ về hiệu suất giữa ba mô hình. Tuy nhiên, mô hình 12 thể hiện hiệu suất tốt nhất với giá trị R^2 cao nhất. Các lỗi thống kê RMSE, MAPE, MABE và r của nó trên bộ dữ liệu thử nghiệm lần lượt là 2,791, 1,117, 2,094 và 0,997.

Đánh giá mô hình

Độ lệch chủ yếu xảy ra khi mô hình dự đoán giá trị radon khoảng 200 (Bq/m³). Ngoài ra, hình dạng của phân phối dự đoán cho ba bộ dữ liệu là tương tự nhau, điều này cho thấy khả năng tổng quát hóa của mô hình ANN được đề xuất trên bộ dữ liệu mới.

Một khả năng dự đoán có ý nghĩa khác là lỗi dự đoán trên mỗi lượng tử được tạo bởi mô hình ANN được đề xuất. Các lượng tử trong phạm vi bằng nhau từ 10 đến 90% để đánh giá độ lệch giữa giá trị đo được và giá trị dự đoán. Nói chung, các giá trị lượng tử được dự đoán tương tự như các giá trị được đo, thể hiện khả năng của mô hình được đề xuất trong dự đoán radon cục bộ đạt độ chính xác. Đánh giá chi tiết được đưa ra trong Bảng 1, trong đó sai số (E) được tính cho mỗi phân vị. Lượng tử cao nhất ở phân vị 10, 20, 60 và 90%. Hơn nữa, độ lệch trung bình của tập dữ liệu thử nghiệm lớn hơn một chút so với độ lệch của tập dữ liệu huấn luyện và toàn bộ.

Ví dụ thứ hai

Ví dụ này được thực hiện và trình bày để kiểm tra tính hiệu quả của mô hình ANN-OHL khi so sánh với các mô hình khác.

Sáng chế sử dụng ba mô hình để so sánh bao gồm: SVM (Support Vector Machine), RF (Random Forest) và ANN hai lớp ẩn. Đây là ba trong số các mô hình học máy chuẩn, đã được sử dụng rộng rãi cho các bài toán dự báo. Đối với

mô hình SVM đang phát triển, hàm hạt nhân cơ sở hướng tâm được triển khai để huấn luyện mô hình và hai tham số C (cost) và δ (sigma) được điều chỉnh để kiểm soát độ chính xác của mô hình. Đối với mô hình RF, kích thước lá tối thiểu (m) và số lượng cây (n_t) được chọn để đánh giá hiệu suất của nó. Quy trình được thực hiện với nhiều thử nghiệm khác nhau về C , δ , m , n_t để tìm ra các tham số tối ưu cho cả hai mô hình. Dựa trên các giá trị RMSE, SVM và RF tốt nhất được xác định lần lượt là $C = 51,623$, $\delta = 0,024$ và $m = 800$, $n_t = 4$.

Như thể hiện trong Bảng 2, mô hình ANN một lớp ẩn được đề xuất (một trong những sự đặc biệt của sáng chế này) vượt trội so với các mô hình còn lại với các giá trị RMSE và MAPE thấp hơn và các giá trị r , R^2 cao hơn. Ngược lại, mô hình SVM mang lại hiệu suất kém nhất trong cả huấn luyện và thử nghiệm. Có thể thấy rằng việc tăng số lượng lớp ẩn, như trường hợp của mô hình ANN hai lớp ẩn, sẽ làm giảm độ chính xác của dự đoán và dẫn đến trạng bị quá mức (overfitting) khi có thể quan sát thấy hiệu suất tốt hơn trong thử nghiệm thay vì huấn luyện. Không chỉ vậy ANN hai lớp ẩn cho thấy sai số cao hơn gấp 3 lần nhưng hiệu quả lại thấp hơn về độ chính xác.

Bảng 2. Hiệu quả của một số mô hình học máy được áp dụng trong tập dữ liệu hiện tại

Hiệu quả	ANN một lớp ẩn đề xuất		Hai lớp ẩn ANN		SVM		RF	
	<i>Huấn luyện</i>	<i>Kiểm tra</i>	<i>Huấn luyện</i>	<i>Kiểm tra</i>	<i>Huấn luyện</i>	<i>Kiểm tra</i>	<i>Huấn luyện</i>	<i>Kiểm tra</i>
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

Phân tích độ nhạy

Mô hình dự báo được đề xuất theo sáng chế được sử dụng để dự báo giá trị hoạt độ phóng xạ của khí radon dựa trên sáu đặc điểm tham số đầu vào là tọa độ

theo trục x, tọa độ theo trục y, khoảng cách, hướng, suất liều gamma, và hàm lượng urani. Sáng chế triển khai phương pháp đánh giá độ nhạy (Permutation importance). Phương pháp này cơ bản được thực hiện như sau:

(1) Nhận mô hình ANN được huấn luyện;

(2) Xáo trộn các giá trị trong một cột tính năng duy nhất và đưa ra dự đoán bằng tập dữ liệu đã tạo. Các giá trị được dự đoán và đo lường được sử dụng để tính toán mức độ mất mát mà RMSE phải chịu do xáo trộn và ước tính tầm quan trọng của các tính năng được xáo trộn.

(3) Trả dữ liệu về thứ tự ban đầu và lặp lại bước (2) với tính năng tiếp theo.

Do tính ngẫu nhiên trong việc xáo trộn, bước (2) được lặp lại năm lần để lấy kết quả trung bình.

Việc xây dựng mô hình dự đoán sự phát tán khí phóng xạ từ khu vực mỏ đồng Sin Quyền, đã cho thấy có sự tương quan chặt chẽ giữa hàm lượng, suất liều gamma tại khu vực mỏ đồng có mối quan hệ urani với nồng độ khí phóng xạ radon tại các khu vực nhà dân cách xa khoảng 1km (hệ số tương quan về độ tin cậy giữa giá trị radon dự báo và giá trị thực được xác định $R^2=0,997$ theo bảng 2). Có thể khẳng định nồng độ khí phóng xạ radon đo được tại các khu vực nhà dân là do sự phát tán từ mỏ đồng Sin Quyền trong quá trình khai thác quặng.

Với kết quả trên, tương ứng với tập huấn luyện độ chính xác của tập kiểm tra đưa vào kiểm tra được kết quả $R^2 \sim 0,997$ tương ứng 99,7% mà sai số thấp.

Mô hình trở nên ít bị thừa hơn vì lỗi huấn luyện với $RMSE = 2,793$ và lỗi kiểm tra $RMSE = 2,791$ không quá khác biệt. Mô hình được đề xuất không chỉ là một cách tiếp cận mô hình hóa đơn giản mà còn là một mô hình dự đoán chính xác với giá trị sai số thấp (tức là $MAPE = 1,121$ (%), $MABE = 2,102$ (%), $r = 0,997$, $R^2 = 0,995$ cho phần huấn luyện và $MAPE = 1,117$ (%), $MABE = 2,094$ (%), $r = 0,997$, $R^2 = 0,995$ cho phần thử nghiệm).

Lợi ích của sáng chế

Phương pháp dự báo phát tán khí phóng xạ radon (^{222}Rn) tại khu vực hoặc mỏ chứa chất phóng xạ theo sáng chế có nhiều ưu điểm vượt trội so với các phương pháp đã biết như:

- Mô hình dự báo được lựa chọn đặc tính kỹ thuật tối ưu hóa, có thể mang lại hiệu quả cao khi khai thác bộ dữ liệu được chuẩn hóa, và phản ánh được bản chất, có tính đại diện mang lại hiệu quả thiết thực giảm thời gian, giảm số lượng tham số đầu vào, giảm chi phí mà vẫn đạt hiệu quả và độ chính xác cao.

- Kiến trúc của mô hình dự báo được tối giản hóa bằng cách sử dụng mạng nơ ron nhân tạo một lớp ẩn (ANN-OHL) trong khi vẫn khai thác và sử dụng hiệu quả cho bộ dữ liệu lớn quan trắc môi trường phóng xạ khí radon và với cấu trúc toàn bộ dữ liệu đã được lựa chọn.

- Với việc sử dụng mô hình ANN-OHL mang lại hiệu quả về thời gian, chi phí và độ chính xác tin cậy, có khả năng áp dụng và mở rộng.

- Phân tích và đánh giá được tham số độ nhạy và có ảnh hưởng đến mô hình dự báo phát tán khí radon ở khu vực điểm chứa chất phóng xạ.

- Việc sử dụng mô hình ANN một lớp ẩn được cho là không chỉ áp dụng tốt trong phạm vi rất rộng đối với tất cả các mỏ, điểm mỏ, khu vực khai thác, khu chế biến quặng, địa điểm sử dụng chất phóng xạ ở Việt Nam mà có thể còn phù hợp các điểm tương tự ở trên thế giới.

Trên đây, sáng chế đã được mô tả chi tiết theo các phương án ưu tiên thực hiện, và có thể kèm theo các phương án thay thế hoặc tương đương hoặc các ví dụ cụ thể, sử dụng các mô tả và thuật ngữ phù hợp để người có hiểu biết trung bình về lĩnh vực kỹ thuật này có thể hiểu và thực hiện được giải pháp theo sáng chế. Vì vậy, người có hiểu biết trung bình về lĩnh vực kỹ thuật tương ứng có thể dễ dàng tạo ra các thay đổi, cải biến, hoặc thay thế tương đương dựa vào các nội dung và phương án được mô tả. Do đó, các thay đổi, cải biến, hoặc thay thế tương đương này được coi là không nằm ngoài phạm vi của sáng chế, và phạm vi bảo hộ của sáng chế hiển nhiên là không bị giới hạn bởi các nội dung và phương án được mô tả mà được xác định trong yêu cầu bảo hộ dưới đây.

YÊU CẦU BẢO HỘ

1. Phương pháp dự báo phát tán khí phóng xạ radon (^{222}Rn) tại khu vực hoặc mỏ chứa chất phóng xạ, phương pháp này được thực hiện trên máy tính hoặc thiết bị tính toán, phương pháp này bao gồm các bước:

sử dụng mô hình dự báo theo kiến trúc ANN (Artificial Neural Network) gồm có lớp đầu vào, lớp đầu ra, và chỉ có một lớp ẩn, trong đó lớp ẩn của mô hình dự báo bao gồm từ mười đến mười ba nơ ron,

trong đó:

lớp đầu vào của mô hình dự báo có số lượng nút mạng đầu vào là sáu tương ứng với tập dữ liệu đầu vào bao gồm tọa độ theo trục x, tọa độ theo trục y, suất liều gamma, khoảng cách, hướng, và hàm lượng Urani, trong đó tọa độ theo trục x, tọa độ theo trục y, suất liều gamma, và hàm lượng urani là các giá trị được xác định tại vị trí cần dự báo, khoảng cách là khoảng cách từ vị trí cần dự báo tới mỏ chứa chất phóng xạ, và hướng là hướng được xác lập giữa vị trí cần dự báo và mỏ chứa chất phóng xạ,

lớp đầu ra có một nút mạng tương ứng với kết quả đầu ra là dự báo hoạt độ phóng xạ của khí radon tương ứng với tập dữ liệu đầu vào đã nêu;

dự báo, bởi mô hình dự báo nêu trên nhờ nhập vào lớp đầu vào của mô hình dự báo toàn bộ tập dữ liệu đầu vào tương ứng với vị trí cần dự báo, để đưa ra kết quả đầu ra tại lớp đầu ra của mô hình dự báo, kết quả đầu ra này được xác định là dự báo hoạt độ phóng xạ của khí radon tại vị trí cần dự báo;

trong đó:

việc huấn luyện mô hình dự báo đã nêu sử dụng bộ dữ liệu chuẩn hóa được phân tích và xử lý từ các dữ liệu thực tế thu thập từ các khu vực có mỏ chứa chất phóng xạ, được phân chia thành hai nhóm gồm có các tập dữ liệu huấn luyện và các tập dữ liệu kiểm tra theo tỉ lệ 80:20, để đưa ra bộ các trọng số được tối ưu hoá cho mô hình dự báo.

2. Phương pháp theo điểm 1, trong đó phương pháp này sử dụng hệ thống định

vị GPS tại vị trí cần dự báo để tọa độ theo trục x, tọa độ theo trục y.

3. Phương pháp theo điểm bất kỳ trong số các điểm từ 1 đến 2, trong đó ít nhất là các tập dữ liệu kiểm tra được sử dụng để đánh giá hiệu suất và độ chính xác của mô hình dự báo được huấn luyện thông qua ít nhất là một trong số các chỉ số đánh giá gồm có sai số gốc bình quân phương (RMSE), sai số phần trăm tuyệt đối trung bình (MAPE), sai số sai lệch tuyệt đối trung bình (MABE), hệ số tương quan (r), và hệ số xác định (R^2), hoặc sự kết hợp của chúng.

4. Phương pháp theo điểm 3, trong đó các chỉ số đánh giá đã nêu được xác định thông qua các biểu thức sau đây:

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

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

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

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

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

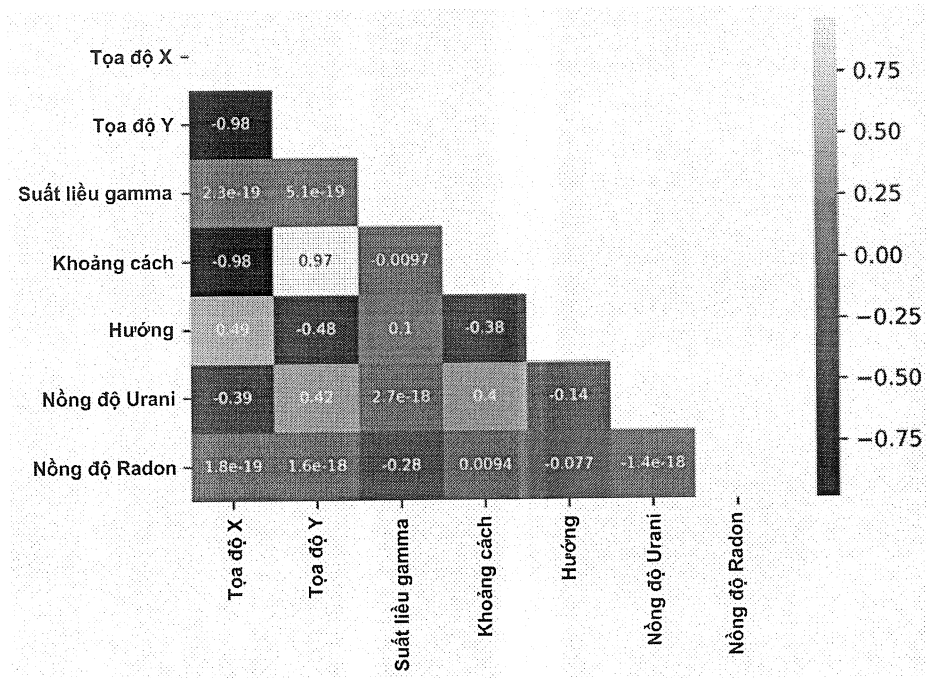
trong đó: y_t và Y_t lần lượt là báo các giá trị hoạt độ phóng xạ của khí radon tại vị trí cần dự báo đo được và dự đoán; $\bar{y}_t$ và $\bar{Y}_t$ là các giá trị hoạt độ phóng xạ của khí radon trung bình đo được và dự đoán; và T là số lượng các tập dữ liệu huấn luyện.

5. Phương pháp theo điểm bất kỳ trong số các điểm từ 1 đến 4, trong đó các dữ liệu thực tế thu thập từ các khu vực có mỏ chứa chất phóng xạ được chuẩn hóa để có giá trị nằm trong khoảng từ 0 đến 1 thông qua biểu thức sau đây:

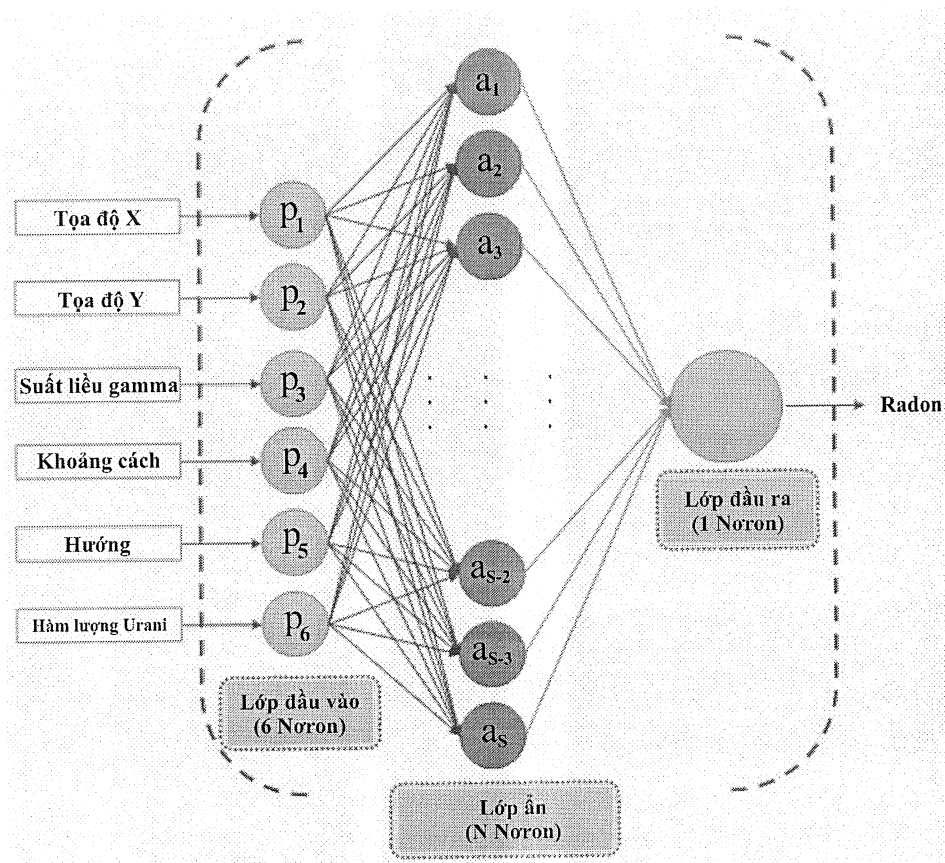
$$z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}$$

trong đó: $x (x_1, \dots, x_n)$ là dữ liệu thực tế, z_i là dữ liệu được chuẩn hóa thứ i .

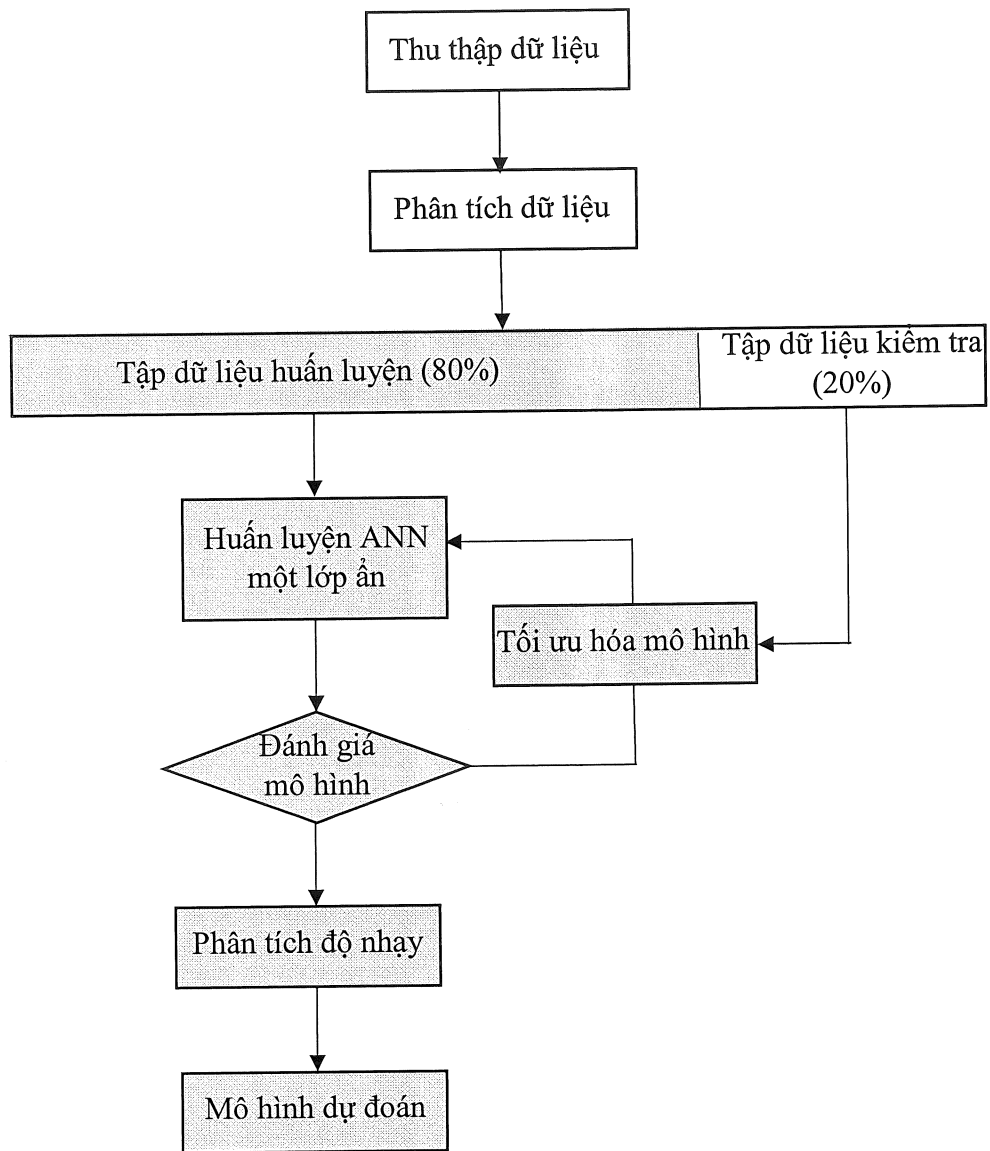
6. Phương pháp theo điểm bất kỳ trong số các điểm từ 1 đến 5, trong đó suất liều gamma được thực hiện thông qua phép đo tích lũy trong một khoảng thời gian xác định trước sao cho lược bỏ được sự ảnh hưởng từ các tham số khí tượng gồm có lượng mưa, nhiệt độ, gió, độ ẩm, và yếu tố địa hình địa mạo.



Hình 1



Hình 2

**Hình 3**