

# ĐỒNG DANH CÁC VỈA THAN MỎ NÚI BÉO, QUẢNG NINH BẰNG PHƯƠNG PHÁP HỒI QUY LOGISTIC VÀ MẠNG TRÍ TUỆ NHÂN TẠO

**Khuong Thế Hùng<sup>1\*</sup>, Tạ Thị Toán<sup>1</sup>, Nguyễn Danh Tuyên<sup>2</sup>**

<sup>1</sup>*Trường Đại học Mở - Địa chất, 18 Phố Viên, Đức Thắng, Bắc Từ Liêm, Hà Nội*

<sup>2</sup>*Công ty cổ phần Địa chất Việt Bắc, 65 An Trạch, Đống Đa, Hà Nội*

*\*Email: [khuongthehung@humg.edu.vn](mailto:khuongthehung@humg.edu.vn)*

## TÓM TẮT

Mỏ than Núi Béo nằm về phía Nam tỉnh Quảng Ninh, thuộc dải than Hòn Gai - Cẩm Phả, nơi được đánh giá là một trong những khu vực có tiềm năng lớn về than khoáng của nước ta. Trên cơ sở tổng hợp tài liệu, xử lý số liệu về các thông số vỉa và chất lượng than bằng thuật toán trí tuệ nhân tạo và phương pháp hồi quy logistic cho phép nhận dạng và khoanh nổi các vỉa than khu mỏ một cách phù hợp. Nhận dạng các vỉa than mỏ Núi Béo theo phương pháp hồi quy logistic cho kết quả đạt 70,87 %, trong khi đó kết quả nhận dạng các vỉa than bằng thuật toán trí tuệ nhân tạo cao hơn, đạt 84,94 %. Nhìn chung, nhận dạng các vỉa than bằng phương pháp trí tuệ nhân tạo cao hơn phương pháp hồi quy logistic, tuy nhiên vẫn còn 15,06 % sai số, đây cũng là kết quả có thể chấp nhận cho việc dùng thuật toán trong công tác đồng danh vỉa. So với phương pháp truyền thống, đồng danh các vỉa than khu mỏ Núi Béo bằng thuật toán cho thấy các vỉa được nổi đơn giản hơn, thoáng hơn và phù hợp với môi trường trầm tích. Kết quả nghiên cứu khẳng định vai trò của các phương pháp toán ứng dụng trong nghiên cứu địa chất, đặc biệt trong đồng danh các vỉa than và phù hợp với môi trường trầm tích khu mỏ.

*Từ khóa:* Đồng danh vỉa than, phương pháp hồi quy logistic, mạng trí tuệ nhân tạo, mỏ Núi Béo, Quảng Ninh.

## 1. MỞ ĐẦU

Các lớp trầm tích cũng như vỉa than phát triển tương đối liên tục trong quá trình thành tạo cho nên chúng có mối liên hệ không gian gần gũi nhau. Các lớp đá hoặc vỉa than gần gũi nhau về không gian sẽ có chiều dày, góc dốc, số lớp kẹp cũng như đặc điểm, tính chất tương tự nhau, chính các yếu tố này là cơ sở cho việc đồng danh, liên kết vỉa, tập vỉa than [2, 3, 5-10]. Sau khi thành tạo, quá trình hoạt động kiến tạo về sau sẽ làm thay đổi thể nằm, tạo uốn nếp, dịch chuyển các vỉa than gây ra sự gián đoạn, phức tạp hóa trong quá trình nổi vỉa. Chính vì vậy, yêu cầu cần phân chia vùng nghiên cứu thành các khu vực có tính đồng nhất tương đối để thực hiện công tác nổi, liên kết vỉa than, lớp trầm tích.

Việc đồng danh vỉa than trên bề than Quảng Ninh có độ tin cậy phụ thuộc vào mật độ công trình thăm dò, đặc điểm cấu trúc địa chất của từng khu mỏ, khối kiến trúc. Đồng danh vỉa giữa các khu mỏ trong cùng khối cấu trúc bậc IV hoặc trên toàn dải than, bề than phụ thuộc chủ yếu vào các phương pháp so sánh đặc điểm địa tầng giữa các vỉa than, so sánh về hình học, các đặc điểm của đá vách, đá trụ vỉa, quy luật biến đổi của vỉa than, cấu tạo vỉa, chiều dày vỉa, chất lượng và các tính chất vật lý của các vỉa, phương pháp tầng đánh dấu cũng đã được chú trọng áp dụng [2, 3, 6]. Tuy nhiên, do tình trạng trạng thiết bị, trình độ khoa học công nghệ hạn chế nên việc tìm ra những dấu hiệu đặc trưng cho riêng một vài lớp, một vài tập nào đó là rất khó khăn. Mục đích của nghiên cứu

này là nhận dạng các vỉa than tại khu vực mỏ Núi Béo, Quảng Ninh bằng thuật toán hồi quy logistic và mạng trí tuệ nhân tạo trên cơ sở phân tích các thông số chiều dày, góc dốc và chất lượng vỉa than nhằm phục vụ công tác đồng danh liên kết vỉa, góp phần định hướng cho công tác thiết kế khai thác được chính xác và phù hợp với điều kiện khu mỏ.

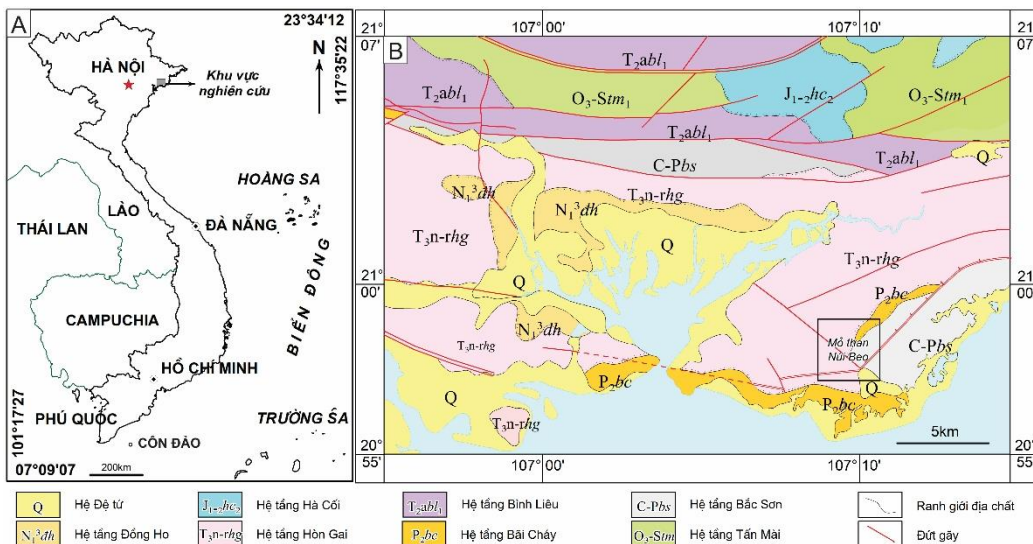
## 2. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU

### 2.1. Đối tượng

Mỏ Núi Béo thuộc dải than Hòn Gai - Cẩm Phả, khu mỏ nằm trong địa phận của 03 phường, đó là phường Hà Phong, phường Hà Tu và phường Hà Trung, thành phố Hạ Long, tỉnh Quảng Ninh. Khu mỏ nằm ở trung tâm thành phố Hạ Long, nằm bên trái đường Quốc lộ 18A từ Hạ Long đi Mông Dương (Hình 1A).

Khu mỏ phổ biến các thành tạo trầm tích Trias thuộc hệ tầng Hòn Gai, phân hệ tầng giữa và các trầm tích bờ rời hệ Đệ tứ [8] (Hình 1B). Thành phần thạch học của phân hệ tầng Hòn Gai giữa bao gồm các lớp cuội kết, cát kết, bột kết, sét kết, sét than và các vỉa than nằm xen kẽ nhau, chiều dày địa tầng khoảng 1.800 m. Phân hệ tầng Hòn Gai giữa là đối tượng chứa các vỉa than công nghiệp. Trầm tích hệ Đệ tứ (Q) phủ trực tiếp lên các thành tạo của phân hệ tầng Hòn Gai giữa, chúng được phân bố ở các khu vực thấp, thung lũng xung quanh khu mỏ. Thành phần trầm tích bao gồm cuội, sỏi, cát, sét bờ rời, đôi nơi là các tầng lãn, đây là sản phẩm phong hoá từ các đá có trước.

Trong khu mỏ Núi Béo phát triển các nếp uốn và hệ thống các đứt gãy, chúng làm phức tạp và gây khó khăn cho công tác đồng danh vỉa và khai thác than. Theo Phạm Tuấn Anh (2009) [1], từ trên bề mặt địa hình trở xuống, trong phạm vi khu mỏ Núi Béo tồn tại các vỉa than sau V14, V13, V11, V10, V9, V7, V6, V5 và V4.



Hình 1. A - Bản đồ Việt Nam và vị trí vùng Hạ Long, B - Sơ đồ địa chất vùng Hạ Long, Quảng Ninh và vị trí khu mỏ Núi Béo [8]

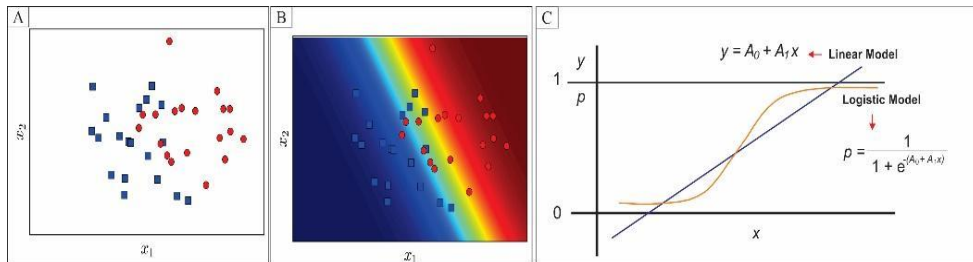
### 2.2. Phương pháp nghiên cứu

Trên cơ sở thu thập các tài liệu địa chất đã tiến hành tại khu vực nghiên cứu, đặc biệt là các tài liệu thăm dò, khai thác triển khai trên khu vực Hà Lầm - Núi Béo và tài liệu bổ sung của 11 lỗ

khoan thực hiện trong năm 2021 thuộc Dự án thăm dò bổ sung than mở Núi Béo do Công ty Cổ phần Địa chất Việt Bắc thực hiện, các số liệu phân tích tại Trung tâm Phân tích Thí nghiệm Địa chất, thuộc Tổng cục Địa chất và Khoáng sản Việt Nam. Các số liệu thu thập từ 223 lỗ khoan (chiều dày than, góc dốc, lớp kẹp), và 438 kết quả phân tích các loại ( $W^{pt}$ ,  $A^k$ ,  $V^{ch}$ ,  $Q^{ch}$ ,...) được xử lý, tổng hợp cho từng vỉa, sử dụng phục vụ công tác nghiên cứu.

### 2.2.1. Phân nhóm theo phương pháp hồi quy logistic (logistic regression)

Theo Freedman (2009) [4], hồi quy logistic (logistic regression) là phương pháp phân chia các đối tượng thành các nhóm khác nhau bởi một ranh giới nào đó.



Hình 2. A - Biểu đồ phân chia các đối tượng theo phương pháp hồi quy logistic: phân bố các nhóm khác nhau, B - Xây dựng các đường phân chia các đối tượng, C - Phân biệt mô hình hồi quy tuyến tính và hồi quy logistic

Như vậy, mô hình hồi quy tuyến tính được xác định theo hàm:  $y = A(0) + \sum_{j=1}^m A(j) \times x(j)$

Mô hình hồi quy logistic được xác định theo hàm:  $P = \frac{1}{1 + e^{-z}}$

trong đó:  $Z = A(0) + \sum_{j=1}^m A(j) \times x(j)$ ;  $P$  là xác suất 0/1 hoặc  $[0,1]$ ;  $y$  - Biến phụ thuộc;  $x(j)$  - Biến độc lập của thông số  $j$ ;  $m$  là số thông số.

Các giá trị 0/1 được xác định theo nguyên lý lấy vỉa hiện tại so sánh với vỉa chuẩn. Ví dụ ta cần xác định vỉa 6 theo phân loại ban đầu. Nếu vỉa này được xác định là chuẩn của vỉa 6 thì sẽ nhận được giá trị 1. Nếu không xác định là vỉa chuẩn thì ta cần dùng phương pháp phân loại để trả lời, các vỉa khác không phải vỉa 6 ban đầu và không là vỉa chuẩn thì nhận giá trị 0. Ngoài ra, chúng ta có thể hiệu chỉnh dựa trên không gian địa tầng gần gũi với các vỉa chuẩn cho các vỉa phân loại một giá số nào đó mà không vượt quá 0,5; cách xử lý này sẽ giúp làm tăng sự điều chỉnh nhận dạng các vỉa lân cận.

### 2.2.2. Phân nhóm theo phương pháp trí tuệ nhân tạo (Artificial Neural Network - ANN)

Những năm đầu của thập niên 90, Schmidhuber (1992) [11] đã định nghĩa mạng neural (nơron) hay còn gọi là mạng trí tuệ nhân tạo là một hệ thống xử lý thông tin có một vài đặc tính thực thi giống như các mạng nơron sinh học. Mạng nơron được phát triển như là tổng quát hóa của các mô hình toán học của nhận thức của con người hay hệ thần kinh sinh học, dựa trên các giả định sau: Thông tin xử lý tại nhiều phần tử đơn giản, gọi là các nơron; tín hiệu được truyền giữa các nơron thông qua các kết nối; mỗi kết nối có một trọng số, thông thường được nhân với tín hiệu truyền qua; mỗi nơron sử dụng một hàm kích hoạt để xác định tín hiệu đầu ra theo tổng các tín hiệu đầu vào. Một mạng nơron được đặc trưng bởi ba đặc tính sau:

- 1) Hình thái của các kết nối giữa các nơron (còn gọi là kiến trúc mạng);
- 2) Phương pháp xác định các trọng số của các kết nối (còn gọi là thuật toán học hay thuật toán huấn luyện - learning/training algorithm);
- 3) Hàm kích hoạt (activation function).

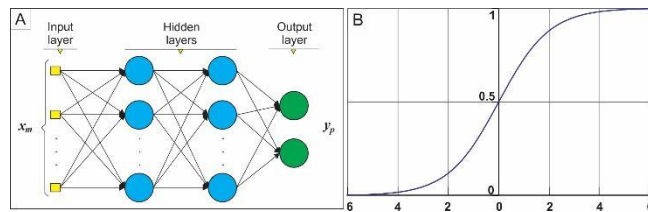
Một mạng nơron điển hình bao gồm một lớp đầu vào (input layer), một hay nhiều lớp ẩn (hidden layers) và một lớp đầu ra (output layer). Trong các ứng dụng thực tế, một trong số các mạng nơron được dùng nhiều nhất là mạng nơron tiếp tục cung cấp (feed-forward) với thuật toán huấn luyện lan truyền ngược (back-propagation) và hàm kích hoạt sigmoid.

Hàm kích hoạt sigmoid: 
$$y = \frac{1}{1 + e^{-x}}$$

Với mạng nơron có  $n_1$  đầu vào  $x$ ,  $n_2$  ở lớp ẩn thì ở đầu ra  $y$  của mạng nơron có thể biểu diễn một cách toán học như sau:

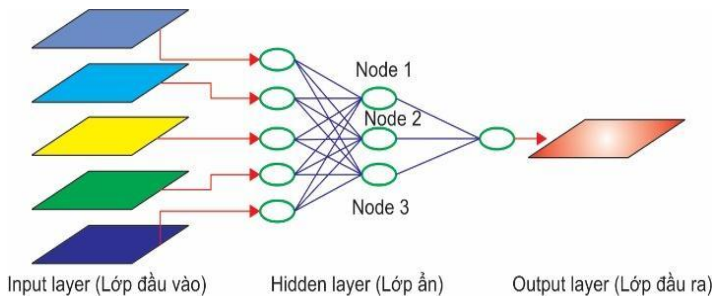
$$y = f \left[ a_o + \sum_{(j=1-n_2)} a_j f_j (b_{oj} + \sum_{(i=1-n_1)} b_{ij} x_i \right]$$

trong đó:  $f$  là hàm kích hoạt;  $a, b$  là các trọng số của các kết nối;  $a_o, b_o$  gọi là các thiên lệch (bias).



Hình 3. A - Mạng nơron điển hình; B - Hàm kích hoạt sigmoid

Hàm kích hoạt sigmoid là hàm nhận dạng 0/1 (sai = 0, đúng = 1) dùng để phân loại các đối tượng, giá trị chỉ biến đổi trong khoảng (0,1).



Hình 4. Mô hình mạng trí tuệ nhân tạo (ANN)

Trong nghiên cứu này chúng tôi sử dụng mô hình ANN 3 lớp ẩn, 1 lớp đầu ra, còn lớp đầu vào tùy thuộc vào số lớp dữ liệu được xây dựng. Do chiều dày than, lớp kẹp và góc dốc via là những thông số có độ tin cao, tiếp đến là độ ẩm ( $W^{Pt}$ ) và độ tro ( $A^k$ ) là 2 thông số chiếm tỷ trọng thông tin, kết quả đầu ra là via than. Chính vì vậy, phương pháp ANN sử dụng số lớp thông tin đầu vào (Input layer) 5; số lớp thông tin ẩn (Hidden layers) 3; số lớp thông tin đầu ra (Output layer) 1; với phương pháp nội suy là standard sigmoid; hệ số luyện (Learning rate) 0,5.

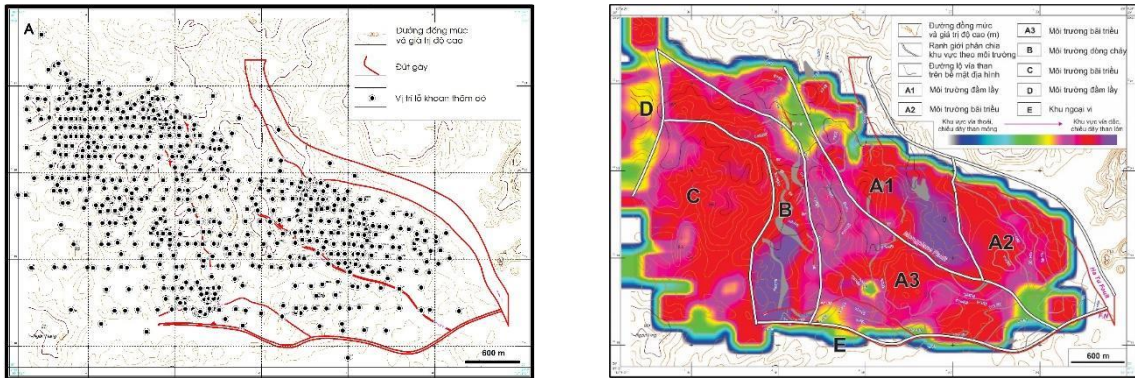


Phương pháp ANN cũng tính ra các hệ số  $A(j)$  và tính toán phi tuyến tính cho các lớp thông tin. Từ các lớp thông tin đầu vào (chiều dày, góc dốc, lớp kẹp, đá vây quanh than, biến đổi đá trên than, dưới than, các thông số  $W^{pt}$ ,  $A^k$ ,  $V^{ch}$ ,  $Q^{ch}$ ,  $d$ ,  $m$ ,...) tương ứng với giá trị đầu ra (tên vỉa,...), ta cần dạy cho chương trình nhận dạng, tìm được các hệ số  $A(j)$  sao cho hệ số tương quan cao nhất. Kết quả cuối cùng là hệ số của các lớp ẩn, để tính ra các giá trị đầu ra.

### 3. KẾT QUẢ VÀ THẢO LUẬN

#### 3.1. Môi trường tạo than và thông số nhận dạng vỉa

Theo Khương Thế Hùng và cộng sự [6], mỏ than Núi Béo được phân thành các khu đồng nhất tương đối theo môi trường thành tạo: Khu A1 mang đặc tính môi trường đầm lầy; khu A2-Bãi triều; khu A3-Bãi triều; khu B-Dòng chảy; khu C-Bãi triều; khu D-Đầm lầy; khu E-Ngoại vi. Khu A1 và A2 lấy ranh giới bằng đứt gãy F-Monplane; giữa A1 và A2 lấy ranh giới của các yếu tố phân nhóm. Khu B và khu A là đứt gãy F-K và F-C đồng thời là đới có độ dốc lớn; khu C và khu B được giới hạn bởi sự hình thành biểu hiện dòng chảy qua khu B. Khu D và khu C có ranh giới là đới độ dốc lớn; khu E tách biệt, lấy đứt gãy F-N làm ranh giới (Hình 5).



Hình 5. A - Sơ đồ vị trí các lỗ khoan thăm dò mỏ than Núi Béo;  
B - Sơ đồ phân khu mỏ than Núi Béo theo môi trường tạo than [5]

Bảng 1. Bảng đánh giá thống kê tính đồng bộ của các thông số

Thông số	Góc dốc	Chiều dày khoan	Số lớp kẹp	$W^{pt}$	$A^k$	$V^{ch}$	$Q^{ch}$	$Q^{kh}$	$d$	$S$
Có mặt (%)	71,07	70,8	70,8	45,73	51,07	44,53	42,27	42,27	41,73	33,47
Dãy thông số	$Q^{kh}$	$Q^{ch}$	$A^k$	Số lớp kẹp	Chiều dày khoan	Góc dốc	$d$	$V^{ch}$	$W^{pt}$	$S$
Đồng bộ (%)	42,27	42,27	42,27	42,27	32,93	32,93	32,93	31,73	29,33	27,07
Thứ tự	Số lớp kẹp	Góc dốc	Chiều dày khoan	$A^k$	$Q^{ch}$	$Q^{kh}$	$V^{ch}$	$W^{pt}$	$d$	$S$
Đánh giá (%)	56,535	52	51,865	46,67	42,27	42,27	38,13	37,53	37,33	30,27

Từ kết quả Bảng 1 cho phép có thể chọn thông số: Số lớp kẹp, Góc dốc, Chiều dày khoan làm cơ sở để phân loại. Nếu sử dụng kết quả phân tích thì chọn thêm chỉ tiêu độ tro ( $A^k$ ), cần nữa thì thêm nhiệt lượng ( $Q^{ch}$ ) và  $Q^{kh}$ . Tuy nhiên, nếu chọn tiếp thì các thông số không đủ để phân loại nên việc sử dụng các thông số cần chọn theo nhóm.

Bước 1 - Dựa trên bộ 3 thông số khoan: Số lớp kẹp, Góc dốc, Chiều dày khoan.

Bước 2 - Dựa trên 3 thông số khoan, kết hợp với thông số phân tích.

### 3.2. Phân loại theo phương pháp hồi quy logistic và phương pháp ANN

#### 3.2.1. Phương pháp hồi quy logistic

Bộ dữ liệu không gian được chia thành 2 tập dữ liệu là tập dữ liệu huấn luyện và tập dữ liệu xác nhận theo tỷ lệ tương ứng là 70 % và 30 %, được sử dụng để huấn luyện mô hình. Thực quan hóa phương pháp luận xây dựng mô hình ANN đồng danh via than như Hình 6. Phương pháp đi xây dựng phương trình hồi quy logistic bằng cách xác định các hệ số A(j) của phương trình, kết quả tính toán như sau (Bảng 2).

Bảng 2. Hệ số cho phương trình hồi quy logistic các via than theo từng khu vực

Tên via	Khu A1		Khu A2		Khu A3	
	Hệ số	Giá trị	Hệ số	Giá trị	Hệ số	Giá trị
Via 14	A(1)	0,088022	A(1)	0,002707	A(1)	0,032732
	A(2)	0,272148	A(2)	1,12108	A(2)	0,217405
	A(3)	-0,09336	A(3)	-3,9892	A(3)	-0,14573
	A(0)	-3,85164	A(0)	-4,22798	A(0)	-1,61819
	Error	0,296827	Error	0,284721	Error	0,240848
	R	0,774902	R	0,518068	R	0,401638
Via 13	A(1)	0,000244	A(1)	0,012333	A(1)	0,001388
	A(2)	-0,12559	A(2)	-0,52641	A(2)	-0,14068
	A(3)	0,001976	A(3)	0,005615	A(3)	0,334966
	A(0)	-0,02099	A(0)	0,446213	A(0)	-0,34768
	Error	0,364171	Error	0,357316	Error	0,423513
	R	0,409526	R	0,392098	R	0,175268
Via 12	A(1)	-0,04838	A(1)	-0,03699	A(1)	0,010142
	A(2)	-0,02873	A(2)	-3,63904	A(2)	-0,49209
	A(3)	-1,23011	A(3)	2,530731	A(3)	1,088893
	A(0)	-1,91559	A(0)	3,523149	A(0)	-2,63852
	Error	0,139129	Error	0,145752	Error	0,151792
	R	0,144595	R	0,607569	R	0,251075
Via 11	A(1)	-0,03501	A(1)	-0,02514	A(1)	-0,0065
	A(2)	0,031669	A(2)	-0,18835	A(2)	-0,0007
	A(3)	0,177201	A(3)	0,426931	A(3)	-0,04324
	A(0)	-0,29457	A(0)	0,099162	A(0)	-0,4608
	Error	0,415585	Error	0,450393	Error	0,40148
	R	0,23632	R	0,215475	R	0,057493
Via 10	A(1)	0,013269	A(1)	0,016448	A(1)	0,001596
	A(2)	0,084949	A(2)	-0,04197	A(2)	0,078735
	A(3)	-0,67708	A(3)	0,028689	A(3)	-0,07391
	A(0)	-0,60885	A(0)	-0,66645	A(0)	-0,55605
	Error	0,413784	Error	0,477694	Error	0,392729
	R	0,278119	R	0,103531	R	0,189156
Via 9	A(1)	-0,00343	A(1)	-0,00994	A(1)	-0,02757
	A(2)	-0,18373	A(2)	0,083062	A(2)	-0,57098
	A(3)	-0,14044	A(3)	-0,39948	A(3)	0,540446
	A(0)	-2,05877	A(0)	-1,03753	A(0)	-0,69387

Đồng danh các vỉa than mở Núi Béo, Quảng Ninh bằng phương pháp hồi quy logistic và mạng...

	Error	0,162378	Error	0,40397	Error	0,133338
	R	0,157772	R	0,184709	R	0,286658
Vĩa 7	A(1)	-0,02621	A(1)	0,007157	A(1)	0,027608
	A(2)	0,085705	A(2)	0,202912	A(2)	0,487379
	A(3)	0,709586	A(3)	0,319085	A(3)	0,117666
	A(0)	-0,22664	A(0)	-2,57114	A(0)	-2,98334
	Error	0,3935	Error	0,406092	Error	0,301582
	R	0,519063	R	0,553502	R	0,756735
Vĩa 6	A(1)	0,016889	A(1)	-0,01601	A(1)	-0,04377
	A(2)	-0,44566	A(2)	-1,13996	A(2)	-0,41193
	A(3)	0,166863	A(3)	0,977674	A(3)	0,398476
	A(0)	0,441225	A(0)	3,317148	A(0)	1,08518
	Error	0,372311	Error	0,278603	Error	0,355721
	R	0,545029	R	0,795297	R	0,378055

### 3.2.2. Phương pháp ANN

Kết quả huấn luyện theo phương pháp ANN cho các thông số nhận dạng vỉa như sau:

Bảng 3. Hệ số ANN các vỉa than theo các khu khác nhau của mỏ Núi Béo

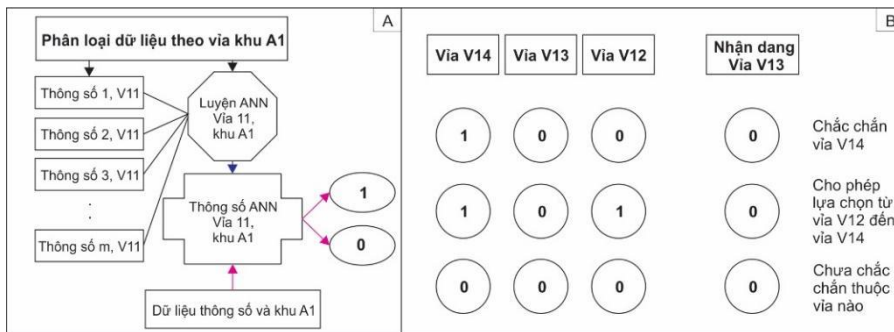
Tên vỉa	Node	Giá trị	Parameter	Node 1	Node 2	Node 3
Vĩa 14 khu A1	Node 0	-6,026	A(0)	-14,013	0,997	4,693
	Node 1	39,072	Gocdoc	13,124	-0,404	-11,446
	Node 2	-4,106	Bedaykhoan	72,170	-3,218	-68,944
	Node 3	-60,617	Solopkep	37,456	-0,784	41,778
Vĩa 14 khu A2	Node 0	3,090	A(0)	12,649	-2,235	27,815
	Node 1	-61,696	Gocdoc	-16,510	9,563	-64,695
	Node 2	-12,038	Bedaykhoan	-19,357	-26,471	-28,311
	Node 3	57,050	Solopkep	17,867	19,227	-9,867
Vĩa 14 khu A3	Node 0	-0,184	A(0)	-16,251	13,630	-36,573
	Node 1	32,286	Gocdoc	10,786	-21,771	-0,1395
	Node 2	-28,079	Bedaykhoan	16,797	-108,206	47,783
	Node 3	-27,145	Solopkep	5,119	-8,227	26,245
Vĩa 13 khu A1	Node 0	57,668	A(0)	-32,916	26,389	-16,778
	Node 1	-60,105	Gocdoc	38,485	-54,325	34,723
	Node 2	-58,410	Bedaykhoan	213,145	-114,016	-17,828
	Node 3	-64,556	Solopkep	20,325	-71,768	-35,794
Vĩa 13 khu A2	Node 0	0,756	A(0)	4,673	-35,135	19,149
	Node 1	71,082	Gocdoc	-0,138	29,141	-13,230
	Node 2	-17,665	Bedaykhoan	-33,303	50,246	-86,592
	Node 3	-69,055	Solopkep	16,575	16,951	36,563
Vĩa 13 khu A3	Node 0	22,781	A(0)	-3,704	10,496	2,748
	Node 1	-71,486	Gocdoc	-19,853	28,598	-7,834
	Node 2	-23,844	Bedaykhoan	16,222	-38,473	-244,395
	Node 3	56,246	Solopkep	9,887	-57,716	82,105
Vĩa 12 khu A1	Node 0	34,076	A(0)	-93,498	76,595	-10,009
	Node 1	-74,522	Gocdoc	223,120	-207,531	52,589
	Node 2	-70,253	Bedaykhoan	355,277	-280,301	-38,492
	Node 3	-33,051	Solopkep	22,231	13,670	3,00842
Vĩa 12	Node 0	13,586	A(0)	-14,231	-5,156	-3,358

khu A2	Node 1	23,149	Gocdoc	21,721	17,867	0,246
	Node 2	-19,639	Bedaykhoan	2,167	7,924	30,150
	Node 3	-25,953	Solopkep	-2,716	-10,784	-13,425
Via 12 khu A3	Node 0	52,209	A(0)	19,731	36,279	-6,923
	Node 1	-13,033	Gocdoc	-16,444	-112,782	23,192
	Node 2	-38,634	Bedaykhoan	7,852	8,029	38,885
	Node 3	-44,807	Solopkep	-30,314	11,587	-3,240
Via 11 khu A1	Node 0	-110,308	A(0)	17,613	-40,398	-16,654
	Node 1	110,323	Gocdoc	-62,328	-0,821	66,048
	Node 2	107,283	Bedaykhoan	-26,532	61,280	14,646
	Node 3	109,064	Solopkep	20,547	17,902	-25,894
Via 11 khu A2	Node 0	-7,691	A(0)	45,852	2,531	45,075
	Node 1	227,750	Gocdoc	-13,222	-257,766	-9,539
	Node 2	53,920	Bedaykhoan	-43,355	-59,130	-23,305
	Node 3	-221,553	Solopkep	-45,444	49,392	-79,685
Via 11 khu A3	Node 0	-0,7190	A(0)	0,403	16,825	-4,430
	Node 1	-42,125	Gocdoc	0,110	-143,780	-31,192
	Node 2	107,956	Bedaykhoan	-300,414	159,748	54,596
	Node 3	-111,513	Solopkep	48,307	-297,982	-43,014
Via 10 khu A1	Node 0	-0,487	A(0)	68,157	-56,792	97,966
	Node 1	-2,854	Gocdoc	-95,959	20,066	-134,341
	Node 2	-22,115	Bedaykhoan	-309,702	53,216	-61,133
	Node 3	1,676	Solopkep	110,936	58,292	110,652
Via 10 khu A2	Node 0	-1,422	A(0)	130,041	-52,288	322,058
	Node 1	-89,504	Gocdoc	-59,265	135,365	-189,735
	Node 2	95,256	Bedaykhoan	-456,543	-145,823	-510,676
	Node 3	2,577	Solopkep	65,775	-172,659	-159,772
Via 10 khu A3	Node 0	-5,216	A(0)	9,247	107,141	36,342
	Node 1	6,033	Gocdoc	-8,674	-93,171	-37,827
	Node 2	-5,136	Bedaykhoan	12,679	-556,479	-297,256
	Node 3	4,108	Solopkep	-29,278	0,265	280,465
Via 9 khu A1	Node 0	14,555	A(0)	2,004	56,890	-44,887
	Node 1	25,703	Gocdoc	-23,369	-97,444	69,961
	Node 2	-54,818	Bedaykhoan	32,056	-106,075	87,130
	Node 3	-46,798	Solopkep	-4,416	0,6139	-3,738
Via 9 khu A2	Node 0	83,564	A(0)	49,723	-11,153	-12,008
	Node 1	-82,564	Gocdoc	-30,487	-61,237	34,351
	Node 2	-84,336	Bedaykhoan	-108,704	46,020	-4,891
	Node 3	-83,799	Solopkep	-23,535	66,113	-5,690
Via 9 khu A3	Node 0	-2,239	A(0)	28,182	-0,505	21,858
	Node 1	46,389	Gocdoc	-80,800	-11,001	-62,225
	Node 2	-46,359	Bedaykhoan	-95,629	102,163	-193,109
	Node 3	-81,237	Solopkep	-92,688	-68,057	-17,563
Via 7 khu A1	Node 0	2,552	A(0)	-5,953	55,010	-28,101
	Node 1	63,229	Gocdoc	9,840	-21,039	87,205
	Node 2	-3,562	Bedaykhoan	-10,088	-56,791	-184,945
	Node 3	-60,237	Solopkep	1,601	-68,135	-15,636
Via 7 khu A2	Node 0	27,359	A(0)	75,451	0,203	-31,903
	Node 1	-29,188	Gocdoc	-60,232	-297,329	69,328



Đồng danh các vỉa than mỏ Núi Béo, Quảng Ninh bằng phương pháp hồi quy logistic và mạng...

	Node 2	60,550	Bedaykhoan	-100,446	-14,675	-38,922
	Node 3	-36,343	Solopkep	-7,262	4,213	-39,419
Vía 7 khu A3	Node 0	31,307	A(0)	-3,151	-17,527	23,123
	Node 1	-3,323	Gocdoc	46,929	26,250	133,670
	Node 2	4,479	Bedaykhoan	-92,055	50,958	1,310
	Node 3	-32,636	Solopkep	-182,860	-33,650	-134,781
Vía 6 khu A1	Node 0	0,674	A(0)	-9,067	-14,424	-13,812
	Node 1	38,024	Gocdoc	-21,374	-21,285	-8,849
	Node 2	-30,719	Bedaykhoan	19,598	204,418	20,498
	Node 3	-20,987	Solopkep	28,172	-69,085	19,407
Vía 6 khu A2	Node 0	40,964	A(0)	29,206	-36,029	-11,684
	Node 1	-11,048	Gocdoc	-124,377	111,579	-5,866
	Node 2	-40,214	Bedaykhoan	40,952	275,315	11,006
	Node 3	-37,402	Solopkep	-10,440	-178,306	30,046
Vía 6 khu A3	Node 0	-0,295	A(0)	-22,282	6,133	0,0101
	Node 1	-21,257	Gocdoc	24,132	-59,315	-14,106
	Node 2	34,235	Bedaykhoan	64,167	-40,888	-5,586
	Node 3	-58,540	Solopkep	-33,882	66,8301	12,650



Hình 6. A - Mô hình xử lý dữ liệu theo phương pháp ANN vỉa than V11 cho khu A1, B - Các trường hợp nhận dạng vỉa than V13

Trên cơ sở các thông số chạy theo vỉa và các khu khác nhau (Bảng 3), ta đưa vào chương trình tính toán. Đầu vào là các thông số, đầu ra là giá trị 1 và giá trị 0.

Giả sử thuật toán chạy cho vỉa 11 ở khu A1; giá trị 1 xác định chắc chắn là vỉa 11, giá trị 0 chưa chắc là vỉa 11, như vậy ta có các giá trị 1 tương ứng với vỉa 11. Tính toán cho toàn bộ các lỗ khoan cho kết quả các vỉa được nhận dạng. Tuy nhiên, kết quả các vỉa đã được nhận dạng có thể đúng, có thể bị nhận dạng sai, các vỉa chưa chắc chắn (0) chưa thuộc vỉa nào sẽ tương ứng với vỉa chuẩn (Hình 6). Kết quả đào tạo cho thấy mô hình đã thực hiện rất tốt với tập dữ liệu huấn luyện, mức độ chính xác của mô hình được huấn luyện bởi tập dữ liệu huấn luyện rất cao là 95,10 %. Mức độ phù hợp của mô hình và bộ dữ liệu đào tạo là tốt ở mức 0,91 (Kappa) với sai số trung phương thấp bằng 0,193 %.

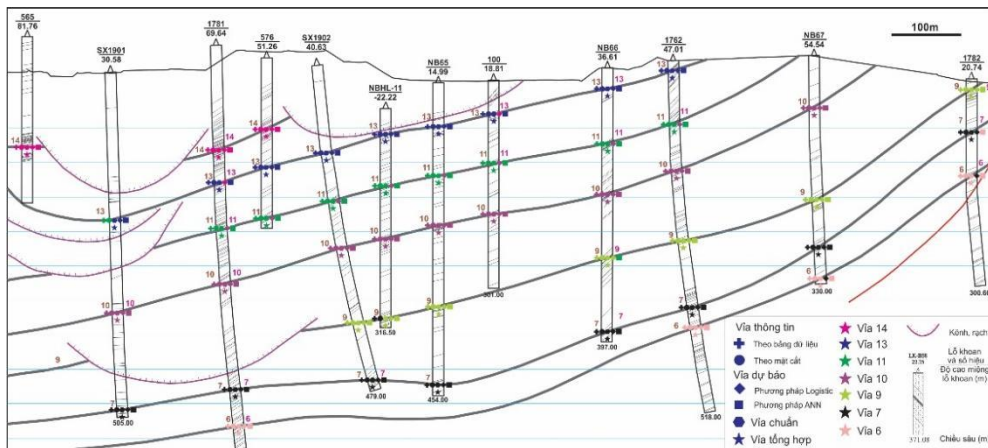
Kết quả nhận dạng theo thuật toán cho thấy có trường hợp cần phải xem xét, xử lý. Các dữ liệu theo thông số và theo khu được đưa vào chương trình để nhận dạng vỉa (ví dụ vỉa 11), nếu kết quả cho giá trị 1 (nhận dạng được vỉa), giá trị 0 (chưa nhận dạng được vỉa). Như vậy, các vỉa trong khu được chạy ngẫu nhiên theo đầu ra là vỉa than. Nếu chương trình chỉ có một trường hợp nhận dạng ra vỉa, ta xác định đây là vỉa chắc chắn; trường hợp chương trình nhận dạng ra nhiều vỉa khác nhau (vỉa nào cũng nhận vỉa đó thuộc về mình) thì kết quả có thể nhận dạng vỉa trong khoảng 2 vỉa đó.

Ngoài ra, các trường hợp chương trình không nhận ra via thì cũng có thể xếp vào via chuẩn (Hình 6A,B).

Chính vì vậy, sau khi nhận dạng via than bằng phương pháp ANN ta có bước xử lý tiếp theo là hiệu chỉnh các kết quả nhận dạng do thuật toán đưa ra.

Bảng 4. Đánh giá kết quả dự báo via than theo phương pháp hồi quy logistic và ANN

Via	Loại	Số mẫu	Nhận dạng đúng		Nhận dạng sai		Tỷ lệ đạt (%)	
			Hồi quy	ANN	Hồi quy	ANN	Hồi quy	ANN
14	Khu A1	55	36	51	19	4	65,45	92,73
14	Khu A2	19	14	17	5	2	73,68	89,47
14	Khu A3	21	15	10	6	11	71,43	<b>47,62</b>
13	Khu A1	91	65	73	26	18	71,43	80,22
13	Khu A2	53	41	48	12	5	77,36	90,57
13	Khu A3	59	33	48	26	13	55,93	78,69
12	Khu A1	123	121	121	2	2	98,37	98,37
12	Khu A2	86	86	86	0	0	<b>100</b>	<b>100</b>
12	Khu A3	113	79	112	34	1	69,91	99,11
11	Khu A1	94	59	73	35	21	62,77	77,66
11	Khu A2	107	73	85	34	22	68,22	79,44
11	Khu A3	123	74	84	49	39	60,16	68,29
10	Khu A1	73	43	56	30	17	58,90	76,71
10	Khu A2	93	42	76	51	17	45,16	81,72
10	Khu A3	109	53	77	56	32	48,62	70,64
9	Khu A1	89	86	88	3	1	96,63	98,88
9	Khu A2	116	73	96	43	20	62,93	82,76
9	Khu A3	149	147	149	2	0	98,66	<b>100</b>
7	Khu A1	35	21	30	14	5	60,00	85,71
7	Khu A2	60	24	53	36	7	<b>40,00</b>	88,33
7	Khu A3	82	47	71	35	11	57,32	86,59
6	Khu A1	29	22	25	7	4	75,86	86,21
6	Khu A2	33	23	31	10	2	69,70	93,94
6	Khu A3	77	61	67	16	10	79,22	87,01
<b>Trung bình</b>							<b>70,87</b>	<b>84,94</b>



Hình 7. Đồng danh via than theo kết quả tính toán logistic và ANN trên mặt cắt T.VI

Kết quả trên cho thấy việc nhận dạng theo phương pháp hồi quy logistic đạt tỷ lệ 70,87 %, phương pháp ANN có tỷ lệ 84,94 %; cao hơn so với hồi quy logistic (Bảng 4). Kết quả nhận dạng vỉa theo phương pháp hồi quy logistic và ANN đều được đưa lên mặt cắt các tuyến thăm dò để phục vụ đồng danh vỉa than, tập vỉa than (Hình 7).

Đối chiếu với kết quả nổi vỉa của Phạm Tuấn Anh và đồng nghiệp [1], kết quả đồng danh các vỉa than khu mỏ Núi Béo bằng thuật toán cho thấy trên tổng 751 vị trí vỉa than thì có 545 vị trí trùng khớp, tương ứng tỷ lệ 72,57 %. Tuy nhiên, xét về mặt cấu trúc, các vỉa được nổi theo thuật toán đơn giản hơn, thoáng hơn và phù hợp hơn với môi trường trầm tích. Một số vị trí mất vỉa than (lỗ khoan SX1901), dự đoán nguyên nhân là do đứt gãy nhưng thực tế đi lò khai thác lại hoàn toàn không gặp, điều này khá trùng hợp với lý giải môi trường kênh rạch ở đây.

#### 4. KẾT LUẬN

Qua công tác nghiên cứu đồng danh, nổi vỉa than ở khu vực mỏ Núi Béo bằng phương pháp hồi quy logistic và mạng trí tuệ nhân tạo (ANN), cho phép rút ra một số kết luận sau:

1. Sơ đồ xác định nổi vỉa mới bằng thuật toán so với nổi vỉa hiện có theo tài liệu thăm dò đảm bảo trùng khớp 72,57 %. Điều này chứng tỏ rằng cấu trúc khu mỏ theo tài liệu cũ đảm bảo mức độ cơ bản, định hướng cho phương pháp nổi vỉa bằng thuật toán.

2. Phương pháp toán dựa trên đặc trưng của vỉa cho nên không đảm bảo 100% nhận dạng đúng, phương pháp hồi quy logistic đạt 70,87 %, phương pháp ANN đạt 84,94 %. Nhận dạng vỉa than bằng phương pháp ANN cao hơn phương pháp hồi quy logistic, tuy nhiên vẫn còn 15,06 % sai số; đây cũng là kết quả khả quan nhất có thể chấp nhận dùng thuật toán trong công tác nổi vỉa.

3. So sánh với phương pháp nổi vỉa truyền thống, đồng danh các vỉa than khu mỏ Núi Béo bằng thuật toán cho thấy các vỉa được nổi đơn giản hơn, thoáng hơn và phù hợp với môi trường trầm tích hơn. Đây cũng là nguyên tắc tối thiểu phải nghiên cứu môi trường thành tạo than khi thực hiện quá trình nổi vỉa.

#### TÀI LIỆU THAM KHẢO

1. Phạm Tuấn Anh (Chủ biên), (2009). Báo cáo chuyển đổi cấp trữ lượng và cấp tài nguyên than khu mỏ Hà Lâm, Lưu trữ tại Tập đoàn Công nghiệp Than - Khoáng sản Việt Nam.
2. Duan H, Xie W, Zhao J, Jia T, (2021). Sequence stratigraphy and coal accumulation model of the Taiyuan Formation in the Tashan Mine, Datong Basin, China. *Open Geosciences*, 13 (1); p. 1259–1272.
3. Einsele G, (2020). Sedimentary Basins, Evolution, Facies, and Sediment Budget. Springer-Verlag Berlin Heidelberg;. Available from: <https://doi.org/10.1007/978-3-662-04029-4>.
4. Freedman DA, (2009). Statistical models: theory and practice. Cambridge University Press. 12 p.
5. Hou H, Shao L, Tang Y, Li Y, Liang G, Xin Y, et al, (2009). Coal seam correlation in terrestrial basins by sequence stratigraphy and its implications for palaeoclimate and palaeoenvironment evolution. *J Earth Sci.*; p. 1–24.
6. Khương Thế Hùng, Nguyễn Danh Tuyên, (2022). Nhận dạng các vỉa than và môi trường trầm tích khu mỏ Núi Béo, Quảng Ninh bằng thuật toán K-means và phương pháp hồi quy. *Tạp chí Khoa học và Phát triển công nghệ, chuyên san Khoa học trái đất và Môi trường* (đã chấp nhận đăng).

7. Khuong The Hung, Nguyen Phuong, Nguyen Thi Cuc, Pham Nhu Sang, Nguyen Danh Tuyen, (2021). Identifying Correlation of Coal Seams in the Tien Hai Area, Northern Vietnam by Using Multivariate Statistic Methods. *Inżynieria Mineralna – Journal of the Polish Mineral Engineering Society*, 2(46), 2021; p. 129–148. Available from: <http://doi.org/10.29227/IM-2021-02-11>.
8. Lê Hùng (Chủ biên), (1996). Bản đồ địa chất và khoáng sản nhóm tờ Hòn Gai - Cẩm Phả, tỷ lệ 1:50.000. Lưu trữ Địa chất - Tổng cục Địa chất và Khoáng sản Việt Nam, Hà Nội.
9. Meng X, Ge M, Tucker ME, (1997). Sequence stratigraphy, sea-level changes and depositional systems in the Cambro-Ordovician of the North China carbonate platform. *Sedimentary Geology*, 114 (1-4); p. 189–222. Available from: [https://doi.org/10.1016/S0037-0738\(97\)00073-0](https://doi.org/10.1016/S0037-0738(97)00073-0).
10. Nalendra S, Kuncoro B, Burhanudin A, (2017). Thickness Variation of Coal Seams in Loa Janan Anticline: Implications for Exploration and Mining Activities. *Proceedings joint convention Malang 2017*, HAGI-IAGI-IAFMI-IATMI (JCM 2017) Ijen Suites Hotel, Malang, September 25-28.
11. Schmidhuber J, (1992). Learning complex, extended sequences using the principle of history compression. *Neural Computation*, 4(2), 234-242.

## **IDENTIFYING THE CORRELATION OF COAL SEAMS IN NUI BEO MINE, QUANG NINH PROVINCE USING LOGISTIC REGRESSION AND ARTIFICIAL INTELLIGENCE METHODS**

**Khuong The Hung<sup>1\*</sup>, Ta Thi Toan<sup>1</sup>, Nguyen Danh Tuyen<sup>2</sup>**

<sup>1</sup>Hanoi University of Mining and Geology 18 Vien street, Duc Thang, Bac Tu Liem, Ha Noi

<sup>2</sup>Viet Bac Geology Joint Stock Company, 65 An Trach, Dong Da, Ha Noi

\*Email: [khuongthehung@hmg.edu.vn](mailto:khuongthehung@hmg.edu.vn)

### **ABSTRACT**

Nui Beo mine is located in the southern part of the Quang Ninh province. It belongs to the Hon Gai-Cam Pha coal zone with high coal resources exploitation potential. In this study, the data of the coal seam parameters and coal quality were synthesized and processed using artificial intelligence and logistic regression methods. The results allow identifying the correlation of the coal seams in Nui Beo mine. Correlation identifying result using logistic regression methods is 70.87 % while correlation identifying result using artificial intelligence is higher, reaching 84.94 % with 15.06 % errors. This error range is accepted in using mathematical methods for identifying correlation of the coal seams. In comparison with traditional methods, the identifying correlation of coal seams using artificial intelligence and logistic regression algorithm is more simple and suitable with Nui Beo mine structure. The results confirm that applying mathematical methods in geological science are effective, especially in identifying coal seams correlation and in sedimentary environments.

*Keywords:* correlation of coal seams, logistic regression, artificial neural network, Nui Beo mine, Quang Ninh.