



## Journal of Mining and Earth Sciences

Website: <http://jmes.humg.edu.vn>



# Building a Random Forest predictive modeling of mineral perspectivity and Mapping gold mineral prospects in Tam Ky - Phuoc Son, Quang Nam



Tinh Thanh Bui <sup>1,\*</sup>, Dung Tien Nguyen <sup>1</sup>, Khang Quang Luong <sup>1</sup>, Bac Hoang Bui <sup>1</sup>, Sang Viet Bui <sup>2</sup>

<sup>1</sup> Hanoi University of Mining and Geology, Hanoi, Vietnam

<sup>2</sup> Intergeo Division, Hanoi, Vietnam

### ARTICLE INFO

#### Article history:

Received 23<sup>rd</sup> Apr. 2022

Revised 17<sup>th</sup> Aug. 2022

Accepted 21<sup>st</sup> Sept. 2022

#### Keywords:

Gold,  
Mapping,  
Random Forest,  
Tam Ky - Phuoc Son  
The predictive modeling of  
mineral prospectivity.

### ABSTRACT

*Tam Ky - Phuoc Son area has great potential for gold mineral with 98 gold occurrences, but the evaluation of the entire gold-mineralization potential of the area is still very limited, while this is considered as a basis for planning, exploration, and mining. The paper uses an Artificial Intelligence model which has a name Random Forest to build predictive modeling of mineral perspectivity and to map the gold mineral prospect of the study area. 12 influencing factors are selected to build the dataset for model training and mapping gold minerals prospect, including Geology, fault systems (NE-SW faults, NW-SE faults, sub meridian faults, sub-latitude faults), Bouguer geophysical anomaly, a geochemical anomaly of silver (Ag), gold (Au), lead (Pb), zinc (Zn), copper (Cu) and distance to the geologic boundary of complexes related to gold mineralization. The data which are generated from these factors are 12 fuzzy maps. This data combines with 98 occurrences' locations to create a dataset that is used to train a model of mineral perspectivity using the Random Forest algorithm. After training the model is evaluated by validation. The results of the Random Forest predictive modeling of mineral prospects are well trained with an accuracy of 95.99% on the training set and 83.05 on the validation set, the performance of the model is excellent on both datasets with AUC of 0.993 and 0.95, respectively. Finally, a mineral perspectivity map is built using the trained model. The study area is divided into 3 types of areas: high, medium, and low prospects. The area of high prospect is 982.8 km<sup>2</sup>, covering 71% of the gold occurrences.*

Copyright © 2022 Hanoi University of Mining and Geology. All rights reserved.

\*Corresponding author

E - mail: [buihanhtinh@humg.edu.vn](mailto:buihanhtinh@humg.edu.vn)

DOI: 10.46326/JMES.2022.63(5).08



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

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

# Xây dựng mô hình rừng ngẫu nhiên dự báo tài nguyên và phân vùng triển vọng khoáng sản vàng khu vực Tam Kỳ - Phước Sơn, Quảng Nam

Bùi Thanh Tịnh<sup>1,\*</sup>, Nguyễn Tiến Dũng<sup>1</sup>, Lương Quang Khang<sup>1</sup>, Bùi Hoàng Bắc<sup>1</sup>, Bùi Viết Sáng<sup>2</sup>

<sup>1</sup> Trường Đại học Mỏ - Địa chất, Hà Nội, Việt Nam

<sup>2</sup> Liên đoàn địa chất Intergeo, Hà Nội, Việt Nam

### THÔNG TIN BÀI BÁO

Quá trình:  
 Nhận bài 23/4/2022  
 Sửa xong 17/8/2022  
 Chấp nhận đăng 21/9/2022

### Từ khóa:

Dự báo tài nguyên,  
 Phân vùng triển vọng,  
 Random Forest,  
 Tam Kỳ - Phước Sơn,  
 Vàng.

### TÓM TẮT

Vùng Tam Kỳ - Phước Sơn có tiềm năng lớn khoáng sản vàng với 98 điểm vàng đã được phát hiện, nhưng việc đánh giá toàn bộ tiềm năng khoáng sản vàng trong vùng còn rất hạn chế làm cơ sở quy hoạch, thăm dò, khai thác khoáng sản vàng. Bài báo nghiên cứu sử dụng mô hình rừng ngẫu nhiên (Random Forest) xây dựng mô hình dự báo tài nguyên khoáng sản vàng và phân vùng triển vọng khoáng sản vàng vùng nghiên cứu. 12 yếu tố thành phần được lựa chọn để xây dựng bộ dữ liệu cho huấn luyện mô hình và phân vùng triển vọng khoáng sản vàng bao gồm: địa chất, hệ thống các đứt gãy (ĐBTN, TBĐN, AVT, AKT), dị thường địa vật lý Bughe, dị thường địa hóa bạc (Ag), vàng (Au), chì (Pb), kẽm (Zn), Đồng (Cu) và yếu tố nội suy khoảng cách ranh giới địa chất của các phức hệ địa chất liên quan đến khoáng hóa vàng. Bộ dữ liệu được tạo ra từ các yếu tố thành phần là 12 bản đồ fuzzy. Bộ dữ liệu này kết hợp với vị trí của 98 điểm vàng tạo ra bộ dữ liệu để huấn luyện mô hình dự báo tài nguyên khoáng sản bằng thuật toán Random Forest. Mô hình sau khi huấn luyện được đánh giá bằng tập dữ liệu xác nhận. Kết quả mô hình dự báo tài nguyên khoáng sản Random Forest được huấn luyện tốt với độ chính xác là 95,99% trên tập huấn luyện và 83,05% trên tập dữ liệu xác nhận, hiệu suất của mô hình thực hiện xuất sắc trên cả tập dữ liệu huấn luyện và tập dữ liệu xác nhận, chỉ số AUC lần lượt bằng 0,993 và 0,95. Cuối cùng bản đồ phân vùng triển vọng khoáng sản được thành lập bằng cách sử dụng mô hình dự báo tài nguyên khoáng sản đã được huấn luyện. Phân chia vùng nghiên cứu thành 3 kiểu diện tích là triển vọng cao, trung bình và thấp. Diện tích triển vọng cao là 982,8 km<sup>2</sup> bao phủ 71% số điểm mỏ, khoáng sản vàng đã biết.

© 2022 Trường Đại học Mỏ - Địa chất. Tất cả các quyền được bảo đảm.

\*Tác giả liên hệ

E - mail: [buithanhtinh@humg.edu.vn](mailto:buithanhtinh@humg.edu.vn)

DOI: 10.46326/JMES.2022.63(5).08

## 1. Mở đầu

Theo các công trình nghiên cứu trước đây vùng Tam Kỳ - Phước Sơn, tỉnh Quảng Nam là một trong những khu vực có tiềm năng hàng đầu tại Việt Nam về quặng vàng (Võ, 2018). Trong vòng ba thập kỷ trở lại đây đã có nhiều dự án điều tra, đánh giá cũng như nghiên cứu về đặc điểm quặng hóa vàng trong vùng (Võ, 2018; Hoàng, 2015; Đình, 2014; Đình, 2013; Lê, 2002a; Đỗ, 2001; Lê, 1999; Cát, 1999; Đỗ, 1998; Bùi, 1986). Vùng nghiên cứu đã thống kê được 98 điểm mỏ, điểm quặng vàng gốc, ngoài ra còn nhiều diện tích tập trung sa khoáng vàng quy mô nhỏ nằm rải rác trong vùng. Tuy nhiên, đến thời điểm hiện tại mới thăm dò và tính trữ lượng cho 02 mỏ là mỏ vàng Phước Sơn (từ năm 1996) và mỏ vàng Bồng Miêu (từ năm 1991) với tổng trữ lượng, tài nguyên quặng vàng lần lượt là 8.808 kg và 4.155 kg vàng (Lê, 1985; Phạm, 2001; Lê, 2011, 2002b).

Mặc dù đã có nhiều nhà địa chất nghiên cứu, công nhận tiềm năng khoáng sản vàng, ở vùng Tam Kỳ - Phước Sơn nhưng việc đánh giá tiềm năng, quy mô khoáng sản vàng trên toàn bộ diện tích còn rất hạn chế. Mới có những báo cáo điều tra, đánh giá khoáng sản vàng ở những diện tích nhỏ trong vùng và mới chỉ có 2 diện tích được thăm dò chi tiết. Gần đây nhất, có công trình "Nghiên cứu, dự báo tiềm năng khoáng sản vàng ẩn sâu ở các trường quặng vàng đới Tam Kỳ - Phước Sơn vùng Trung Trung Bộ" của Võ Quảng Bình đã dự báo tiềm năng khoáng sản vàng ẩn sâu cho toàn vùng nghiên cứu. Tuy nhiên, việc khoanh định các diện tích triển vọng vẫn thực hiện thủ công bằng tay theo các nguyên tắc khoanh định truyền thống (Võ, 2018). Do đó, việc nghiên cứu khoanh định các diện tích triển vọng, nâng cao độ chính xác trong đánh giá toàn diện tiềm năng khoáng sản vàng vùng nghiên cứu là cần thiết phục vụ công tác quy hoạch, thăm dò, khai thác khoáng sản.

Trong những năm gần đây, trên thế giới phát triển các phương pháp máy học, mô hình trí tuệ nhân tạo xây dựng mô hình dự báo về tiềm năng khoáng sản, sử dụng mô hình đó để phân vùng triển vọng khoáng sản đã được công bố trên các tạp chí quốc tế uy tín như các công trình của Carranza và cộng sự (2015a). Mô hình phân tích và tổng hợp các lớp dữ liệu không gian có liên quan đến đối tượng nghiên cứu thu được từ các bộ

dữ liệu GIS để phân định và xếp hạng các khu vực có triển vọng thăm dò, các mỏ khoáng sản được tìm kiếm (Bonham-Carter, 2014). Trong đó, nổi bật có mô hình thuật toán rừng ngẫu nhiên (Random Forest - RF) đang được sử dụng phổ biến trong việc lập bản đồ dự đoán các khu vực được quan tâm theo các tiêu chí nhất định. Ví dụ, áp dụng (RF) để dự đoán các khu vực thích hợp cho sản xuất lúa (Laborte và nnk., 2012); phân loại lớp phủ đất (Rodriguez-Galiano và nnk., 2012), lập bản đồ phân bố loài (Bradter và nnk., 2013). Các ứng dụng của RF tích hợp dữ liệu không gian để lập bản đồ địa chất cũng đang phát triển bao gồm các công trình nghiên cứu của Cracknell và Anya (2013; 2014). Công trình của Reddy và Bonham-Carter (1991) đã chứng minh việc sử dụng mô hình RF để lập bản đồ khả năng sinh khoáng cho các mỏ kim loại gốc ở khu vực Hồ Tuyết ở Manitoba (Canada) (Reddy và Bonham-Carter, 1991). Mới gần đây, Rodriguez-Galiano và cộng sự (2014) đã chứng minh khả năng ứng dụng của RF vào mô hình dự báo về tiềm năng của khoáng sản bằng cách sử dụng 49 vị trí các mỏ vàng gốc ở quận Rodalquilar, Tây Ban Nha (Rodriguez-Galiano và nnk., 2014) và công trình của Carranza và Alice (2015b) lập bản đồ dự đoán tiềm năng khoáng sản vàng ở quận Baguio (Philippines) dựa trên 19 các vị trí điểm. Tuy nhiên, việc ứng dụng các phương pháp sử dụng mô hình Random Forest vào lĩnh vực địa chất khoáng sản ở Việt Nam còn rất mới mẻ và chưa phổ biến.

Trong công trình này, nhóm tác giả đã nghiên cứu sử dụng mô hình RF để xây dựng mô hình dự báo tài nguyên khoáng sản vàng và phân vùng triển vọng khoáng sản vàng vùng Tam Kỳ - Phước Sơn. Nghiên cứu này đã đề xuất phương pháp mới tiếp cận theo xu thế phát triển của thế giới, góp phần hoàn thiện các phương pháp tìm kiếm khoáng sản và bổ sung dữ liệu cho lĩnh vực thăm dò khoáng sản.

## 2. Mô hình Random Forest

Random Forest (RF) là phương pháp phân lớp thuộc tính với thuật toán học có giám sát (supervised learning), có thể được sử dụng cho cả hai bài toán phân lớp và hồi quy. Đây là thuật toán linh hoạt và dễ sử dụng. Thuật toán bao gồm nhiều cây quyết định, cây quyết định được tạo ra dựa trên các mẫu dữ liệu được chọn ngẫu nhiên, việc

dự đoán được thực hiện từ mỗi cây và chọn ra kết quả cuối cùng bằng cách bỏ phiếu.

Thuật toán RF hoạt động theo 4 bước (Breiman, 2015):

Bước 1. chọn số lượng các cây thành phần sẽ được xây dựng từ tập dữ liệu đã cho;

Bước 2. chọn ngẫu nhiên số lượng các thuộc tính sẽ được dùng để phân chia tại mỗi node của cây, thường nhỏ hơn tổng số các thuộc tính có trong tập mẫu. Giá trị số lượng các thuộc tính này được giữ không đổi trong suốt quá trình dựng cây;

Bước 3. thiết lập cây quyết định cho từng tập mẫu nhỏ và nhận kết quả dự đoán từ mỗi quyết định cây. Mỗi cây được bỏ phiếu bình chọn cho kết quả cuối cùng;

Bước 4. sau khi xây dựng được RF, để phân lớp cho đối tượng, thu thập kết quả bình chọn cho đối tượng phân loại trên tất cả các cây quyết định và sử dụng kết quả được bình chọn nhiều nhất làm kết quả cuối cùng của thuật toán.

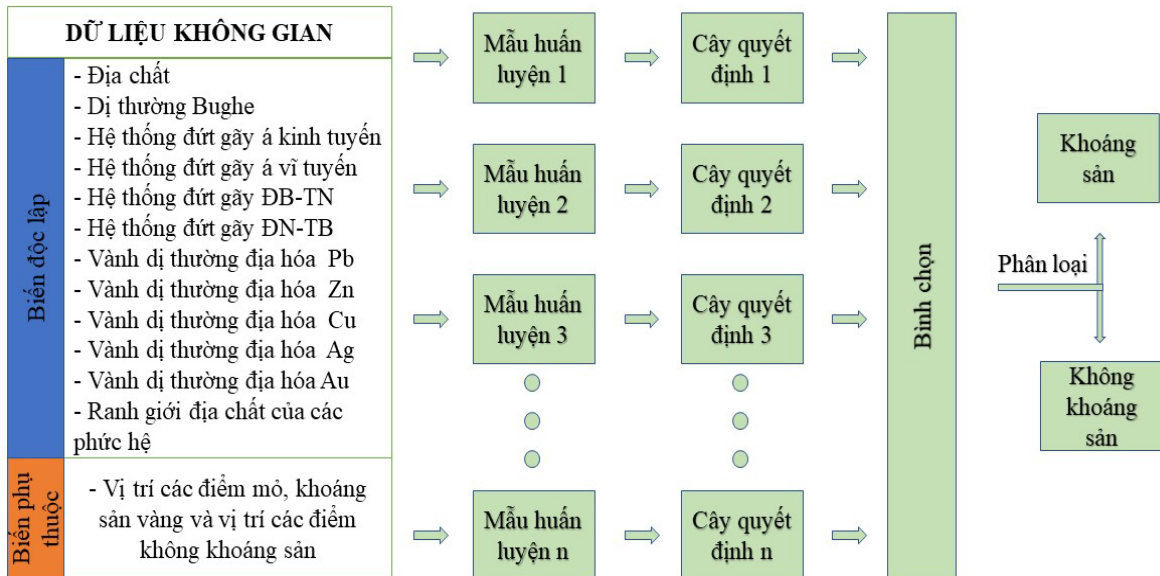
Trên cơ sở thuật toán RF, xây dựng mô hình dự báo tài nguyên khoáng sản từ bộ dữ liệu không gian bao gồm 12 biến độc lập và 1 biến phụ thuộc được mô tả trong mục 3.3. Bộ dữ liệu không gian được chia thành 2 tập dữ liệu là tập dữ liệu huấn luyện và tập dữ liệu xác nhận theo tỷ lệ tương ứng là 70% và 30%, được sử dụng để huấn luyện mô hình. Trực quan hóa phương pháp luận xây dựng mô hình RF dự báo tài nguyên khoáng sản như Hình 1.

### 3. Ứng dụng mô hình RF dự báo phân vùng khoáng sản vàng cho vùng nghiên cứu

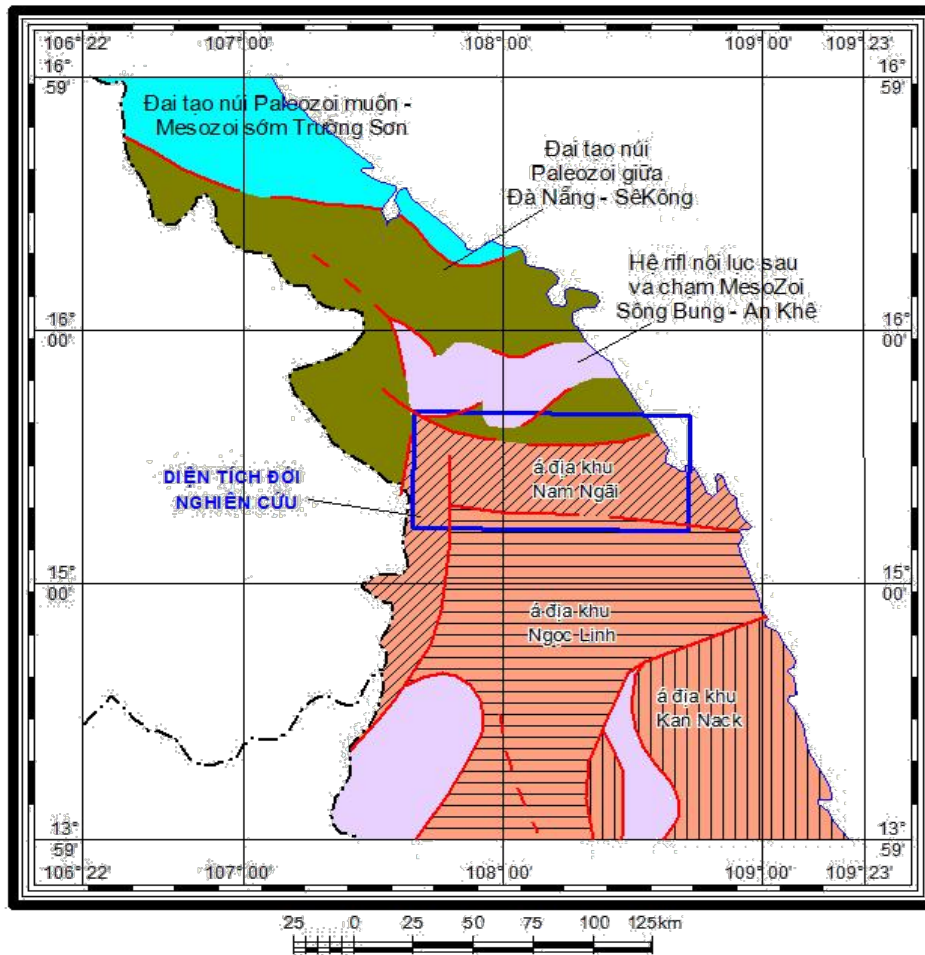
#### 3.1. Khái quát về khu vực nghiên cứu

Vùng nghiên cứu thuộc rìa bắc địa khu biến chất cao Kon Tum, là một phần diện tích của á địa khu Nam - Ngãi với các đá nguyên sinh tuổi Proterozoi - Paleozoi sớm thuộc khối kiến trúc Khâm Đức; có phương á vĩ tuyến với bề rộng của đới khoảng 60 km, dài 120 km, diện tích khoảng 7.020 km<sup>2</sup>. Được giới hạn bởi các điểm góc có tọa độ địa lý 15°08'15" ÷ 15°40'15" độ vĩ bắc; 107°34'50" ÷ 108°42'30" độ kinh đông (Hình 2).

Tham gia vào cấu trúc địa chất bao gồm chủ yếu các thành tạo trầm tích biến chất thuộc hệ tầng Đắc My (MP<sub>đm</sub>), phức hệ Khâm Đức - Núi Vú (NP<sub>3</sub> - E<sub>1kv</sub>), hệ tầng A Vương (E<sub>2</sub>-O<sub>1av</sub>), hệ tầng Suối Cát (O-Ssc). Các đá trầm tích biến chất này bị biến dạng mạnh tạo nên các vi uốn nếp và bị xuyên cắt bởi các đá xâm nhập mafic và siêu mafic gabroamphibolit, gabroproxenit, apodunit, apoharzburgit thuộc phức hệ Tà Vi (vNP-PZ<sub>1tv</sub>), phức hệ Hiệp Đức - Plei Weik (σPZ<sub>1hp</sub>); diorit, granodiorit, tonalit, granit hornblend-biotit, granit biotit, granosyenit thuộc phức hệ Bến Giằng - Quế Sơn (γδP<sub>2-3bq</sub>), phức hệ Diên Bình - Trà Bồng (γδO<sub>3</sub>-S<sub>1dt</sub>), phức hệ Đại Lộc (gyS<sub>4</sub>-D<sub>1đl</sub>), phức hệ



Hình 1. Khái niệm phương pháp luận xây dựng mô hình RF dự báo tài nguyên khoáng sản.



Hình 2. Vị trí vùng nghiên cứu trong bình đồ cấu trúc khu vực (Trần và Vũ, 2009).

Hải Vân ( $\gamma P_3 - T_1, hv$ ), phức hệ Măng Xim ( $\gamma T_2 mx$ ) và các đai mạch lamprophyr kiềm địa khu Kon Tum (Hình 3a).

Các hệ thống đứt gãy phương tây bắc - đông nam, đông bắc - tây nam đôi khi chuyển sang á vĩ tuyến, phát triển khá mạnh và được coi là hệ đứt gãy khổng lồ chế quặng vàng (Võ, 2018). Các mỏ quặng vàng gốc, điểm quặng, điểm khoáng vàng hầu hết bám dọc theo đới dập vỡ phá huỷ của hệ thống đứt gãy này. Trong phạm vi nghiên cứu còn nhiều khu vực có cấu trúc địa chất phức tạp, với các tiền đề, dấu hiệu liên quan đến khoáng hóa vàng.

### 3.2. Đặc điểm khoáng sản vàng

Qua các công trình nghiên cứu trước đây, kết hợp cùng với kết quả phân tích mẫu thạch học, ronghen, thạch học đá biến đổi,... (Võ, 2018) có thể thấy về mặt không gian, khoáng hoá vàng trong đới nghiên cứu phân bố trong nhiều thành

tạo khác nhau như các đá biến chất của hệ tầng Khâm Đức, hệ tầng A Vương, hệ tầng Núi Vú, hệ tầng Đăk My, các thành tạo magma xâm nhập phức hệ Chu lai, Bà Nà, Trà Bồng. Nhưng chiếm đa số là khoáng hoá vàng nằm trong hệ tầng Khâm Đức.

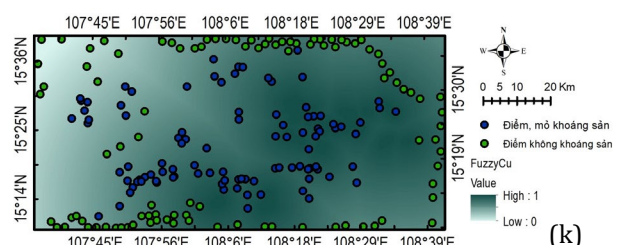
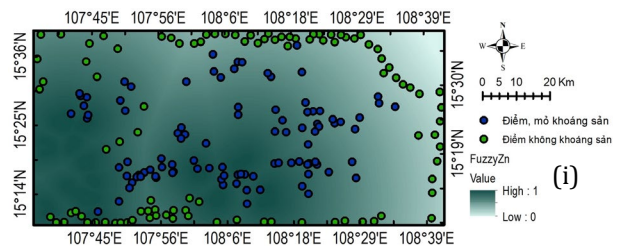
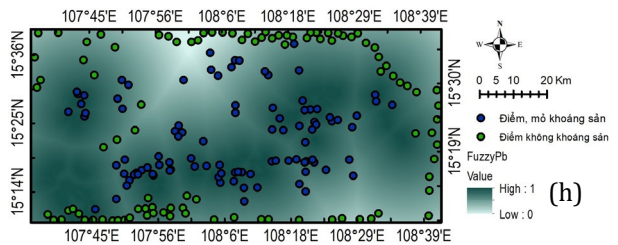
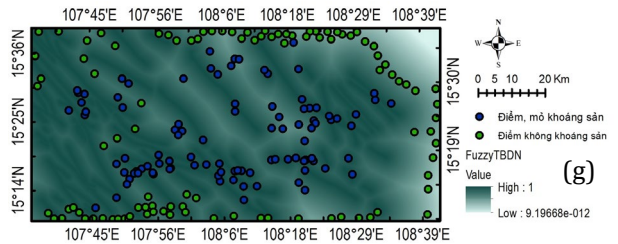
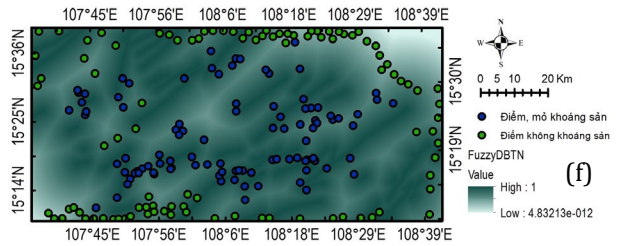
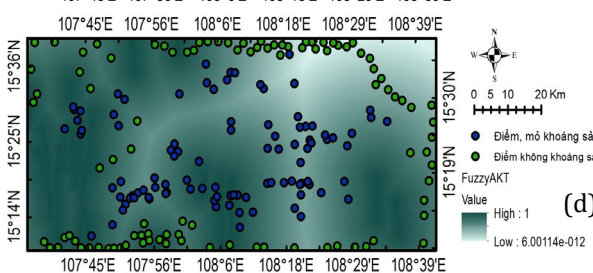
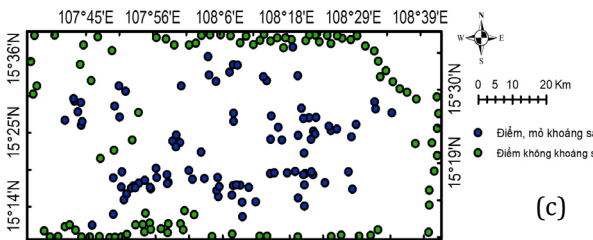
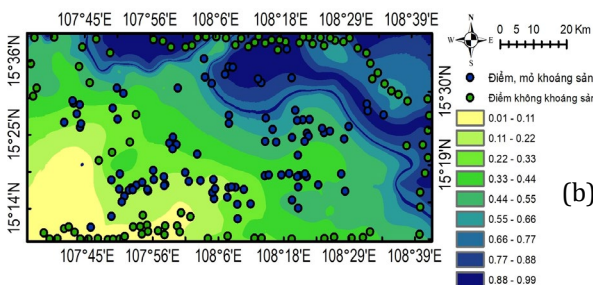
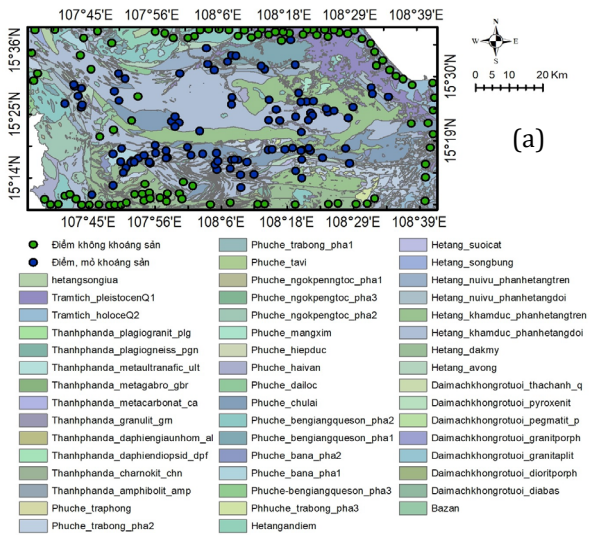
Các thân quặng vàng trong đới nghiên cứu có dạng mạch nhỏ, đới vi mạch thạch anh - sulfur chứa vàng, dạng ổ, thấu kính nhỏ, dạng mạng mạch. Chúng phân bố theo các đới dập vỡ, cà nát của các hệ thống khe nứt, đứt gãy phương tây bắc - đông nam, phương đông bắc - tây nam, á kinh tuyến.

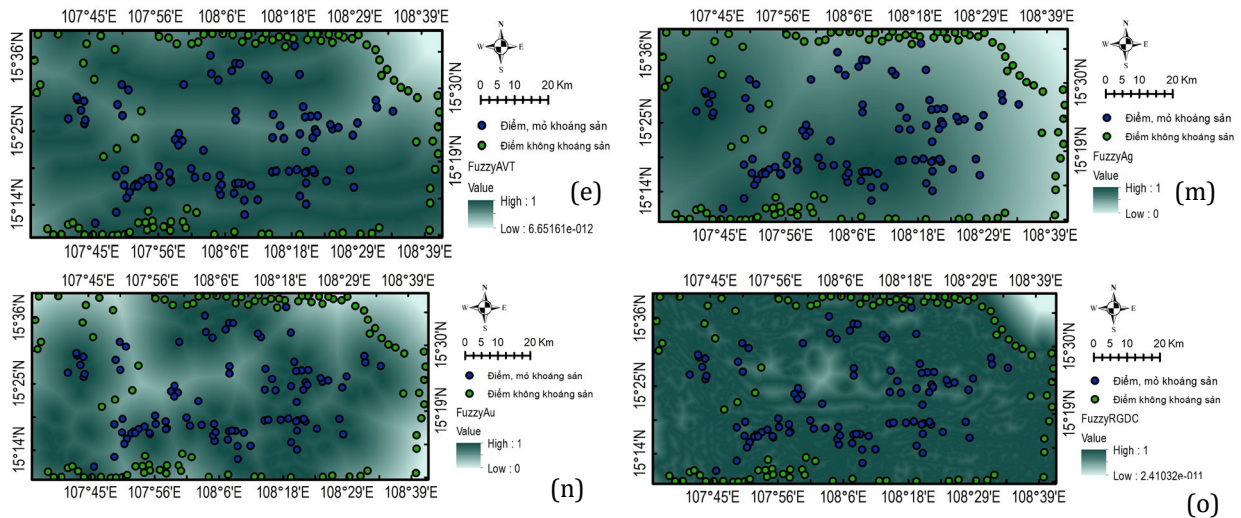
Phần lớn các thân quặng trong đới nghiên cứu có phương phát triển theo hướng đông bắc - tây nam, kéo dài 50÷800 m, góc dốc tương đối thoải 20÷40°, một số thân quặng dốc đứng đến 70÷80°. Nhìn chung, trung bình là 40÷60°. Quặng vàng trong vùng nghiên cứu có hàm lượng vàng 1,2÷8,2 g/T chiếm đa số, còn dao động lớn 0,6÷41,0 g/T,

có mẫu đạt đến 74,0 g/T. Quặng có cấu tạo dạng xâm tán. Khoáng vật quặng chủ yếu là pyrit, galena, sphalerit, chalcopyrit, arsenopyrit. Vàng tồn tại dưới dạng vàng tự sinh và electrum. Quặng vàng trong vùng nghiên cứu chủ yếu thuộc kiểu quặng vàng - thạch anh - sulfur đa kim, ít hơn là kiểu quặng vàng - thạch anh - pyrit.

### 3.3. Dữ liệu không gian

Từ những mô tả đặc điểm khoáng sản trên có thể thiết lập được các yếu tố địa chất liên quan đến khống chế tồn tại khoáng sản vàng là tiền đề, dấu hiệu địa chất tìm kiếm, điều tra khoáng sản vàng vùng Tam Kỳ - Phước Sơn, đây là cơ sở để xây dựng bộ dữ liệu cho huấn luyện mô hình dự báo khoáng sản.





Hình 3. Địa chất (a); dị thường Bughe (b); sự phân bố của các điểm quặng (c); á kinh tuyến (d); á vĩ tuyến (e); hệ thống đứt gãy theo phương đông bắc-tây nam (f); đông nam-tây bắc (g); vành dị thường địa hóa Pb(h); vành dị thường địa hóa Zn (i); vành dị thường địa hóa Cu (k); vành dị thường địa hóa Ag (m); vành dị thường địa hóa Au (n); yếu tố nội suy khoảng cách ranh giới địa chất của các phức hệ địa chất liên quan đến khoáng hóa vàng (o).

Yếu tố địa chất, khoáng hóa vàng nằm trong các thể địa chất khác nhau phân bố trong diện tích nghiên cứu, đây là yếu tố liên quan đến thời gian thành tạo vàng.

Yếu tố đứt gãy, dọc theo các đứt gãy có các đới dập vỡ là nơi tích tụ các khoáng sản vàng. Các hệ thống đứt gãy theo các phương khác nhau có mức độ tích tụ khoáng sản vàng khác nhau.

Yếu tố địa hóa, trong quặng vàng có các nguyên tố cộng sinh với vàng, đây là dấu hiệu cho việc tìm kiếm khoáng sản vàng. Trong vùng nghiên cứu các nhà địa chất đã khoan định được các vành phân tán địa hóa của các nguyên tố này.

Yếu tố địa vật lý được sử dụng để tìm kiếm khoáng sản và phát hiện các thể địa chất dưới bề mặt trái đất. Mỗi thể địa chất phản xạ lại các trường địa vật lý khác nhau, đây là cơ sở nhận biết đối tượng địa chất ở dưới sâu.

Trên cơ sở các yếu tố liên quan đến khoáng hóa vàng vùng nghiên cứu 12 yếu tố thành phần được lựa chọn trong nghiên cứu này, bao gồm: địa chất (Hình 3a), hệ thống đứt gãy theo phương đông bắc - tây nam (Hình 3f), đông nam - tây bắc (Hình 3g), á kinh tuyến (Hình 3d), Á vĩ tuyến (Hình 3e), dị thường Bughe (Hình 3b), vành dị thường địa hóa Ag (Hình 3m), vành dị thường địa hóa Au (Hình 3n), vành dị thường địa hóa Pb (Hình 3h), vành dị thường địa hóa Zn (Hình 3i), vành dị

thường địa hóa Cu (Hình 3k) và yếu tố nội suy khoảng cách ranh giới địa chất của các phức hệ địa chất liên quan đến khoáng hóa vàng (Hình 3o). Đây gọi là các biến thành phần dự báo và được xây dựng thành cơ sở dữ liệu GIS, thể hiện dưới dạng các bản đồ thành phần với tên tương ứng như trong Hình 3. Ngoài ra còn có dữ liệu các điểm mỏ, điểm khoáng sản vàng được gọi là biến mục tiêu (Hình 3c).

Trong nghiên cứu này, toàn bộ dữ liệu được xây dựng mới hoàn toàn để phù hợp với phương pháp sử dụng, không trùng lặp với nghiên cứu nào trước đây. Việc thu thập dữ liệu, xử lý và tổng hợp được thực hiện bằng các phần mềm ArcGIS (phiên bản 10.4) và Mapinfo (phiên bản 15.01). Các công việc gồm có số hóa bản đồ từ file dữ liệu ảnh bằng phần mềm Mapinfo. Thành lập bộ cơ sở dữ liệu gồm điểm, đường, vùng theo từng bản đồ thành phần bằng phần mềm arc catalog. Gán dữ liệu thuộc tính cho các đối tượng địa chất và thành lập các bản đồ Fuzzy-map thành phần theo các yếu tố địa chất tương ứng. Bản đồ sau đó được chuyển đổi sang dạng dữ liệu raster với kích thước ô pixel là 30x30 m. Đây là bộ cơ sở dữ liệu được sử dụng để xây dựng mô hình dự báo tài nguyên khoáng sản và phân vùng triển vọng khoáng sản vàng vùng Tam Kỳ - Phước Sơn.

## 4. Kết quả và thảo luận

### 4.1. Đánh giá độ chính xác

Mô hình RF dự báo tài nguyên khoáng sản được đào tạo bằng cách sử dụng tập dữ liệu huấn luyện gồm 12 biến thành phần. Từ bảng kết quả đào tạo cho thấy mô hình đã thực hiện rất tốt với tập dữ liệu huấn luyện. Theo Bảng 1, có thể thấy mức độ chính xác của mô hình được huấn luyện bởi tập dữ liệu huấn luyện rất cao là 95,99%. Mức độ phù hợp của mô hình và bộ dữ liệu đào tạo là tốt ở mức 0.92 (Kappa) với sai số trung phương thấp bằng 0,1938%. Giá trị độ nhạy phản ánh tỷ lệ phần trăm của các pixel khoáng sản được phân chia chính xác đạt 96,3%; giá trị độ chuyên phản ánh tỷ lệ phần trăm cho các pixel không khoáng sản được phân chia chính xác là 95,68%. Tỷ lệ dương thực là xác suất phân loại pixel của mô hình đối với lớp khoáng sản rất cao ở mức 95,59%; tỷ lệ âm thực là xác suất phân loại pixel của mô hình đối với lớp không khoáng sản là 96,38%.

Bảng 1. Các thông số đánh giá độ chính xác của mô hình RF được huấn luyện.

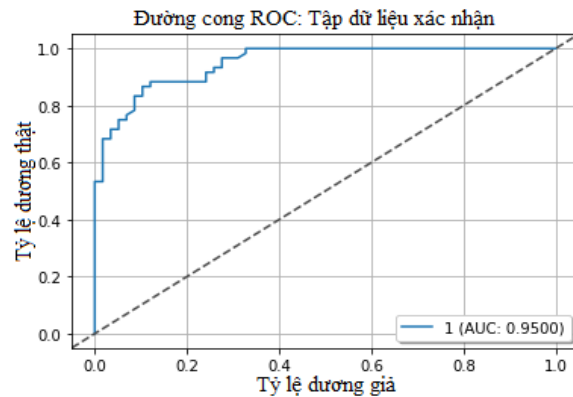
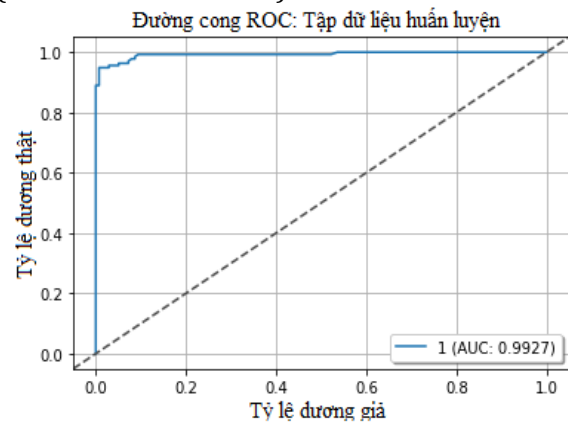
Các chỉ số thống kê	Tập dữ liệu huấn luyện	Tập dữ liệu xác nhận
Dương thực	130	53
Âm thực	133	45
Dương giả	6	7
Âm giả	5	13
Tỉ lệ dương thực (%)	95,59	88,33
Tỉ lệ âm thực (%)	96,38	77,59
Độ nhạy (%)	96,3	80,30
Độ chuyên (%)	95,68	86,54
Độ chính xác (%)	95,99	83,05
Sai số trung phương (%)	0,1938	0,3
Kappa	0.92	0.66
AUC	0.993	0,95

Sau khi mô hình dự báo triển vọng khoáng sản được đào tạo với tập dữ liệu huấn luyện, mô hình này được tiếp tục đánh giá với tập dữ liệu xác nhận và kết quả được hiển thị cùng trong Bảng 1, cho thấy độ chính xác của mô hình là khá cao bằng 83,05%. Kappa của mô hình là 0,66; sai số trung phương thấp bằng 0,3. Tỷ lệ dương thực là 88,33%; tỷ lệ âm thực là 77,59%. Tỷ lệ các pixel khoáng sản được dự đoán chính xác là 80,30%; tỷ

lệ pixel không khoáng sản được mô hình dự đoán chính xác là 86,54%.

### 4.2. Hiệu suất của mô hình

Khả năng dự đoán của mô hình dự báo tài nguyên khoáng sản được đo bằng đường cong ROC và trị số AUC (Hình 4). Kết quả AUC của mô hình trong tập dữ liệu huấn luyện là 0.993 và trong tập dữ liệu xác nhận là 0,95. Từ kết quả trên có thể kết luận rằng mô hình RF dự báo tài nguyên khoáng sản có thể thực hiện xuất sắc việc dự báo các vị trí có khoáng sản cho khu vực nghiên cứu này theo như phạm loại chỉ số AUC của Cantor (Cantor và Kattan 2000).

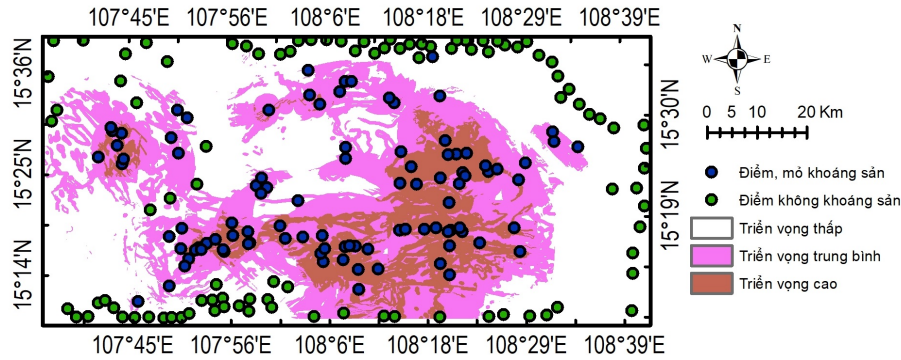


Hình 4. Đường cong ROC của mô hình: tập dữ liệu huấn luyện và tập dữ liệu xác nhận.

### 4.3. Xây dựng bản đồ phân vùng triển vọng khoáng sản vàng vùng Tam Kỳ - Phước Sơn

Mô hình dự báo tài nguyên khoáng sản đã được huấn luyện theo mô hình RF sử dụng tập dữ liệu huấn luyện để tính toán các chỉ số độ nhạy cảm xuất hiện khoáng sản cho khu vực nghiên cứu. Tất cả các bản đồ thành phần đã được chuyển





Hình 5. Bản đồ phân vùng triển vọng vàng vùng Tam Kỳ - Phước Sơn.

đổi sang định dạng raster và sau đó được đưa vào mô hình dự báo tài nguyên khoáng sản để tạo ra các chỉ số nhạy cảm, được gọi là chỉ số xác suất khoáng sản. Các chỉ số này được phân loại dựa trên mức độ ảnh hưởng của các yếu tố thành phần đến khả năng xuất hiện khoáng sản. Cuối cùng, bản đồ phân vùng triển vọng khoáng sản vàng vùng Tam Kỳ - Phước Sơn được xây dựng bởi một loạt các chỉ số xác suất khoáng sản như Hình 5. Cụ thể, dữ liệu để xây dựng bản đồ phân vùng triển vọng khoáng sản vàng là 12 bản đồ dưới dạng raster, được chuyển hóa thành 1 ma trận 2 chiều ( $N \times K$ ) với  $N$  - số điểm pixel trên một bản đồ và  $K$  - số tính chất bằng 13 tương ứng với 12 biến độc lập và 1 biến phụ thuộc. Các bản đồ này có kích thước tương ứng với diện tích vùng nghiên cứu. Một bản đồ raster là tập hợp của 1 chuỗi các pixel, mỗi ô pixel ở đây đã được xác lập kích thước là  $30 \times 30$  m như trình bày ở mục 3.3. Bộ dữ liệu huấn luyện và xác nhận được xây dựng trên cơ sở vị trí 98 điểm mỏ, khoáng sản vàng thành 1 ma trận có kích thước ( $N' \times K$ ),  $N'$  bằng số lượng các pixel vị trí của 98 điểm mỏ, khoáng sản. Mô hình dự báo tài nguyên khoáng sản vàng được huấn luyện bằng bộ dữ liệu này để nhận biết đúng vị trí có khoáng sản và vị trí không có khoáng sản vàng. Mô hình huấn luyện thành công được sử dụng để chạy dự đoán các pixel có khoáng sản cho toàn bộ vùng nghiên cứu. Đây là điểm khác biệt lớn của phương pháp này so với các phương pháp nghiên cứu truyền thống, có tính ưu việt và khách quan hơn trên cơ sở đánh giá độ chính xác và hiệu suất của mô hình như đã trình bày ở trên.

Bản đồ kết quả phân vùng triển vọng khoáng sản vàng vùng Tam Kỳ - Phước Sơn được chia làm 3 vùng diện tích, cụ thể: diện tích triển vọng thấp chiếm 56% tổng diện tích vùng nghiên cứu, diện

tích triển vọng trung bình chiếm 30% và diện tích triển vọng cao chiếm 14%. Có thể thấy, diện tích triển vọng cao bao phủ 70 trong tổng số 98 điểm mỏ vàng đã biết trong vùng nghiên cứu (chiếm 71%), khu vực này có diện tích vào khoảng  $982,8 \text{ km}^2$ .

## 5. Kết luận

Mặc dù đã có nhiều nghiên cứu ở những diện tích nhỏ trong vùng nghiên cứu, tuy nhiên công trình nghiên cứu đánh giá tổng thể tiềm năng khoáng sản của toàn vùng còn rất hạn chế. Đặc biệt chưa có công trình nào nghiên cứu sử dụng các phương pháp nghiên cứu hiện đại có độ chính xác cao, đang phát triển trên thế giới để thực hiện đánh giá này. Trong nghiên cứu này, nhóm tác giả đã nghiên cứu sử dụng mô hình RF xây dựng thành công mô hình dự báo tài nguyên khoáng sản vàng. Kết quả của nghiên cứu này cho thấy mô hình được xây dựng rất tốt, với độ chính xác là 95,99% trong tập dữ liệu huấn luyện và 83,05% trong tập dữ liệu xác nhận. Mô hình RF được đề xuất thực hiện tốt với tập dữ liệu huấn luyện và khá tốt với tập dữ liệu xác nhận có chỉ số ( $Kappa$ ) lần lượt là 0,92 và 0,66. Hiệu suất thực hiện của mô hình xuất sắc trên cả tập dữ liệu huấn luyện và tập dữ liệu xác nhận AUC bằng 0,993 và 0,95.

Mô hình sau khi huấn luyện thành công được sử dụng xây dựng bản đồ phân vùng triển vọng khoáng sản vàng vùng Tam Kỳ - Phước Sơn. Với 3 kiểu diện tích là diện tích triển vọng thấp, diện tích triển vọng trung bình và diện tích triển vọng cao. Trong đó, diện tích triển vọng cao bao phủ 70 trong tổng số 98 điểm mỏ, khoáng sản vàng đã biết trong vùng nghiên cứu (chiếm 71% số điểm vàng đã biết) và có diện tích chiếm 14% trên tổng diện tích vùng nghiên cứu, bằng khoảng  $982,8 \text{ km}^2$ .

## 6. Kiến nghị

Công trình đã nghiên cứu sử dụng mô hình RF xây dựng thành công mô hình dự báo tài nguyên khoáng sản vàng từ đó thành lập ra bản đồ phân vùng triển vọng khoáng sản vàng cho khu vực Tam Kỳ - Phước Sơn. Đây là cơ sở cho việc tính toán dự báo tài nguyên khoáng sản vùng nghiên cứu. Nhóm tác giả đề xuất thực hiện nghiên cứu sâu hơn nữa bao gồm: thực địa kiểm tra kết quả nghiên cứu từ công trình này; nghiên cứu lựa chọn phương pháp dự báo tiềm năng khoáng sản và xác định các thông số tính trữ lượng để tính được tiềm năng tài nguyên của vùng nghiên cứu; thu thập thêm các tài liệu địa chất, địa vật lý liên quan đến tìm kiếm khoáng hóa vàng để xây dựng thêm bộ dữ liệu không gian giúp nâng cao độ chính xác của mô hình dự báo tài nguyên khoáng sản.

## Lời cảm ơn

Bài báo được hoàn thành theo kết quả nghiên cứu của đề tài cơ sở, mã số T21-27. Thông qua bài báo, tác giả xin gửi lời cảm ơn tới Bộ môn Tìm kiếm - Thăm dò, Trường đại học Mỏ - Địa chất đã tạo điều kiện để công trình khoa học hoàn thành đúng tiến độ và chất lượng.

## Đóng góp của các tác giả

Bùi Thanh Tịnh - lên kế hoạch, xây dựng cơ sở dữ liệu, xử lý số liệu, viết bài; Nguyễn Tiến Dũng, Lương Quang Khang, Bùi Hoàng Bắc - góp ý, chỉnh sửa nội dung; Bùi Viết Sáng - thu thập số liệu và các tài liệu, báo cáo nghiên cứu khoáng sản vàng trong vùng Tam Kỳ - Phước Sơn.

## Tài liệu tham khảo

- Bonham-Carter, G. F. (2014). *Geographic information systems for geoscientists: modelling with GIS*. No. 13. Elsevier.
- Breiman, L. (2015). Random forests leo breiman and adele cutler. *Random Forests-Classification Description*, 106.
- Breiman, L. (2002). Manual on setting up, using, and understanding random forests v3. 1. *Statistics Department University of California Berkeley, CA, USA*, 1(58), 3-42.
- Bradter, U., Kunin, W. E., Altringham, J. D., Thom, T. G., & Benton, T. G. (2013). Identifying appropriate spatial scales of predictors in

species distribution models with the random forest algorithm. *Methods in Ecology Benton and Evolution*, 4(2), 167-174.

- Bùi, M. (1986). *Báo cáo tìm kiếm chung khoáng sản vàng vùng Tam Kỳ - Tiên Phước, Quảng Nam - Đà Nẵng*. Trung tâm thông tin, lưu trữ và tạp trí địa chất, Tổng cục Địa chất và Khoáng sản Việt Nam.
- Cantor, S. B., & Kattan, M. W. (2000). Determining the area under the ROC curve for a binary diagnostic test. *Medical Decision Making*, 20, 468-470.
- Carranza, E. J. M., & Alice, G. (2015a). Data-driven predictive mapping of gold prospectivity, Baguio district, Philippines: Application of Random Forests algorithm. *Ore Geology Reviews Laborte*, 71, 777-787.
- Carranza, E. J. M., & Alice, G. (2015b). Random forest predictive modeling of mineral prospectivity with small number of prospects and data with missing values in Abra (Philippines). *Computers Laborte and Geosciences*, 74, 60-70.
- Cracknell, M. J., & Anya, M. (2013). The upside of uncertainty: Identification of lithology contact zones from airborne geophysics and satellite data using random forests and support vector machines. *Geophysics Reading*, 78(3), WB113-WB126.
- Cracknell, M. J., & Anya, M. (2014). Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. *Computers Reading and Geosciences*, 63, 22-33.
- Cát, N. H. (1999). *Báo cáo đo vẽ bản đồ địa chất và tìm kiếm khoáng sản Nhóm từ Tam Kỳ- Hiệp Đức, tỷ lệ 1:50.000*. Trung tâm thông tin, lưu trữ và tạp trí địa chất, Tổng cục Địa chất và Khoáng sản Việt Nam.
- Đinh, Đ. H. (2013). *Báo cáo thăm dò vàng gốc tại khu vực Hồ Ráy, Mỏ vàng Bồng Miêu, xã Tam Lãnh, huyện Phước Ninh, tỉnh Quảng Nam*. Trung tâm thông tin, lưu trữ và tạp trí địa chất, Tổng cục Địa chất và Khoáng sản Việt Nam..

- Đình, T. H. (2014). Báo cáo điều tra đánh giá bổ sung, xác định tài nguyên còn lại mỏ vàng gốc khu vực khe 39, xã Phước Hòa, huyện Phước Sơn, tỉnh Quảng Nam. Trung tâm Thông tin, lưu trữ và Tạp chí địa chất, Tổng cục Địa chất và Khoáng sản Việt Nam.
- Đỗ, Q. B. (2001). Báo cáo đánh giá tiềm năng và các khoáng sản khác liên quan với các thành tạo đá lục, đá phiến đen đối Quảng Nam. Trung tâm Thông tin, lưu trữ và tạp trí địa chất, Tổng cục Địa chất và Khoáng sản Việt Nam.
- Đỗ, V. C. (1998). Báo cáo địa chất và khoáng sản tỷ lệ 1:50.000 nhóm tờ Dakglei - Khâm Đức. Trung tâm thông tin, lưu trữ và tạp trí địa chất, Tổng cục Địa chất và Khoáng sản Việt Nam.
- Hoàng, Q. L. (2015). Báo cáo kết quả điều tra đánh giá bổ sung xác định tài nguyên còn lại mỏ vàng gốc khu vực thôn 1, xã Tiên lập, huyện Tiên Phước, tỉnh Quảng Nam. Trung tâm thông tin, lưu trữ và tạp trí địa chất, Tổng cục Địa chất và Khoáng sản Việt Nam.
- Laborte, A. G., Maunahan, A. A., & Hijmans, R. J. (2012). Opportunities for expanding paddy rice production in Laos: spatial predictive modeling using Random Forest. *Journal of Land Use Science Hijmans*, 7(1), 21-33.
- Livingston, F. (2005). Implementation of Breiman's random forest machine learning algorithm. *ECE591Q Machine Learning Journal Paper Livingston*, 1-13.
- Lê, Đ. H. (1985). Báo cáo đánh giá triển vọng mỏ vàng gốc Bồng Miêu, Quảng Nam - Đà Nẵng. Trung tâm thông tin, lưu trữ và tạp trí địa chất, Tổng cục Địa chất và Khoáng sản Việt Nam.
- Lê, V. Đ. (2002a). Báo cáo kết quả đánh giá khoáng sản vàng và các khoáng sản khác đi cùng vùng Trà Núi, Trà Thủy, Quảng Nam. Trung tâm thông tin, lưu trữ và tạp trí địa chất, Tổng cục Địa chất và Khoáng sản Việt Nam.
- Lê, V. H. (2002b). Báo cáo kết quả thăm dò địa chất khu vực Bãi Đất và Bãi Gõ mỏ vàng Đăk Sa, huyện Phước Sơn, tỉnh Quảng Nam. Trung tâm thông tin, lưu trữ và tạp trí địa chất, Tổng cục Địa chất và Khoáng sản Việt Nam.
- Lê, V. H. (1999). Báo cáo kết quả thăm dò vàng khu vực Tiên Hà, huyện Tiên Phước, Hiệp Đức, Quế Sơn, tỉnh Quảng Nam. Trung tâm thông tin, lưu trữ và tạp trí địa chất, Tổng cục Địa chất và Khoáng sản Việt Nam.
- Lê, V. H. (2011). Báo cáo thăm dò vàng gốc khu vực Phước Sơn, xã Phước Đức và xã Phước Xuân, huyện Phước Sơn, tỉnh Quảng Nam. Trung tâm thông tin, lưu trữ và tạp trí địa chất, Tổng cục Địa chất và Khoáng sản Việt Nam.
- Phạm, V. T. (2001). Báo cáo kết quả đánh giá khoáng sản vàng và các khoáng sản khác vùng Phước Thành - Phước Sơn, tỉnh Quảng Nam.
- Reddy, R. K. T., & Bonham-Carter, G. F. (1991). A decision - tree approach to mineral potential mapping in Snow Lake area, Manitoba. *Canadian Journal of Remote Sensing Bonham-Carter*, 17(2), 191-200.
- Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *Journal of Photogrammetry Rigol-Sanchez and Remote Sensing*, 67, 93-104.
- Rodriguez-Galiano, V. F., Chica-Olmo, M., & Chica-Rivas, M. (2014). Predictive modelling of gold potential with the integration of multisource information based on random forest: a case study on the Rodalquilar area, Southern Spain. *International Journal of Geographical Information Science Chica-Rivas*, 28(7), 1336-1354.
- Trần, V. T., & Vũ, K. (2009). Địa chất và tài nguyên Việt Nam. Nhà xuất bản Khoa học tự nhiên và Công nghệ, Hà Nội.
- Võ, Q. B. (2018). Báo cáo nghiên cứu, dự báo tiềm năng khoáng sản vàng ẩn sâu ở các trường quặng vàng đối Tam Kỳ - Phước Sơn vùng Trung Trung Bộ. Trung tâm thông tin, lưu trữ và tạp trí địa chất, Tổng cục Địa chất và Khoáng sản Việt Nam.
- Zhang, C., & Ma, Y. (2012). *Ensemble machine learning: methods and applications*. Springer.