

# Machine learning model to assess gold mineralization potential mapping for northwest Thanh Hoa province

TRUONG Xuan Quang<sup>1\*</sup>, TRAN Van Anh<sup>2</sup>, TRUONG Xuan Luan<sup>3</sup>

<sup>1</sup> Faculty of Architecture, Urban Design and Sustainable Sciences, VNU School of Interdisciplinary Sciences and Arts, Vietnam National University, Hanoi, Vietnam

<sup>2</sup> Department of Photogrammetry and Remote Sensing, Hanoi University of Mining and Geology, Hanoi, Vietnam

<sup>3</sup> Faculty of Information Technology, Hanoi University of Mining and Geology, Hanoi, Vietnam

\* Corresponding email: txquang@vnu.edu.vn

**Abstract:** *In Vietnam, gold is one of the key minerals for mining. While gold mining in Thanh Hoa province holds economic value, evaluating and predicting its spatial distribution remains challenging. Machine learning is becoming a powerful tool in the field of mineral research and extraction, including gold. The strength of Machine Learning lies in its ability to process and analyze large amounts of data to make more accurate predictions, optimize exploration and extraction processes, and minimize risks. This paper presents a set of machine learning models to identify the best model for generating a gold deposit potential map. The study utilized seven thematic maps, including lithology, magnetic data, gravity data, geological age, faults, lineaments, ore point density, magma distribution, gold mineral potential from remote sensing imagery, and placer gold distribution, as input data for the models. Additionally, 706 points (353 sampling sites with gold placer and 353 sites without gold) were used to generate the training and testing dataset. The study area is situated in the northwestern part of Thanh Hoa province, known for its high potential for gold deposits. Machine learning models such as Random Forest, Logistic Regression, SVM, and Gradient Boosting were implemented to identify the best-fit model for the study area. After comparing the models, the initial results showed that the Random Forest model achieved the highest accuracy with an AUC of 0.82, identifying 4% of the area as having very high potential. The final result, a gold mineral potential map, was compared with field data, and it was found that all points containing gold in the field were located in the areas (very high potential) in the prediction map.*

**Keywords:** *Gold mineral potential mapping, Machine Learning, Northwest Thanh Hoa province.*

## 1. Introduction

Mineral potential mapping (MPM) plays a significant role in mineral exploration [1]. Geographic Information Systems (GIS) are well-known as powerful tools for processing and combining data within maps in mineral potential mapping [2]. MPM uses a classification algorithm to distinguish prospective areas from non-prospective areas [3]. Recently, machine learning has been applied in many fields of study, including mineral deposit research, such as for building mineral potential maps. Several machine learning techniques have been used for MPM, such as logistic regression [4, 5], Random Forest [6, 7, 8], Bayesian network classifiers [9], support vector machine, SVM [10, 11], convolutional neural network, CNN [12, 13], artificial neural network, ANN [2, 14], and multilayer perceptrons [15].

The difficulty of MPM based on machine learning has been discussed in previous research. The input data for MPM modeling have some limitations: geochemical sampling is usually spatially discontinuous, and the limited sampling density cannot adequately support quantitative analysis [12] the data imbalance of geochemical exploration data means that there are only a few known mineralized data points in the study area and numerous data points to be evaluated [16]. Classification of locations for the training dataset, as prone or non-prone to mineral deposits based on expert experience, can also change and become subjective when more information becomes available [2].

This study aimed to avoid the above problems. Different machine learning models (Random Forest, Logistic Regression, SVM, and Gradient Boosting) were used to compare and find the best model to reduce the overfitting problem. Moreover, the dataset was enriched by adding more samples from various sources. The total training and testing of the models used in the study was 353 sampling sites with gold placers, including 19 newly exploited points in 2022 and 2023.

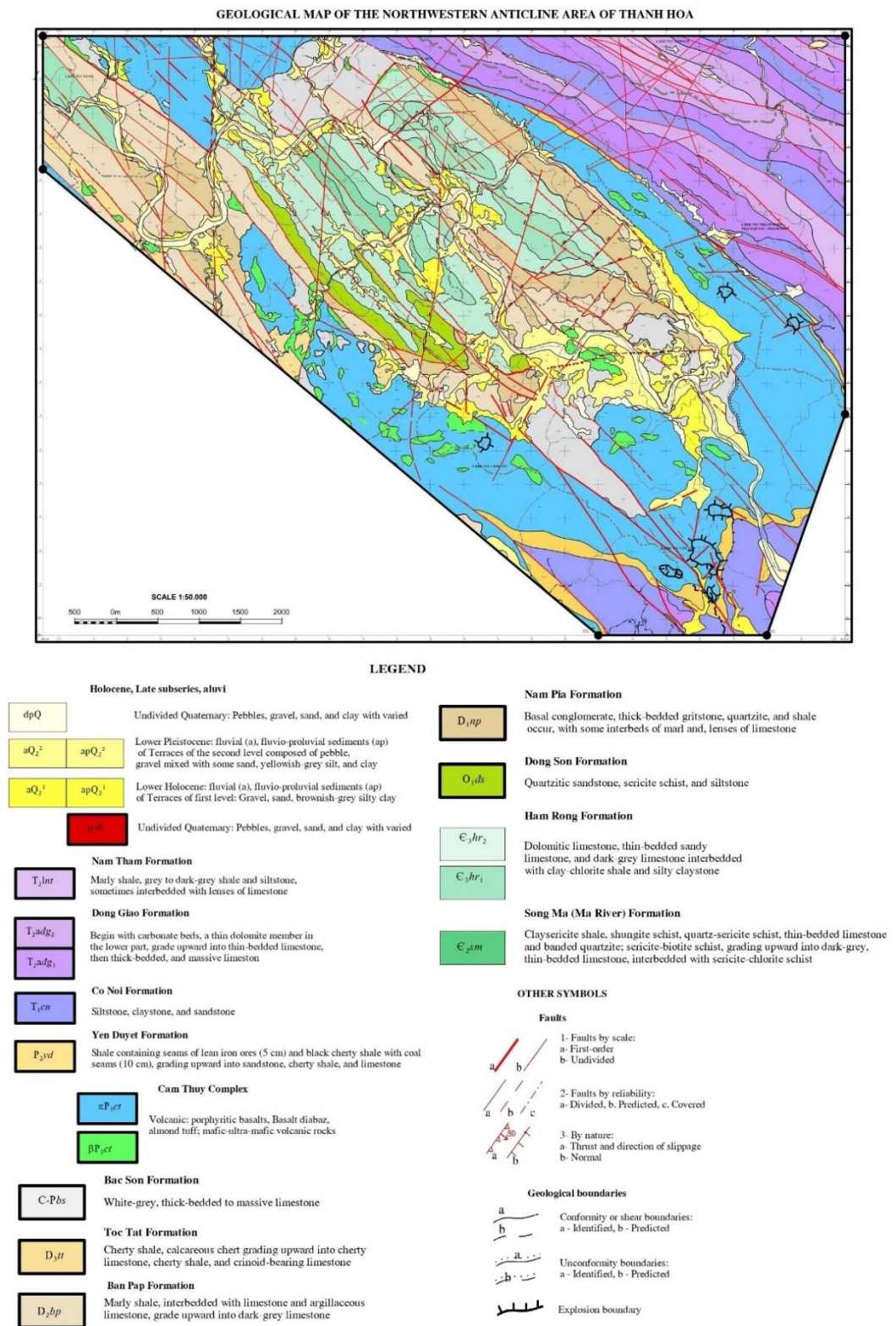
## 2. Geological background

The Northwestern region of Thanh Hoa province, Vietnam, is within the Thanh Hoa structural zone, characterized by a wedge-shaped anticline structure sandwiched between the Mesozoic Son La and Sam Nua terranes [17]. The Early Paleozoic intracontinental orogeny of the study area is composed of the Lower Paleozoic clastic-carbonate sediments of the Song Ma (€2sm), Ham Rong (€3-O1hr), and Dong Son (O1ds) Formations. The Middle Paleozoic stratigraphy is composed of the Nam Pia (D1np), Ban Pap (D2bp), and Ban Cai (D3bc) formations. The Upper Paleozoic stratigraphy is the deep marine carbonate from the Bac Son Formation (C-Pbs). The Late Permian-Early Mesozoic includes basaltic eruptions of the Cam Thuy (P3ct); Yen Duyet (P3yd), Co Noi (T1cn), Dong Giao (T2adg) and Nam Tham (T2nt) Formations; and coal-bearing sediments of the Suoi Bang Formation (T3n-rsb). In the study area, the basalt and diabase units of the Cam Thuy Formation exhibit characteristics of subvolcanic facies, occurring as dikes, ring dikes, and dome-shaped intrusions.

The formations with gold mineralization and associated minerals are Ham Rong, Nam Pia, Ban Pap, Co Noi, Dong Giao, and Nam Tham Formations. The primary forms of Au-mineralisation in the region are as below: (1) Gold-quartz-low-sulfide mineralization is distributed in the fissures and fissure zones of basalt and diabatic rocks, low-grade thermally metamorphosed sedimentary rocks; (2) Gold mineralization in propylitic alteration zones. Propylite alteration is distributed along fracture and fault systems. Areal propylite coincides with explosion centers, overlying basaltic tuff breccias; (3) Gold-antimony mineralization is distributed within carbonate sedimentary units in the form of pockets and veins, with small-scale deposits; (4) Gold and polymetallic mineralization consist of stockwork veins and veinlets within limestone and calcareous shale, and (5) Carlin-like gold mineralization is distributed in the carbonate sedimentary rocks. The orebodies of the mineralization types in the study area often coincided with the intersection of tectonic faults, with the main direction being NW-SE (Figure 1).

### 3. Study areas and data used

The northwestern region of Thanh Hoa covers an area of approximately 2500 km<sup>2</sup>. According to Đovjikov et al. [18], the area is within the Thanh Hóa structural zone, characterized by a wedge-shaped anticline structure sandwiched between the Mesozoic Son La and Sam Nua terranes. Based on the division of tectonic units of Vietnam from Tran Van Tri et al. [17, 19], most of the northwestern Thanh Hoa region lies within the Paleozoic intracontinental orogeny of the Northwest, with a small northeastern area belonging to the Permian-Early Mesozoic Song Da intracontinental rift. The southwestern boundary of the region is the Late Paleozoic-Early Mesozoic Truong Son orogenic belt and the high-grade metamorphic Phu Hoat terrane. The Paleozoic intracontinental orogeny is composed of Lower Paleozoic clastic-carbonate sediments of the Song Ma (€2sm), Ham Rong (€3-O1hr), Dong Son (O1ds), Nam Pia (D1np), Ban Pap (D2bp), and Ban Cai (D3bc) formations. The Upper Paleozoic strata are the deep marine carbonate from the Bac Son Formation (C-Pbs). The Late Permian-Early Mesozoic Song Da intracontinental rift system includes basaltic eruptions of the Cam Thuy formation (P3ct); coal-bearing clastic sediments of the Yen Duyet formation (P3yd); Early Mesozoic clastic-carbonate sediments of the Co Noi (T1cn), Dong Giao (T2adg) and Nam Tham (T2nt) formations. The formations with gold mineralization and associated minerals are Ham Rong, Nam Pia, Ban Pap, Co Noi, Dong Giao, and Nam Tham Formations.



**Fig. 1.** Geological map of the study area (Source: Department of Geology and Minerals of Vietnam)

In the study area, the basalt and diabase units of the Cam Thuy Formation exhibit characteristics of sub-volcanic facies, occurring as dikes, ring dikes, and dome-shaped intrusions [19]. The orebodies of the mineralization types often coincide with the intersection of tectonic faults, with the main direction being NW-SE. The main forms of Au-mineralisation are as follows: (1) The gold-quartz-low-sulfides mineralization is distributed in the fissures and fissure zones of basalt and diabolic rocks, low-grade thermally metamorphosed sedimentary rocks; (2) Gold mineralization in propylitic alteration zones. Propylite is distributed along fracture and fault systems. Propylite areas coincide with volcanic explosion centers, overlying basaltic tuff breccias; (3) Gold – Antimony mineralization ore is distributed within

carbonate sedimentary units in the form of pockets and veins, with small-scale deposits; (4) Gold and polymetallic mineralization consist of stockwork veins and veinlets within limestone and calcareous shale; and (5) Carlin-like gold mineralization is distributed in the carbonate sedimentary rocks [20, 21]. The generation of predictor MPM maps was implemented using 10 factors (Table 1) and (Figure 2).

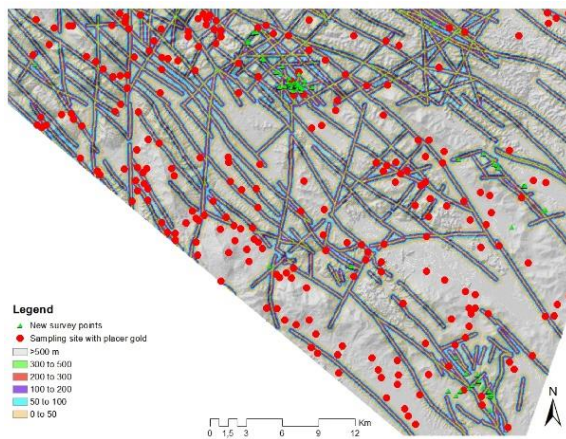
**Tab. 1.** Data used

Map layers	Classes	Description
1. Fault buffer (m)	(1) 0-50; (2) 50-100; (3) 100-200; (4) 200-300; (5) 300-500; (6) >500 (m)	Generate from geological and petrographic-structural maps (1:50.000)
2. Lineament buffer (m)	(1) 0-50; (2) 50-100; (3) 100-200; (4) 200-300; (5) 300-500; (6) >500 (m)	Generate from geological, petrographic-structural maps (1:50 000) and Remote sensing interpreter in both optical (10 m) and SAR images (10 m).
3. Magma	(1) Diabase, gabbro-d diabase, and sedimentary rocks; (2) Aphanitic diabase; (3) Massive basalt with blocky structure; (4) Amygdaloidal basalt; (5) Vesicular basalt; (6) Basaltic tuff; (7) Basaltic tuff breccia	Generate from geological and petrographic-structural maps (1:50,000)
4. Lithology	(1) Conglomerates, quartzites, and quartz-mica schists; (2) Sandstones and siltstones (3) Massive limestones; (4) Mudstones and clayey siltstone; (5) Calcareous shales with organic material; (6) Marly clays and thin-bedded limestone.	Generate from geological and petrographic-structural map (1:50 000)
5. Placer aureole	(1) 1-2 ; (2) 3-5; (3) 6-10; (4) 11-20; (5) >20 gold particles/10dm <sup>3</sup>	Generate from geological and petrographic-structural map (1:50 000)
6. Bouguer gravity anomaly map (mGal)	(1) [(-63.6) – (-40.7)]; (2) [(-40.7) – (-32.2)]; (3) [(-32.2) – (-27.4)]; (4) [(-27.4) – (-21.7)]; (5) [(-21.7) – (-16.3)]; (6) [(-16.3) – (-10.6)]; (7) [(-10.6) – (-3.3)]; (8) [(-3.3) – (5.5)] and (9) [5.5 – 17.3]	Classification method: natural breaks (1:50.000)
7. Magnetic anomaly map (nT)	(1) [(-304.9) – (-188.6)]; (2) [(-188.6) – (-111.8)]; (3) [(-111.8) – (-66.5)]; (4) [(-66.5) – (-29)]; (5) [(-29) – (-3.4)]; (6) [(-3.4) – 14.3]; (7) [14.3 – 34.04]; (8) [34.04 – 61.6] and (9) [61.6 – 197.6]	Classification method: natural breaks (1:50.000)
8. Density of the sampling site with placer gold	The total points of the sampling sites with gold placer were 353, including 15 new gold mineral points discovered in 2024.	The map was established using ArcGIS software.
9. Geological age	(1) Dong Son (O <sub>1ds</sub> ); Toc Tat (D <sub>3tt</sub> ) formations and Cam Thuy complex (2) Yen Duyet (P <sub>3yd</sub> ), Nam Pia (D <sub>1np</sub> ) formations (3) Co Noi (T <sub>1cn</sub> ), Song Ma (€2sm) formations (4) Ban Pap formation (D <sub>1-2bp</sub> ), (5) Bac Son (C-Pbs),	Generated from geological map (1:50.000)

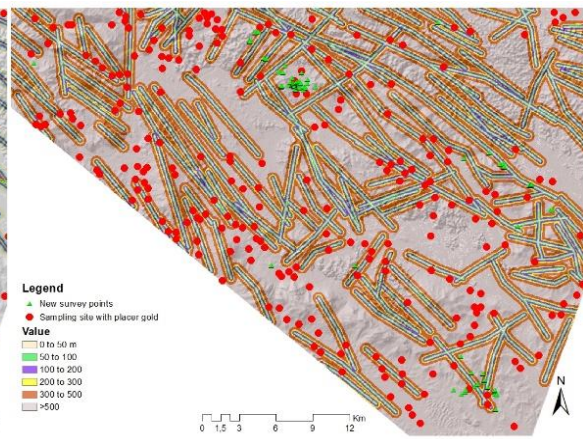
Map layers	Classes	Description
	Dong Giao (T <sub>2</sub> adg) and Nam Tham (T <sub>2</sub> lnt) formations; (6) Ham Rong formation (E <sub>3</sub> hr)	
10. Mineral potential zones	The layer generated by using the Rock Index (RI) includes (1) Very low, (2) Low, (3) Medium, (4) High, and (5) Very high potential.	Implemented based on the Sentinel-2 image (RS); classification method: natural breaks

In this paper, the dataset was divided into a 70% training set and a 30% testing set. It consists of 353 sampling sites with gold placers or gold minerals and 353 sampling sites without gold placers, totaling 706 points, including gold deposit and non-deposit samples.

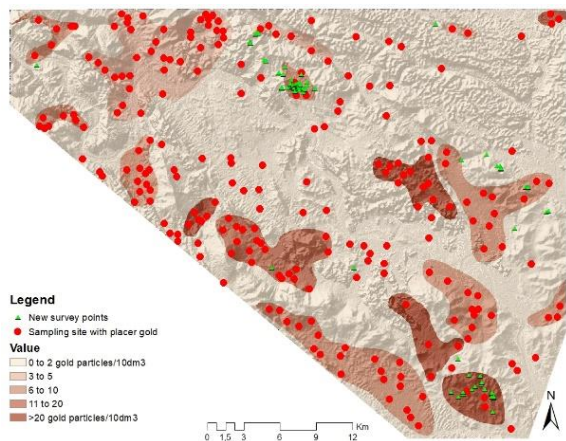
To have sufficient training data, we used 334 sampling sites with gold placer locations from the geological map and updated 19 gold-bearing mineral locations collected in 2022 and 2023. The total number of deposit samples is 353. The general criteria for identifying the 353 non-deposit samples are as follows: The non-deposit samples were collected from uncrushed rocks, unfolded rocks, do not have small geological fractures in various directions, eliminating the potential for hosting hydrothermal veins containing ore minerals, unmetamorphosed rocks, and rocks located distal from craters. Rock groups little likely to contain primary gold in the study area have been identified, including (1) Dong Son, Bac Son, Toc Tat, Nam Tham stratigraphic units and the Holocene or Pleistocene formations; (2) Hard rocks, not subject to brittle deformation and dynamic metamorphism, with low joint density (quartzitic limestone, quartzitic sandstone, massive limestone, white-grey, cherty shale); (3) Rocks not favorable for thermal contact metamorphism, medium-low temperature metasomatism such as quartzite, dolomite, massive limestone, dense basalt; (4) Areas located far from eruption centers, far from volcanic craters, far from major northwest-southeast striking tectonic faults. For training and testing the Machine Learning models, ML, the paper used 706 points, including non-deposit and deposit samples (Figure 3).



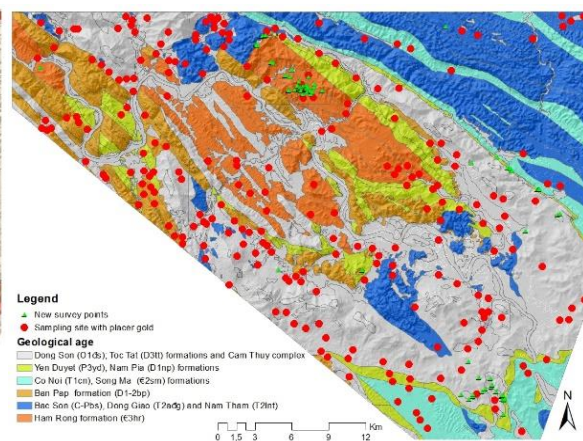
a. Fault buffer map (The base map is a Hillshade map from ALOS PALSAR DEM)



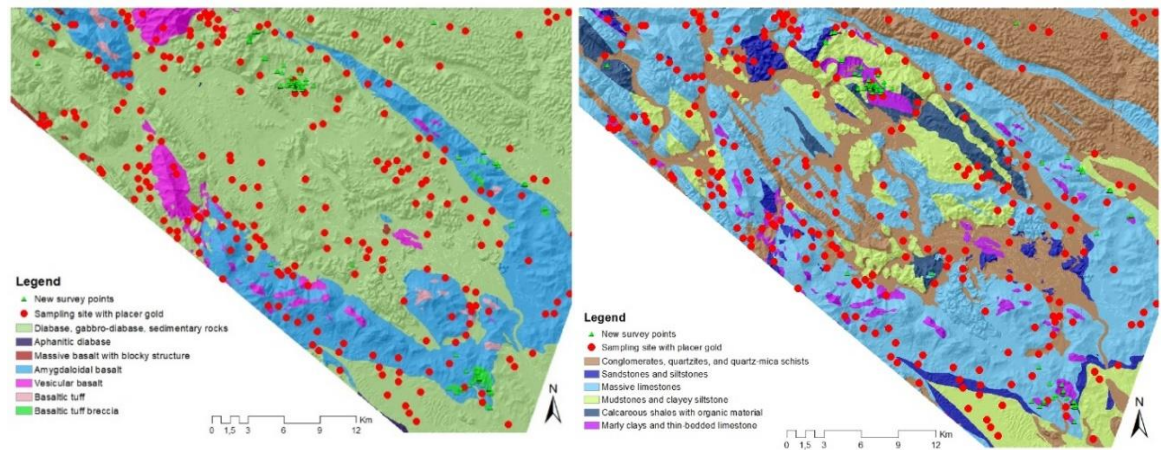
b. Lineament buffer (The base map is a Hillshade map from ALOS PALSAR DEM)



c. Density of the sampling site with placer gold

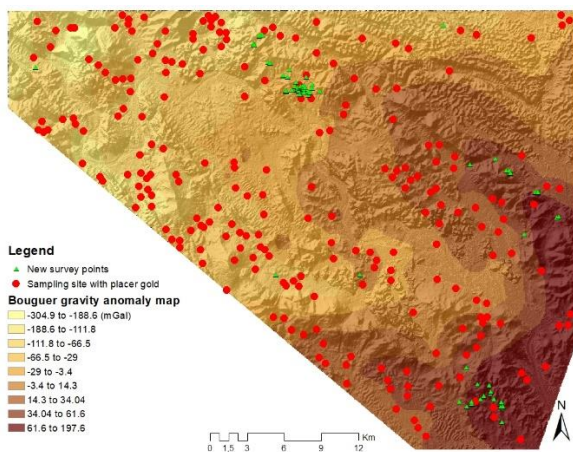


d. Geological age

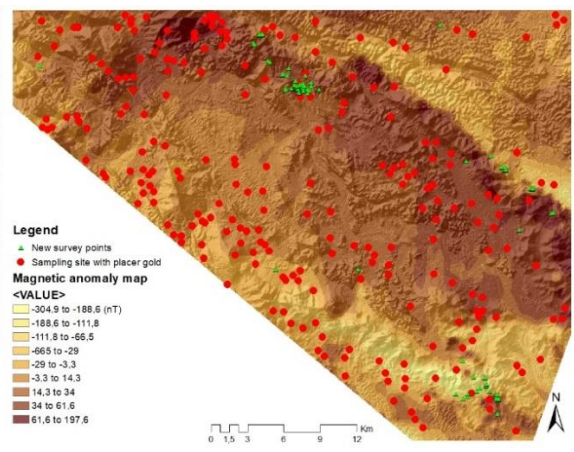


e. Magma

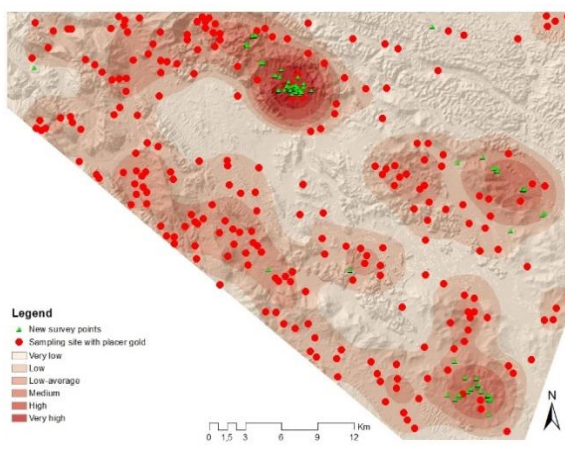
f. Lithology



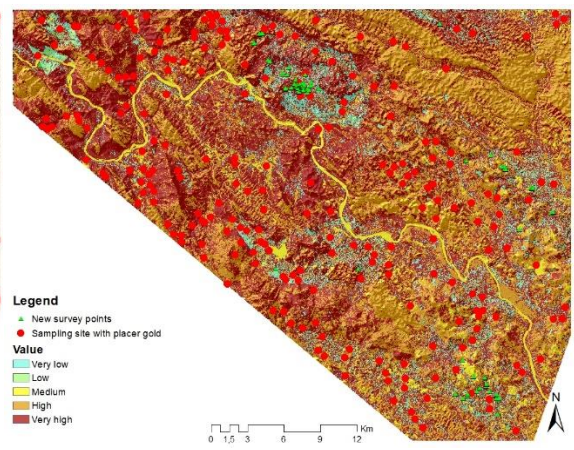
g. Bouguer gravity anomaly map



h. Magnetic anomaly map

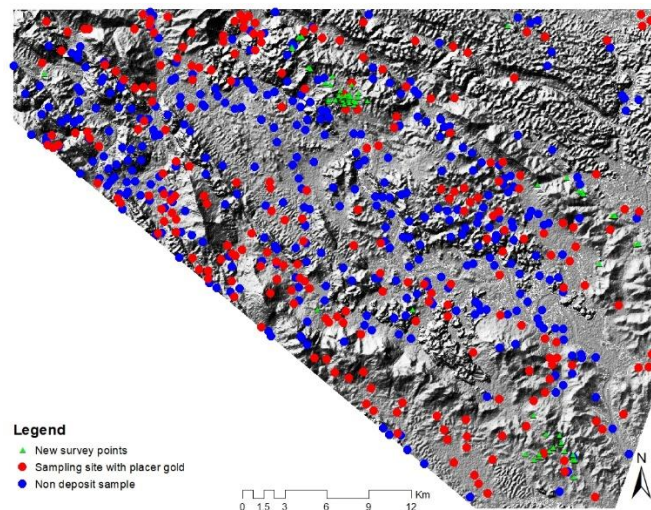


i. Density of deposit samples



k. Mineral potential zones from remote sensing data

**Fig. 2.** Ten map factors of the MPM using machine learning models



**Fig. 3.** Sampling sites with gold placer distributed in the study area

### 3. Methodology

The research aims to prepare and evaluate the gold mineral deposit map (MPM) of the study area located in Northwest Thanh Hoa Province using four machine learning algorithms: Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), and Gradient Boosting. Ten conditioning factors were determined, including geology (fault buffer, magma, lithology, geological age, deposit placer aureole sample density); geophysics (gravity and magnetic maps), and both mineral potential zones and lineaments were extracted from satellite images (Figure 2). The implementation diagram of the research is shown in (Figure 4).

#### a. Logistic Regression (LR)

Logistic regression (LR) is a statistically based model that uses a probabilistic function to establish a relationship between the dependent variable (deposit = 1, non-deposit = 0) and the independent variables (conditioning factors) [2]. The existence or absence of gold minerals should serve as the dependent variable in the logistic regression (LR). Formula (1) expresses the relationship between the independent and dependent variables:

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}} \quad (1)$$

where  $p$  is the probability of the presence or absence of a deposit, which is between 0 and 1, Where  $\beta_0$  is the constant value of the equation and  $\beta_1, \dots, \beta_n$  are the coefficients of each conditioning factors/variables  $x_1, \dots, x_n$  [4, 22].

#### b. Random Forest

Breiman (2001) first introduced the Random Forest (RF) classification algorithm, a supervised learning technique that combines several decision trees for enhanced performance [23]. The fundamental idea of Random Forest is to create a variety of uncorrelated decision trees and apply the bootstrap aggregation technique to create training subsets. After the predictions of each decision tree, the final result is based on either the average (for regression) or the majority vote (for classification). The training data is divided into random subsets with replacements to construct the forest (bagging). Only a random subset of the features in each tree is considered for splitting at each node. More robust trees are produced, and overfitting is less likely due to this randomness. Compared to single decision trees, Random Forest is less prone to overfitting and achieves high accuracy by reducing variation.

#### c. Support Vector Machine

Cortes and Vapnik (1995) developed the statistical theory algorithm known as Support Vector Machines (SVM) to find the optimum hyperplane for separating data points into distinct classes [24]. Finding the hyperplane with the largest margin between the nearest data points in each class is the main concept. Future data points can be better-classified thanks to the buffering effect of this margin. Support vectors are the data points from each class that are closest to the hyperplane. These points impact the hyperplane's position and define the margin. SVMs can handle non-linear data by utilizing kernel functions, even though they are best suited for linearly separable data. SVM is useful with high-dimensional data, works well even with little training data, and is less prone to overfitting than other methods. These functions map the data into a higher-dimensional space where a linear separation becomes possible.

d. Gradient Boosting

Gradient boosting is a robust machine learning technique applied in regression and classification tasks. It is an algorithm that combines multiple weak learners into a stronger model. This process is performed by building each sub-model based on the errors of the previous model and minimizing the loss function through gradient descent. Gradient Boosting can be applied to various types of problems. Gradient Boosting Classifier in scikit-learn was used for implementing the gradient boosting model in the study.

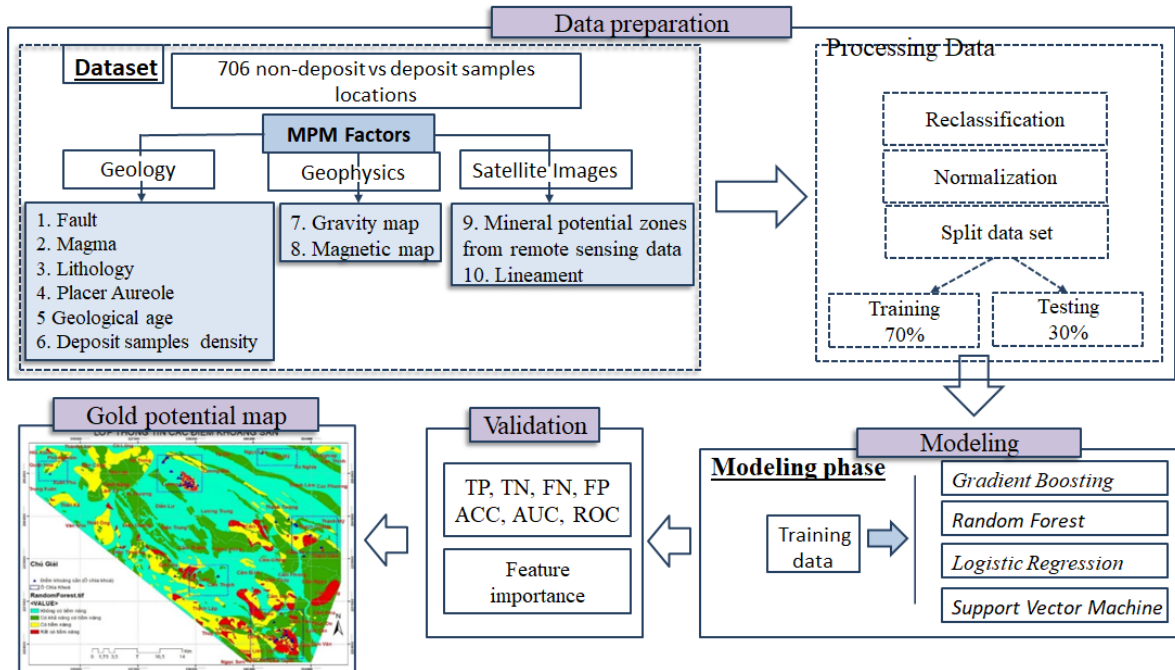


Fig. 4. Implementation diagram of the research.

In this study, the gold mineral potential map was evaluated using the receiver operating characteristic (ROC) curve and other statistical metrics, including accuracy and precision [16]. Commonly used performance indicators were calculated based on the confusion matrix provided by the classifier. The confusion matrix entries include true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN), as shown below.

			Predict		
			Positive	Negative	
Accuracy the percentage correctly given dataset.	Actual	Positive	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>	(2) measures of samples predicted in a The model's
		Negative	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>	

precision (3) measures the percentage of true positive samples among all predicted positive samples. Recall (4) assesses the proportion of true positive samples among all actual positive samples. The F1 score (5) represents the harmonic mean of precision and recall, ranging from 0 to 1.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

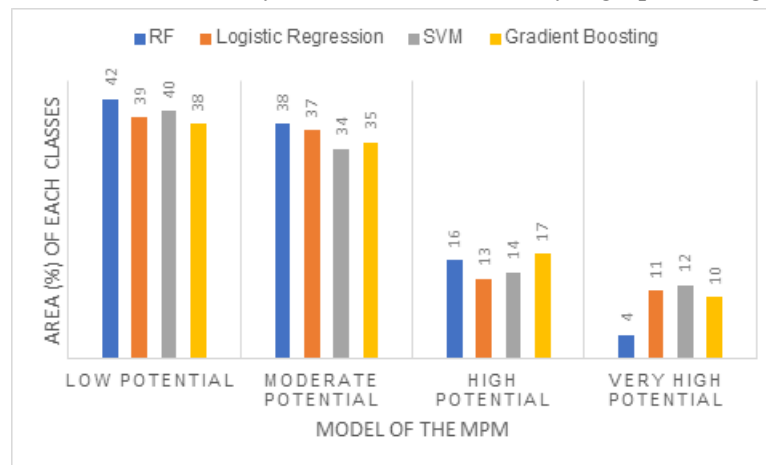
$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

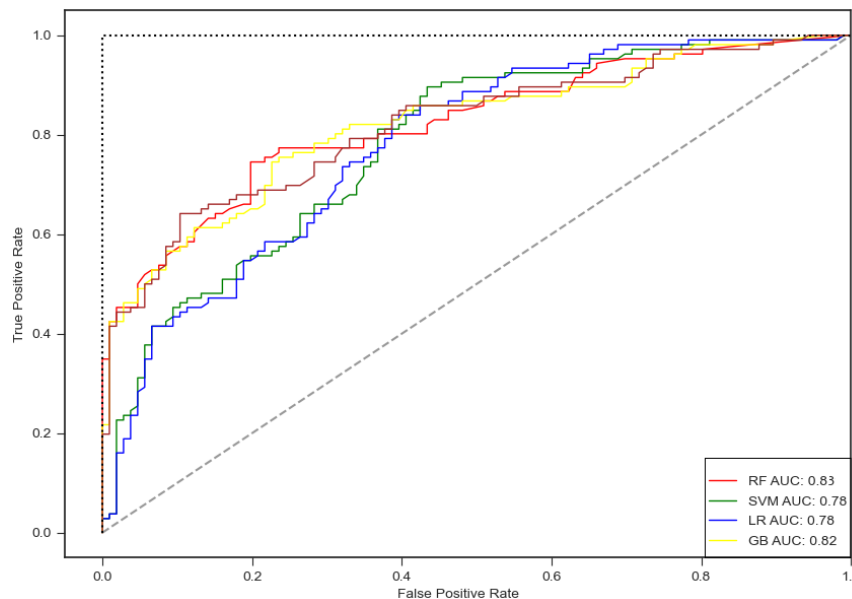
$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (5)$$

4. Results and discussion

Finally, the gold mineral potential map, MPM was implemented using four machine learning methods: Logistic Regression, Random Forest, Support Vector Machine, and Gradient Boosting methods. The MPM indices produced by four methods were established into zones ranging from: very high potential, high potential, moderate, and low potential. The highest accuracy in the RF model was achieved using an open-source Python library, Scikit-Learn [25]. The map was reclassified into four different levels, ranging from low potential, moderate, potential, and very high potential. Justifying the hyperparameters selected for each machine learning model involved using K-fold cross-validation and the GridSearchCV function from sklearn. To identify thresholds of the very high potential zones, the researchers used multiple thresholds, evaluated them and picked the one with the highest F1 score. The Random Forest model indicates that a very high potential of 4% (Figure 5) is achieved with 130 trees; a maximum of iterations: 100; a minimum of samples per leaf node: 10; and with the area under the curve (AUC) of ROC (AUC=0.83) The performance of the Logistic Regression method based on scikit-learn library is controlled by various parameters like: max\_iter: 100; tol: 0.0001 and using Limited memory Broyden Fletcher Goldfarb Shanno (L-BFGS) for optimization, it reduces loss function fluctuations and enhances the convergence rate of the training process [26], AUC=0.78 and percentages of a very high potential zone is about 11%. The Support Vector Machine (SVM) model utilized a sigmoid kernel with a stopping criterion tolerance of 0.0001 and achieved an AUC = 0.79, indicating a 12% probability for very high potential zones. Meanwhile, the Gradient Boosting model, applying hyperparameters such as n\_estimators=100, tol=0.0001, and max\_depth=5, revealed that 10% of the study area is classified as very high-potential gold (Figure 6)



**Fig. 5.** The percentage of MPM area of the four models



**Fig. 6.** AUC values of the gold mineral potential mapping models

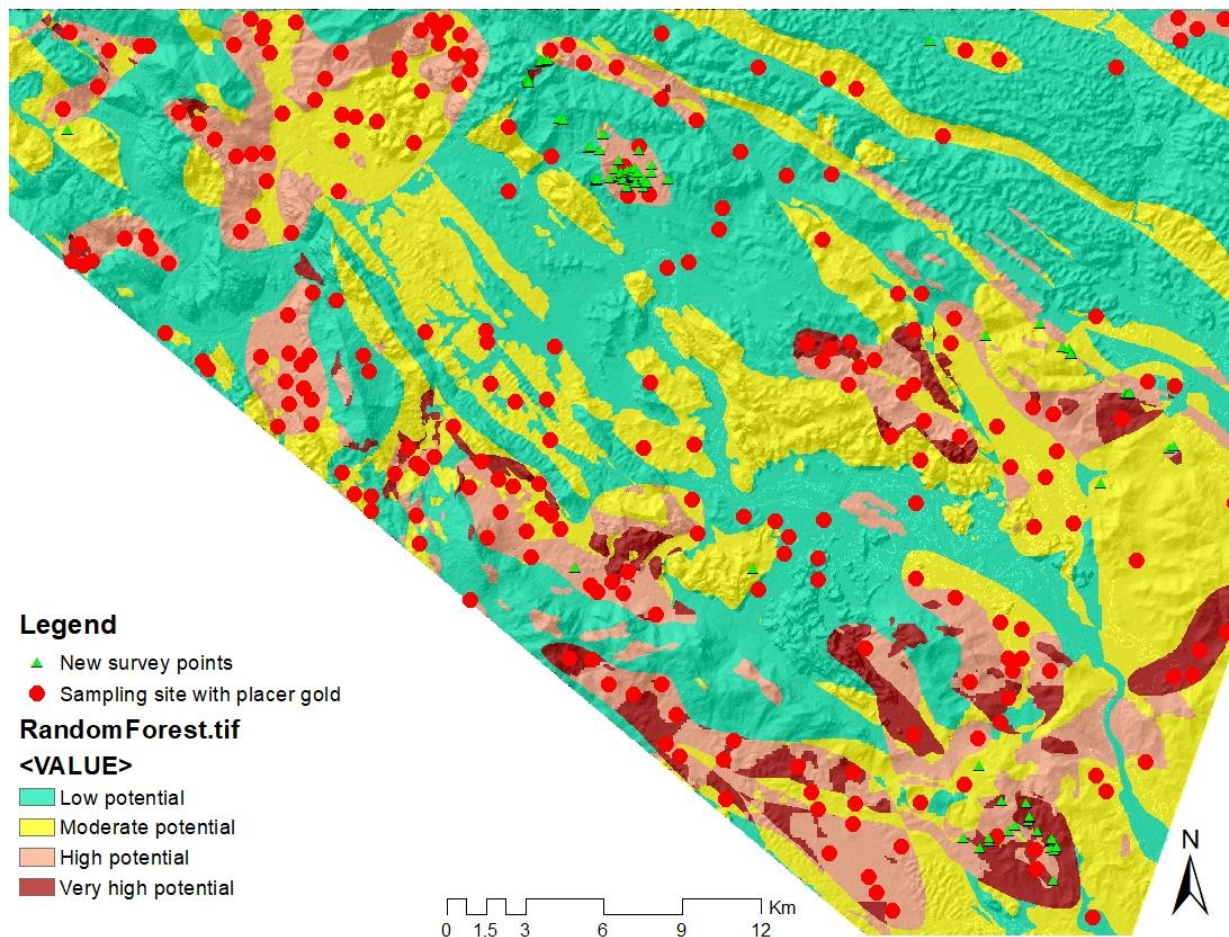
**Tab. 3.** Performance of the machine learning models for MPM

<b>Model</b>	<b>precision</b>	<b>recall</b>	<b>f1-score</b>	<b>Accuracy</b>
Random Forest	0.8	0.82	0.75	0.83
Logistic Regression	0.69	0.78	0.73	0.78
SVM	0.7	0.78	0.74	0.78
Gradient Boosting	0.74	0.82	0.75	0.82

Table 3 shows the model's performance scores, considering gold mineral potential mapping for Sampling sites with gold placer. Random Forest showed the highest accuracy at 0.83, followed by Gradient Boosting (0.82), Support vector machine (0.78), and Logistic Regression (0.78). Using four machine learning/AI models, the random forest was selected to create a gold mineral potential map for the study area on a 1:50,000 scale, the map identifying three levels of potential: Very high potential, high potential, moderate, and low potential. Additionally, areas with no potential for gold mineralization in the study area have been delineated. The reliability level, based on accuracy, reaches 83%. Most of the prospective areas for mineralization align with faults, ancient volcanic vents, and fault intersections and generally extend along the northwest-southeast direction. These prospective areas, which also coincide or nearly coincide with known mineral points, are acceptable. The research results provide accurate guidance for future gold mineralization surveys, particularly in the highly prospective and prospective areas identified on the prediction map but not yet noted by previous geologists (Figure 7).

Areas with high gold mineralization potential are located at the edges of sub-volcanic igneous massifs, where eruptive activities and volcanic vents, as well as thermal contact metamorphic zones, are present. These areas are also found near the northwest-southeast striking faults and deformation systems, particularly at fault intersections. The orientation of the potential areas aligns with the northwest-southeast trending structure of the region, coinciding with the locations of registered deposits and ore occurrences as shown on the geological map. The results of random magnetotelluric measurements at certain locations indicate that the hidden gold prediction results are reliable. These prediction results provide valuable information for planning future gold mineral activities.

The RF classifier predicts that very low potential often coincides with the locations mentioned above, including places such as Cam Thac, Cam Van, and Cam Yen in Cam Thuy district; Thach Dinh, Thach My, and Thach Binh in Thach Thanh district. The gold placer plays a crucial role in the search for primary gold deposits and contributes to the development of mineral potential mapping. However, attention should be given to factors such as shape, size, and especially the abrasion of minerals. These factors provide valuable information about the original ore deposits.



**Fig. 7.** Gold mineral potential map for the study area based on Random Forest

### Acknowledgments

This work is funded by the Ministry of Science and Technology of Vietnam (MOST), with project code number ĐTĐL.CN-85/21. The authors thank the anonymous reviewers for improving the paper.

### Literature - References

1. Yixiao Wu, Bingli Liu, Yaxin Gao, Cheng Li, Rui Tang, Yunhui Kong, Miao Xie, Kangning Li, Shiyao Dan, Ke Qi, Yufei Ren, Zhuo Wu, Mineral prospecting mapping with conditional generative adversarial network augmented data, *Ore Geology Reviews*, Volume 163, 2023, 105787, ISSN 0169-1368, <https://doi.org/10.1016/j.oregeorev.2023.105787>.
2. Lee, Saro & Oh, Hyun-Joo. (2011). Application of Artificial Neural Network for Mineral Potential Mapping. 10.5772/16187.
3. Xiong, Y., & Zuo, R. (2018). GIS-based rare events logistic regression for mineral prospectivity mapping. *Computers & Geosciences*, 111, 18-25. doi:10.1016/j.cageo.2017.10.005
4. Zhang, D., Ren, N., & Hou, X. (2018). An improved logistic regression model based on a spatially weighted technique (ILRBSWT v1.0) and its application to mineral prospectivity mapping. *Geoscientific Model Development*, 11(6), 2525-2539. doi:10.5194/gmd-11-2525-201
5. Zhao PD (2007) Quantitative mineral prediction and deep mineral exploration. *Earth Sci Front* 14(5):1–10. <https://doi.org/10.3321/j.issn:1005-2321.2007.05.001>. (in Chinese with English abstract)
6. Emmanuel John M. Carranza, Alice G. Laborte, Random forest predictive modeling of mineral prospectivity with small number of prospects and data with missing values in Abra (Philippines), *Computers & Geosciences*, Volume 74, 2015, Pages 60-70, ISSN 0098-3004, <https://doi.org/10.1016/j.cageo.2014.10.004>.
7. Emmanuel John M. Carranza, Alice G. Laborte, Data-driven predictive mapping of gold prospectivity, Baguio district, Philippines: Application of Random Forests algorithm, *Ore Geology Reviews*, Volume 71, 2015, Pages 777-787, ISSN 0169-1368, <https://doi.org/10.1016/j.oregeorev.2014.08.010>
8. Chen, M., Xiao, F. Projection Pursuit Random Forest for Mineral Prospectivity Mapping. *Math Geosci* 55, 963–987 (2023). <https://doi.org/10.1007/s11004-023-10070-0>

9. Porwal, A., Carranza, E. J. M., & Hale, M. (2006). Bayesian network classifiers for mineral potential mapping. *Computers & Geosciences*, 32, 1-16. <https://doi.org/10.1016/j.cageo.2005.03.018>
10. Renguang Zuo, Emmanuel John M. Carranza, Support vector machine: A tool for mapping mineral prospectivity, *Computers & Geosciences*, Volume 37, Issue 12, 2011, Pages 1967-1975, ISSN 0098-3004, <https://doi.org/10.1016/j.cageo.2010.09.014>
11. Neda Mahvash Mohammadi, Ardeshir Hezarkhani, Application of support vector machine for the separation of mineralised zones in the Takht-e-Gonbad porphyry deposit, SE Iran, *Journal of African Earth Sciences*, Volume 143, 2018, Pages 301-308, ISSN 1464-343X, <https://doi.org/10.1016/j.jafrearsci.2018.02.005>.
12. Wang, Z., Zuo, R. & Yang, F. Geological Mapping Using Direct Sampling and a Convolutional Neural Network Based on Geochemical Survey Data. *Math Geosci* 55, 1035–1058 (2023). <https://doi.org/10.1007/s11004-022-10023-z>
13. Renguang Zuo, Emmanuel John M. Carranza, Support vector machine: A tool for mapping mineral prospectivity, *Computers & Geosciences*, Volume 37, Issue 12, 2011, Pages 1967-1975, ISSN 0098-3004, <https://doi.org/10.1016/j.cageo.2010.09.014>
14. Alok Porwal, E.J.M. Carranza, M. Hale, Bayesian network classifiers for mineral potential mapping, *Computers & Geosciences*, Volume 32, Issue 1, 2006, Pages 1-16, ISSN 0098-3004, <https://doi.org/10.1016/j.cageo.2005.03.018>
15. Skabar, A.A. (2005). Mapping mineralization probabilities using multilayer perceptrons. *Natural Resources Research*, Vol. 14, No. 2, 109-123, ISSN 15207439
16. Chen, Y., Lu, L. The Anomaly Detector, Semi-supervised Classifier, and Supervised Classifier Based on K-Nearest Neighbors in Geochemical Anomaly Detection: A Comparative Study. *Math Geosci* 55, 1011–1033 (2023). <https://doi.org/10.1007/s11004-022-10042-w>
17. Trần Văn Trị, et al., 2009. *Geology and Resources Vietnam*. NXB Khoa học Tự nhiên và Công nghệ, Hà Nội. Science and Technology Publishing House (book in Vietnamese)
18. Dovjikov A. E (editor), (1965). *Geology of Northern Vietnam* (in Vietnamese). Publish House: Science and Technology, Hanoi
19. Tran V Tri et al., (2023). *Geology and Georesources of Vietnam*. Youth Public house (book in English). Pp 26 -72; 89-93
20. Zhang, N., Zhou, K., Li, D., 2018. Back-propagation neural network and support vector machines for gold mineral prospectivity mapping in the Hatu region, Xinjiang, China. *Earth. Sci. Inform.*, 11, 553-566
21. Tao Sun, Fei Chen, Lianxiang Zhong, Weiming Liu, Yun Wang, GIS-based mineral prospectivity mapping using machine learning methods: A case study from Tongling ore district, eastern China, *Ore Geology Reviews*, Volume 109, 2019, Pages 26-49, ISSN 0169-1368, <https://doi.org/10.1016/j.oregeorev.2019.04.003>
22. Phong, T. V., Phan, T. T., Prakash, I., Singh, S. K., Shirzadi, A., Chapi, K., ... Pham, B. T. (2019). Landslide susceptibility modeling using different artificial intelligence methods: a case study at Muong Lay district, Vietnam. *Geocarto International*, 36(15), 1685–1708. <https://doi.org/10.1080/10106049.2019.1665715>
23. Breiman, L. Random Forests. *Machine Learning* 45, 5–32 (2001). <https://doi.org/10.1023/A:1010933404324>
24. Cortes, C., Vapnik, V. Support-vector networks. *Mach Learn* 20, 273–297 (1995). <https://doi.org/10.1007/BF00994018>
25. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. *Scikit-learn: Machine Learning in Python*. *J. Mach. Learn. Res.* 2011, 12, 2825–2830
26. Najafabadi, M.M., Khoshgoftaar, T.M., Villanustre, F. et al. Large-scale distributed L-BFGS . *J Big Data* 4, 22 (2017). <https://doi.org/10.1186/s40537-017-0084-5>