Application of explainable machine learning to characterizing apatite fertility in porphyry-skarn deposits

Le Wang, Le; Qin, Ben; Chiaradia, Massimo; Qin, Ke-Zhang; Cao, Ming-Jian

How to cite

LE WANG, Le et al. Application of explainable machine learning to characterizing apatite fertility in porphyry-skarn deposits. In: Ore geology reviews, 2025, vol. 186, p. 106926. doi: 10.1016/j.oregeorev.2025.106926

This publication URL: https://archive-ouverte.unige.ch/unige:188326
Publication DOI: 10.1016/j.oregeorev.2025.106926

Ore Geology Reviews 186 (2025) 106926

Contents lists available at ScienceDirect

Ore Geology Reviews

journal homepage: www.elsevier.com/locate/oregeorev

Application of explainable machine learning to characterizing apatite fertility in porphyry-skarn deposits

Le Wang [a,b,c,*], Ben Qin [a,d,*], Massimo Chiaradia [c], Ke-Zhang Qin [a,b], Ming-Jian Cao [a,b]

[a] State Key Laboratory of Lithospheric and Environmental Coevolution, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100029, China
[b] College of Earth and Planetary Sciences, University of Chinese Academy of Sciences, Beijing 100049, China
[c] Department of Earth Sciences, University of Geneva, Rue des Maraîchers 13, Geneva 1205, Switzerland
[d] School of Resources and Environmental Engineering, Inner Mongolia University of Technology, Hohhot 010051, China

ARTICLE INFO

Keywords:
Porphyry-skarn deposits
Igneous apatite
Mineralization potential evaluation
Machine learning

ABSTRACT

Over the past century, numerous near-surface porphyry-skarn deposits have been discovered, mined, and depleted, driving the need for innovative exploration methods to support a green-energy-driven society. Whole-rock analyses blend signals from diverse minerals, while zircon, forming late in magma evolution, lacks information on magmatic volatiles critical for mineralization. This study introduces a novel method using global igneous apatite compositional data, analyzed with supervised machine learning algorithms—Extreme Gradient Boosting (XGBoost) and Random Forest (RF). Optimized via grid search cross-validation, these models achieved classification accuracies of 94 % (XGBoost) and 89 % (RF), with feature importance analysis identifying Mn, Sr, Cl, La, REE, F/Cl, Ce/Ce*, and Eu/Eu* as key indicators distinguishing fertile from barren apatites. To validate practical utility, the models were tested on apatites from barren host granitic rocks and fertile dacite porphyry at the Güzelyayla porphyry Cu-Mo deposit in Turkey. XGBoost and RF attained accuracies of 92 % and 98 %, demonstrating their ability to serve as reliable tools for early-stage exploration. By integrating explainable machine learning with host rock geochemistry, the study links apatite fertility indicators to critical formation factors of porphyry-skarn deposits, offering a robust framework for identifying new resources in greenfield areas. Our findings demonstrate that fertile magmas possess higher Cl concentrations than barren magmas under equivalent extents of sulfide saturation and oxygen fugacity. Additionally, high oxygen fugacity and extensive fractionation favors the formation of porphyry-skarn Cu/Mo/Au mineralization.

1. Introduction

Global metal demand doubles every 20–30 years, with copper mining over the next 26 years expected to surpass all historical production, yet discovering new Cu deposits is increasingly difficult and costly due to the depletion of large, high-grade, and near-surface deposits (Schodde, 2018, 2019). New exploration models and methods are essential to help geologists identify deeper deposits and explore new regions (Holliday and Cooke, 2007; Cooke et al., 2020; Grunsky and de Caritat 2020). Porphyry and skarn deposits, providing most of world's copper and molybdenum, usually show close relations with *syn*-ore porphyry stocks / dikes (Meinert et al., 2005; Sillitoe, 2010; Shu et al., 2019). These deposits form from fluids exsolved from hydrous and oxidized magmas (Richards, 2011; Chiaradia et al., 2009; Chiaradia, 2021; Nathwani et al., 2024). The compositions of whole rock (Baldwin and Pearce, 1982; Loucks, 2014; Nathwani et al., 2022) and zircon, a key member of porphyry indicator minerals (PIMs; Lu et al., 2016; Shu et al., 2019; Cooke et al., 2020; Zou et al., 2022), are widely utilized to assess the metallogenic potential of magmatic rocks in porphyry fertility studies. However, whole rocks reflect mixed information of many minerals crystallized at different levels (Chiaradia et al., 2025) and their geochemistry is vulnerable to post-ore hydrothermal alteration, metamorphism, and weathering. Zircon generally crystallizes in more felsic rocks (Hoskin et al., 2000) which may bias the inferred host rock toward more evolved compositions (Jennings et al., 2011) and cannot record the information of volatiles, which are critical for porphyry mineralization (Grondahl and Zajacz, 2022).

Apatite ($Ca_5[PO_4]_3[F, Cl, OH]$), however, is generally an early crystallizing phase over the whole duration of magma cooling (Hoskin et al., 2000). It can structurally incorporate a broad range of elements

* Corresponding author at: No. 19 Beitucheng Western Road, Chaoyang District 100029, Beijing, PR China.
E-mail addresses: wangle@mail.iggcas.ac.cn (L. Wang), benqin@imut.edu.cn (B. Qin).

https://doi.org/10.1016/j.oregeorev.2025.106926
Received 20 July 2025; Received in revised form 1 October 2025; Accepted 2 October 2025
Available online 3 October 2025

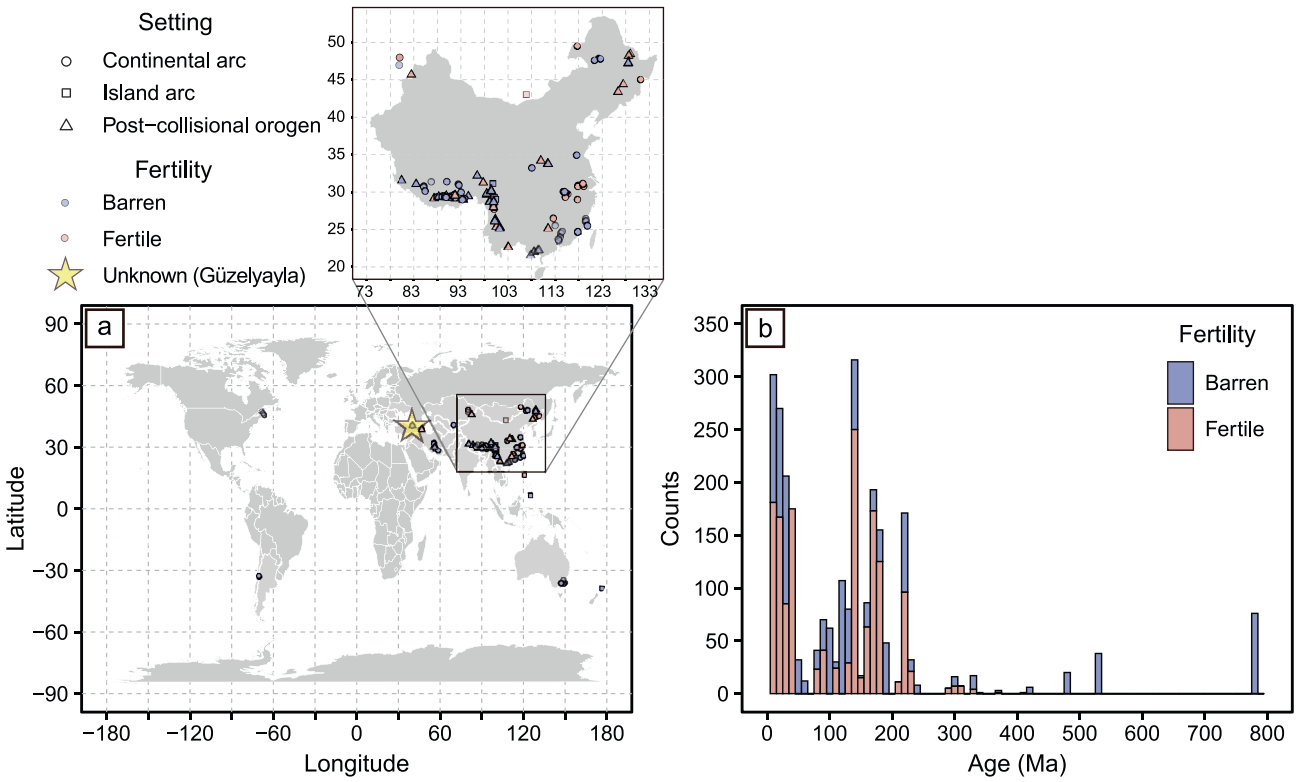


Fig. 1. Spatiotemporal distribution of apatite samples. (a) Spatial distribution of global data compilation of apatite compositions from the literature (Table S2). Fertile dacite porphyry and barren host granitic rocks from the Güzelyayla porphyry Cu-Mo prospect (yellow star) were taken as unknown sample to test the constructed machine learning models. The map was created with R version 4.4.0. (b) Histogram of age-frequency distribution of apatite samples (age bin = 10 Myr).

through the following substitutions (Elliot, 1994; Pan and Fleet, 2002; Hughes and Rakovan, 2015): (Sr^{2+}, Mn^{2+}, Fe^{2+}, rare earth elements REE^{3+}, Y^{3+}, Na^{+}) ↔ Ca^{2+}; (Si^{4+}, V^{5+}, As^{5+}, S^{6+} and C^{4+}) ↔ P^{5+}; (Cl^{-}, OH^{-}) ↔ F^{-}, recording the compositional characteristics and physico-chemical conditions (e.g., temperature, oxidation state, water content, differentiation degrees) of the associated magma. Apatite lattice spacings (Williams and Cesbron, 1977), volatiles (F, Cl, SO_3; Williams and Cesbron, 1977; Piccoli and Candela, 2002; Imai, 2004; Brugge, 2021; Cao et al., 2021a; Xu et al., 2021), trace elements (Li, Mn, Fe, Sr, Cu; Belousova et al., 2002; Mao et al., 2016; Loader, 2017; Azadbakht et al., 2018; Duan and Jiang, 2018), and their ratios (e.g., Cl/F, Sr/Y, V/Y, V/Sc > 50, La/Yb_N, Eu/Eu*, Ce/Ce*, La/Sm_N, Yb/Sm_N, LREE/HREE (LREE: light REE, HREE: Heavy REE; Mao et al., 2016; Rukhlov et al., 2016; Chen et al., 2017; Loader, 2017; Azadbakht et al., 2018; Duan and Jiang, 2018; Moghadam et al., 2022; Zhou et al., 2022a) have been widely used as proxy to discern fertile magmas from barren ones or to distinguish different mineralization systems, making it a pivotal member of PIMs (Bouzari et al., 2011, 2016; Rukhlov et al., 2016; Cooke et al., 2020). However, recent studies noticed that one or two-dimensional apatite elemental concentration or element ratios for porphyry Cu-Mo-Au deposits in different regions cannot distinguish the fertile and barren magma very well (e.g., Li et al., 2021; Xing et al. 2021; Huang et al., 2022). This could result from a limited use of parameters of apatite geochemical parameters. The geochemical composition of apatite can be conceptualized as a multi-dimensional space, which traditional data visualization methods struggle to fully address and represent.

Supervised machine learning algorithms have been empirically validated across multiple studies as highly effective for solving high-dimensional, multi-element classification problems. These algorithms consistently outperform traditional methods, such as binary discriminant plots or elemental ratio diagrams, in terms of classification accuracy and robustness (Petrelli and Perugini, 2016; Zhao et al., 2019; Wang et al., 2021; Qin et al., 2022; Chen et al., 2023; ZhangZhou et al., 2024). Furthermore, investigations into the explanation of machine learning offer novel insights and corroborating evidence for established ore type prediction models (Zhang et al., 2023; Qiu et al., 2024) and magma fertility evaluation (Nathwani et al., 2022; Zhou et al., 2022b; Zou et al., 2022; Zheng et al., 2024; Ge et al., 2025; Liang et al., 2025). However, there remains insufficient attention to the interpretability of machine learning models and their linkage to geological processes.

In this contribution, two supervised machine learning algorithms, Extreme Gradient Boosting (XGBoost) and Random Forest (RF), were utilized to evaluate the fertility of magmas related to porphyry-skarn Cu/Mo/Au deposits based on apatite geochemical composition as XGBoost and RF can offer intrinsic feature importance measures and have better performance during our experimental evaluation. The algorithms were first trained using the volatile-trace-element chemistry of apatite grains from barren and fertile rocks worldwide. After performance evaluation, the fertility of apatites from host and ore-forming magmatic rocks sampled from the Güzelyayla porphyry Cu-Mo deposit, Eastern Sakarya Zone arc, Turkey (Moghadam et al., 2025) was assessed using the trained models. We demonstrate that the two new models outperform traditional methods and can effectively support porphyry-skarn exploration in greenfield areas.

2. Data compilation and labeling

2.1. Data compilation

All the apatite geochemical data utilized for model training and construction were collected from the literature (Tables S1, S2). To guarantee the data quality, the following criteria were applied to filter the collected data. First, only analyses including all elements of interest (F, S, Cl, Mn, Sr, Y, REE, see below) in this study were retained. Second, analyses which are reported with any kind of alteration were rejected as

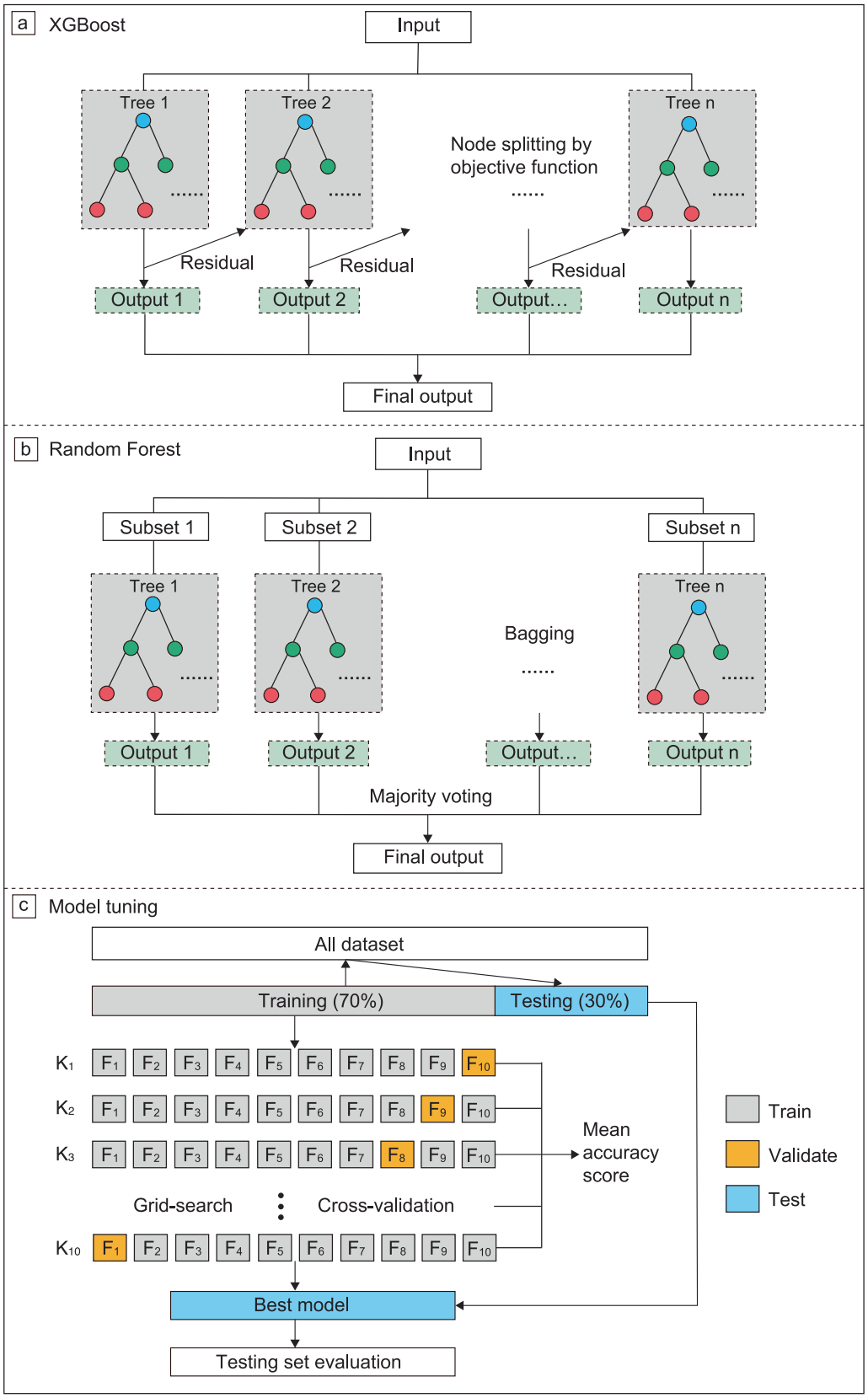
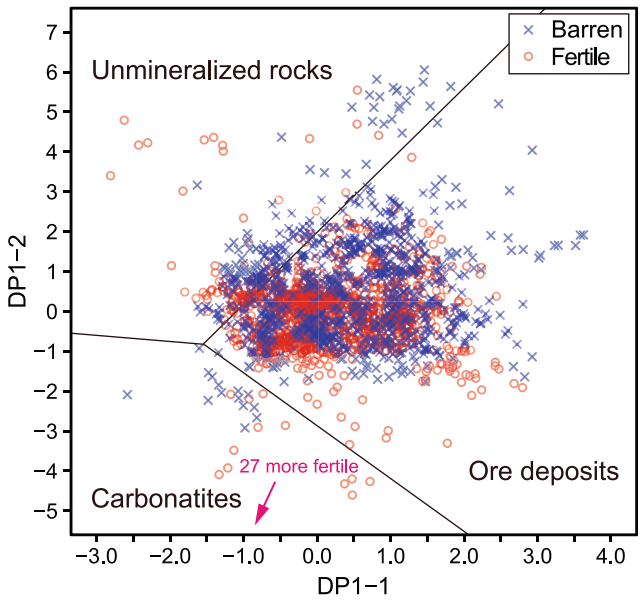


Fig. 2. Schematics of ensemble machine learning algorithms and model tuning process used in this study. (a) XGBoost, an optimized gradient boosting framework that uses a series of decision trees and residual learning to improve prediction accuracy. (b) Random Forest, where a dataset is sub-sampled n times, and a classification tree is constructed for each subset to predict the label (e.g., fertility). The final prediction is determined by averaging the results of all trees. (c) A tenfold cross-validation workflow.



Counts of TP, FP, TN, FN	Evaluation parameters
SUM = 1853 (1043 fertile + 810 barren) TP = 923 FP = 655 FN = 120 TN = 155	Accuracy = 58.2% Precision = 58.5% TPR (recall) = 88.5% FPR = 80.9% F1 = 70.4%

Fig. 3. The first discriminant function (DP1-1) versus the second discriminant function (DP1-2) discrimination diagram for apatite (Step 1) grouped by fertility after Mao et al. (2016). DP1-1 = -0.06461*log(Mn) − 1.55971*log(Sr) + 2.60894*log(Y) + 0.36314*log(La) − 1.76617*log(Ce) + 0.62425*log(Eu) − 3.64151*log(Dy) + 0.70865*log(Yb) − 1.17835*log(Pb) + 0.41613*log(Th) + 0.96296*log(U) + 6.58860; DP1-2 = 0.20726*log(Mn) − 1.03478*log(Sr) + 15.10377*log(Y) + 4.99528*log(La) − 5.80367*log(Ce) + 0.17409*log(Eu) − 8.77133*log(Dy) − 4.32634*log(Yb) + 2.02248*log(Pb) − 0.67191*log(Th) + 0.02096*log(U) − 10.44899 (all elements in ppm; Mao et al., 2016). The boundaries are from Rukhlov et al. (2016). TP = true positive, FP = false positive, FN = false negative, TN = true negative.

Table 1
Performance evaluation of XGBoost and RF models.

Algorithm	Training set 10-fold cross validation accuracy	Training set 10-fold cross validation F1 score	Testing set accuracy	Testing set F1 score
XGBoost	0.9995	0.9995	0.9373	0.9438
RF	0.9700	0.9737	0.8943	0.9075

hydrothermal alteration could modify the contents of fluid-mobile elements in apatite (Bouzari et al., 2016; Cao et al., 2021b; Qu et al., 2022). Third, apatite grains with abnormally high REE values (over twenty-folds higher) than spots from the same sample were discarded as the laser likely hits REE-rich inclusions (monazite, xenotime; Harlov, 2015). Finally, F concentrations exceeding 36,390 ppm were excluded while building the models, as anomalously high F concentrations in apatite may indicate analytical errors or F-carbonate complexation (Piccoli and Candela, 2002). After filtering, the dataset comprises 2995 items in total, with 1670 apatite data from fertile rocks and 1325 from barren rocks and includes 20 elements (F, S, Cl, Mn, Sr, Y, La, Ce, Pr, Nd, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu). Besides, a dataset comprising 1,853 apatite analyses, including concentrations of Mn, Sr, Y, La, Ce, Eu, Dy, Yb, Pb, Th, and U, along with calculated DP1-1 and DP1-2 values following Mao et al. (2016), was utilized to evaluate the performance of the proposed method against established approaches (Table S3).

The F/Cl, Sr/Y, La/Sm_N, Gd/Lu_N, La/Yb_N, Ce/Ce* and Eu/Eu*, LREE (La, Ce, Pr, Nd), MREE (Sm, Eu, Gd, Tb, Dy), HREE (Ho, Er, Tm, Yb, Lu), REE, LREE/HREE were calculated and added to the dataset that was used to train and test the models. Since both high S, As, V and low Mn content in apatite indicate high oxygen fugacity (Miles et al., 2014, 2016; Rukhlov et al., 2016), we also calculated S/Mn ratio, which could enlarge the signal of redox state.

In addition to the global apatite dataset, apatite element data of barren host granitic rocks and ore-forming dacite porphyry from the Güzelyayla porphyry Cu-Mo deposit in the Eastern Sakarya Zone arc, Turkey (Fig. 1; Table S4; Moghadam et al., 2025) were utilized to evaluate the built machine learning classifiers.

2.2. Labelling and classifying fertility

In this study, fertile magmas refer to magmas that are associated with ore formation, while barren magmas are not (Wilkinson, 2013; Richards, 2022). Apatite element data collected from representative mineralized intrusions from 61 economic porphyry and skarn deposits around the world were labeled as "fertile apatite", whereas 94 apatite data of intrusions or volcanic rocks from non-mineralized areas (>~5 km from ore bodies) as well as pre-ore (>2 Myr earlier than mineralization age) and post-ore (>2 Myr later than mineralization age) igneous rocks from known ore districts without mineralization were labeled as "barren apatite" (c.f., Nathwani et al., 2022). The geodynamic settings involved include island arc (n = 384), continental arc (n = 1203), and post-collisional orogen (n = 1527) (Fig. 1a). The ages for the fertile apatite are mostly Mesozoic and Cenozoic, with minor Paleozoic, and the barren apatite data show a similar age span but with a slightly higher number of Paleozoic data (Fig. 1b).

3. Methods

3.1. Supervised machine learning algorithms

All machine learning methods were applied using a Python-coded package, *scikit learn* (Pedregosa et al. 2011).

3.1.1. Extreme gradient boosting (XGBoost)

XGBoost is constructed by a gradient-boosting decision tree method (Fig. 2a; Friedman, 2001). XGBoost employs a gradient boosting framework wherein new predictive models are iteratively generated to estimate errors from prior iterations. These models' outputs are subsequently ensembled to yield the ultimate prediction. The term "gradient boosting" originates from the technique's use of gradient descent optimization to incrementally reduce loss functions during sequential model integration. Operationally, the algorithm iteratively introduces new decision trees that correct residual errors left by preceding predictions, with final outputs compiled as the additive combination of all constituent trees' contributions. For a set of data with $i = 1, \cdots, n$ observations, label element $y_i \in \mathbb{R}$, and m features with feature element $x_i \in R^m$, the predicted value $\hat{y}_i$ corresponds to the summation of scores from the leaf nodes, formally computed as (Chen and Guestrin, 2016):

$$\hat{y}_i = \phi(\boldsymbol{x}_i) = \sum_{k=1}^{K} f_k(\boldsymbol{x}_i), \text{with } f_k \in \mathscr{F} \quad (1)$$

where $\mathscr{F} = \{f(\boldsymbol{x}) = w_{q(\boldsymbol{x})}\}(q: \mathbb{R}^m \to T,\ w \in \mathbb{R}^T)$ is the space of the regression trees, with additive function tree $k = 1, \cdots, K$, and each f_k corresponds to an independent tree structure q and leaf weight w. Here, q represents the structure of each tree that maps an observation to a corresponding leaf, with T representing the total number of leaves in the tree. $w_{q(x)}$ represents the set of scores computed in all leaf nodes in a tree.

The hyperparameters we focused on when tuning the XGBoost model

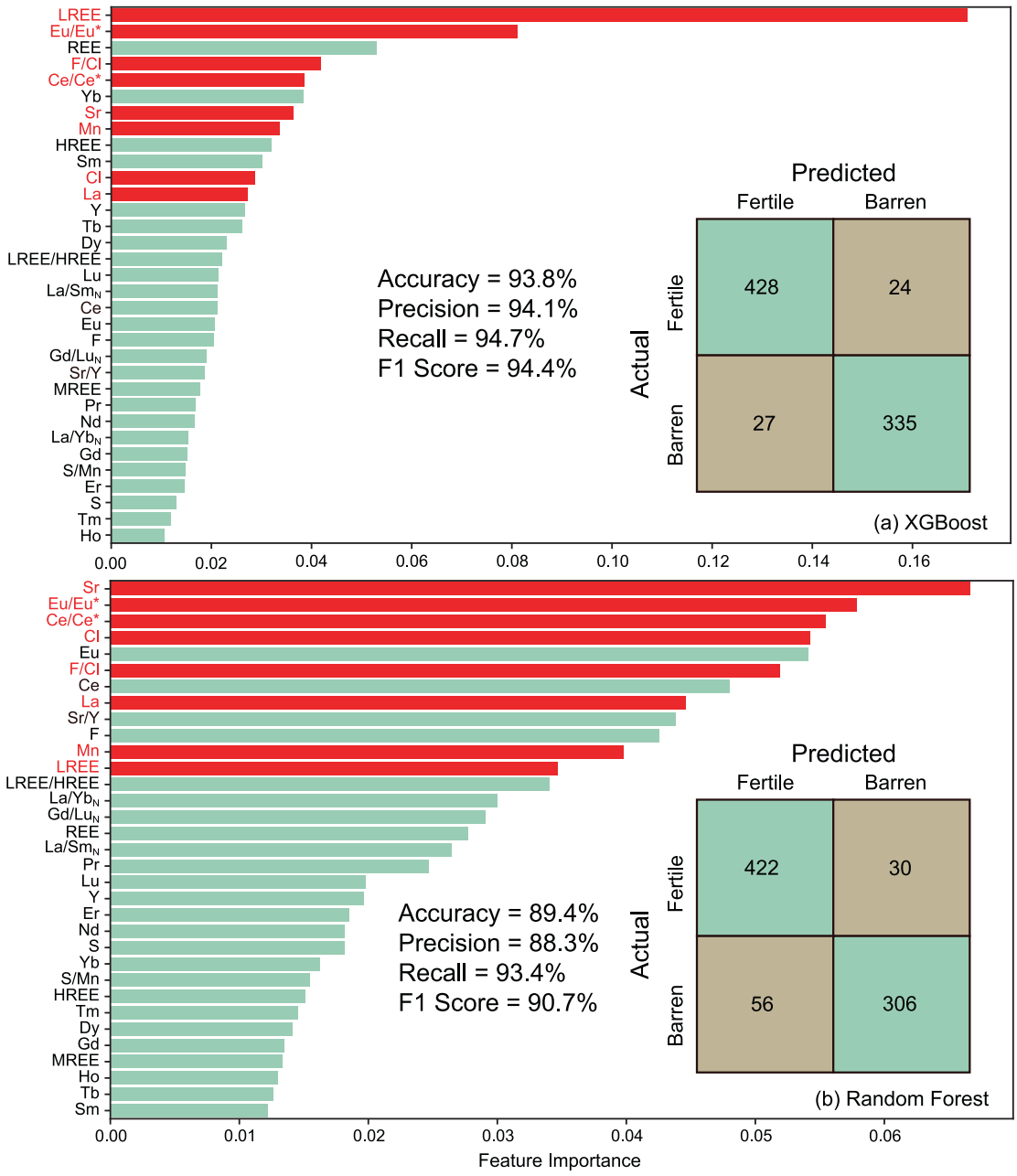


Fig. 4. Results of the XGBoost and RF model trained on the labeled dataset to classify fertile and barren apatite. Relative feature importance of apatite geochemical parameters in (a) XGBoost and (b) RF models. Inserts show confusion matrices for test subset classification by each model. Red variables denote the eight key parameters that consistently rank among the top 15 in both methodologies.

included the learning rate (eta), minimum split loss (gamma), and maximum tree depth (max_depth), each summarized below. Learning rate (eta) determines the step size at each iteration while updating the model's weights. The update rule with learning rate can be represented as:

$$W_{new} = W_{old} - \eta \times \nabla \text{Loss} \quad (2)$$

where W_{new} represents the updated weight, W_{old} is the previous weight, η is the learning rate, and ∇Loss denotes the gradient of the loss function.

Minimum split loss (gamma) specifies the minimum loss reduction required for node splitting, acting as a regularization parameter that constrains tree growth. Higher values reduce splits, lowering complexity to mitigate overfitting. Splits occur only if loss reduction exceeds gamma. Maximum tree depth (max_depth) limits tree vertical expansion, balancing complex feature capture against overfitting risk. Shallower trees enhance generalization by restricting node-to-leaf path length. Both parameters optimize model performance through controlled complexity.

3.1.2. Random forest (RF)

RF is an ensemble learning technique that combines several decision trees to identify the most popular classification (the final output). In RF, each decision tree is built by a bootstrap resampling of the training dataset, thereby reducing the variance of the decision tree (Fig. 2b; Breiman, 2001). Hyperparameter tuning is the process of optimizing the parameters that govern the learning process of the RF algorithm to achieve the best possible model performance. We incorporate several key hyperparameters, including n_estimators (the number of trees in the forest), max_depth (the maximum depth of each tree), min_samples_split

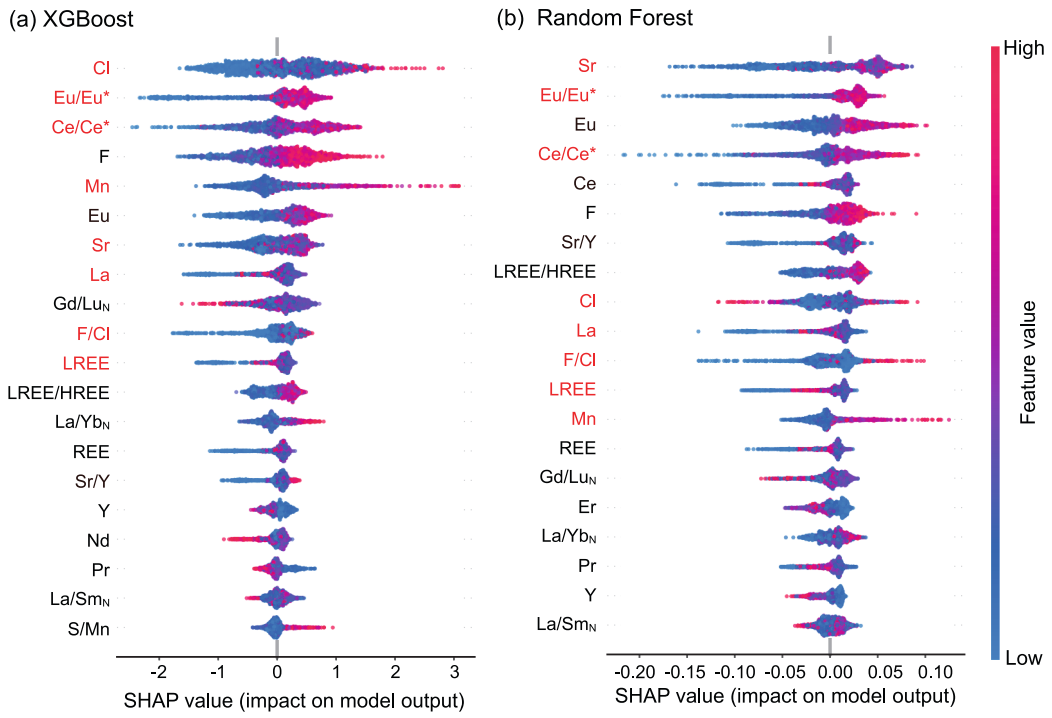


Fig. 5. SHAP summary plots for two different methods in this study. (a) XGBoost-based and (b) RF-based SHAP plot for apatites. Red variables denote the eight key parameters that consistently rank among the top 15 in feature importance (Fig. 4).

(the minimum number of samples required to split an internal node), min_samples_leaf (the minimum number of samples required to be at a leaf node), and max_features (the number of features considered when looking for the best split).

3.2. Model construction

3.2.1. Feature selection and model hyperparameter tuning

After comparative evaluation of different feature subsets, all 33 features (F, S, Cl, Mn, Sr, Y, La, Ce, Pr, Nd, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu, F/Cl, S/Mn, Sr/Y, La/Sm$_N$, Gd/Lu$_N$, La/Yb$_N$, Ce/Ce*, Eu/Eu*, LREE, MREE, HREE, REE, and LREE/HREE) in the dataset were ultimately selected for model training to optimize both performance and interpretability.

Model tuning was accomplished using a grid search strategy with the ten-fold cross-validation method (Fig. 2c; Kohavi, 1995), which exhaustively generates candidates from a grid of parameter values. This process involves training the model on nine folds and validating it on the excluded fold, repeated ten times. This strategy will evaluate the model according to different parameter values within the search range and find the best parameter combination by promoting the output scores over the training set. The best parameter combination with the highest output score suggested by the grid search was chosen as the final hyperparameters, which were used to retrain the final chosen classifiers.

3.2.2. Model evaluation

The processed dataset was randomly split 7:3 into training (70 %) and test sets (30 %; Fig. 2c). Classifiers were trained exclusively on the training data and subsequently evaluated using the held-out test set.

Model performance was assessed through classification outcomes: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN), where TP/TN represent desired correct predictions and FP/FN denote classification errors (Kubat, 2021). Based on these outcomes, all models were quantitatively evaluated using accuracy, precision, recall (true positive rate, TPR), false positive rate (FPR), and F1 scores, calculated per standard binary classification formulas (Kubat, 2021):

$$\text{Accuracy} = (N_{TP} + N_{TN}) / (N_{TP} + N_{TN} + N_{FP} + F_{FN}) \quad (3)$$

$$\text{Precision} = N_{TP} / (N_{TP} + N_{FP}) \quad (4)$$

$$\text{Recall (TPR)} = N_{TP} / (N_{TP} + N_{FN}) \quad (5)$$

$$\text{FPR} = N_{FP} / (N_{FP} + N_{TN}) \quad (6)$$

$$\text{F1} = 2 \cdot (\text{Precision} \cdot \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (7)$$

Accuracy quantifies overall correctness (TP + TN versus total predictions), while precision measures true positives among positive classifications, and recall (sensitivity/TPR) calculates the proportion of actual positives identified. The false positive rate (FPR) inversely corresponds to TPR, whereas the F1-score harmonizes precision and recall via harmonic mean – serving as the optimal metric for imbalanced data where elevated values indicate superior model performance.

4. Results and discussion

4.1. Conventional apatite fertility indicators

As a general principle, igneous apatite can effectively reflect their whole-rock geochemical signatures. Although many workers have qualitatively discriminated fertile igneous suites from barren ones using several apatite fertility indicators (e.g., high S, Brugge, 2021; Frei, 1996; high Sr, Loader, 2017; high Cl, SO_3, Li, (La/Yb)$_N$, Eu/Eu*, Duan and Jiang, 2018; high Cl and low F/Cl, Zhong et al., 2018; high Sr/Y, V/Y, (Eu/Eu*)$_N$, Moghadam et al., 2022), only one quantitative apatite fertility indicator (a scatterplot) was constructed (Mao et al., 2016).

Based on our compiled dataset, the performance of step 1 classifier using DP1-1 vs DP1-2 (Fig. 3; Mao et al., 2016) was reassessed. Except for eliminating data without Pb, Th, or U, the data plotted here are mostly the same with those used for training the machine learning models, thus allowing comparison between published scatterplot classifier and machine learning classifiers. This scatterplot has high TPR (88.5 %), namely can identify most of the fertile igneous suites if the

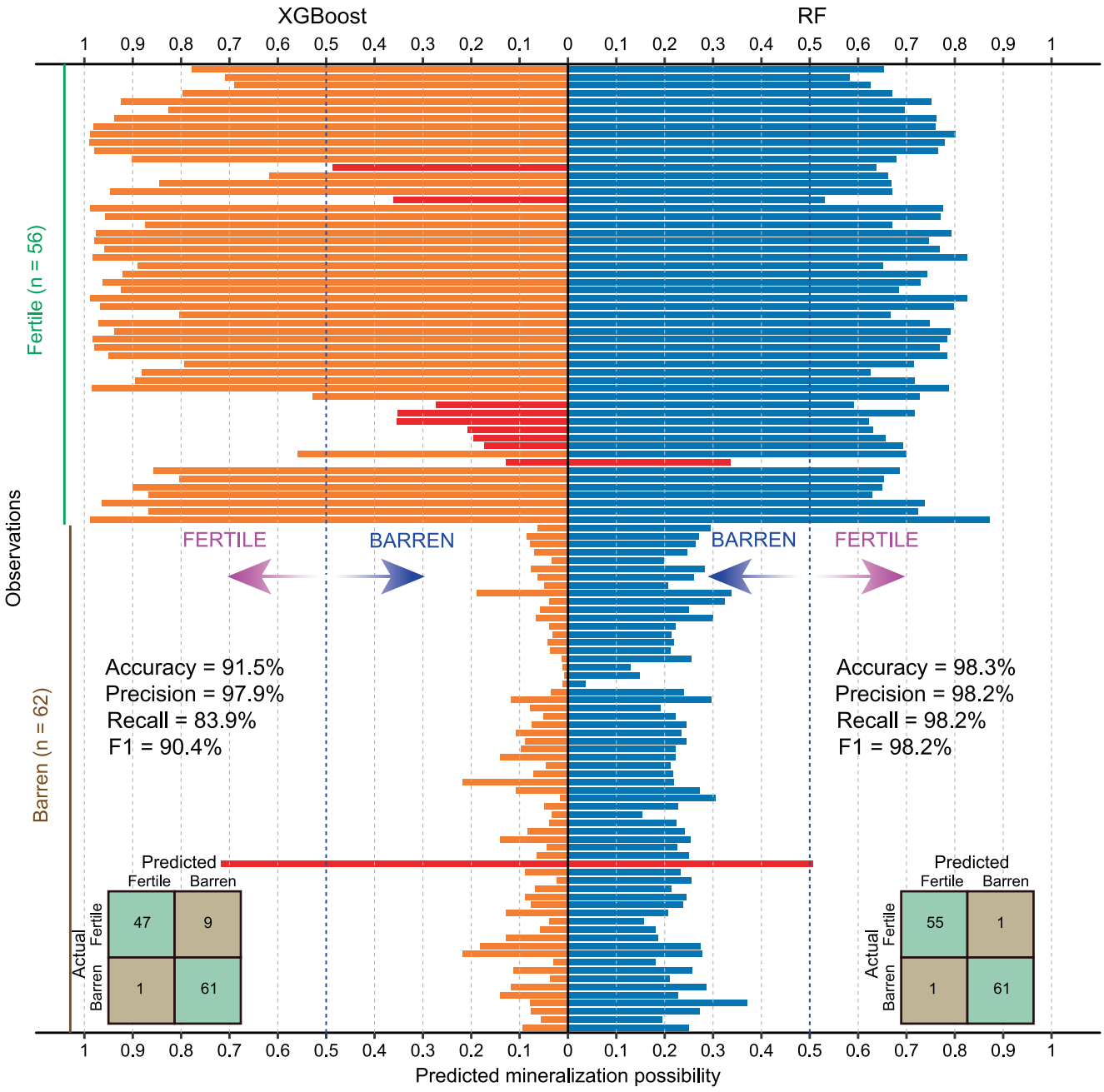


Fig. 6. Prediction results of the trained XGBoost and RF models applied to 118 unlabeled apatite grains from magmatic rocks in the Güzelyayla porphyry Cu-Mo deposit, with confusion matrices shown as inserts for each model, where red columns indicate erroneous predictions.

unknow samples are all fertile apatite. However, both the accuracy (58.2 %) and precision (58.5 %) are relatively low. Visually we can see that most of the barren apatites (655 out of 810, blue cross) were plotted within the area of "Ore deposits" (Fig. 3), indicating that this scatterplot has 80.9 % probability (FPR) to mis-identify a barren apatite as a fertile apatite. By contrast, this scatterplot also produces false negative observations (120 out of 1043, i.e., 11.5 %). These misclassifications could result from the fact that some of the elements used (such as Mn, Sr, Ce, Eu, and Th) may not only be governed by the water content and oxidation state of magmas but also be affected by the crystallization of REE-bearing minerals before apatite (Rukhlov et al., 2016; Loader, 2017; Brugge, 2021). Applying a discrimination diagram with a high false positive rate (FPR = 0.81) to mineral exploration could yield erroneous results, leading to inaccurate mineralization targets and substantial financial losses. To develop more reliable exploration tools, this study integrates additional constraints (volatiles and elemental ratios) into apatite chemistry analysis, employing machine learning algorithms for enhanced accuracy.

4.2. Performances of machine learning models

The XGBoost model demonstrated superior performance in ten-fold cross-validation, achieving perfect accuracy (1.00) and a F1-score of 1.00, outperforming the RF model's respective scores of 0.97 accuracy and 0.97 F1-score (Table 1). In our models, both area under curve (AUC) and average precision (AP) values are consistent across the training ($AUC_{XGBoost} = 1.00$, $AUC_{RF} = 0.99$; $AP_{XGBoost} = 1.00$, $AP_{RF} = 0.99$) and testing datasets ($AUC_{XGBoost} = 0.98$, $AUC_{RF} = 0.96$; $AP_{XGBoost} = 0.98$, $AP_{RF} = 0.97$), with differences between training and testing sets no greater than 0.03 for each metric and model (Fig. S1). This close

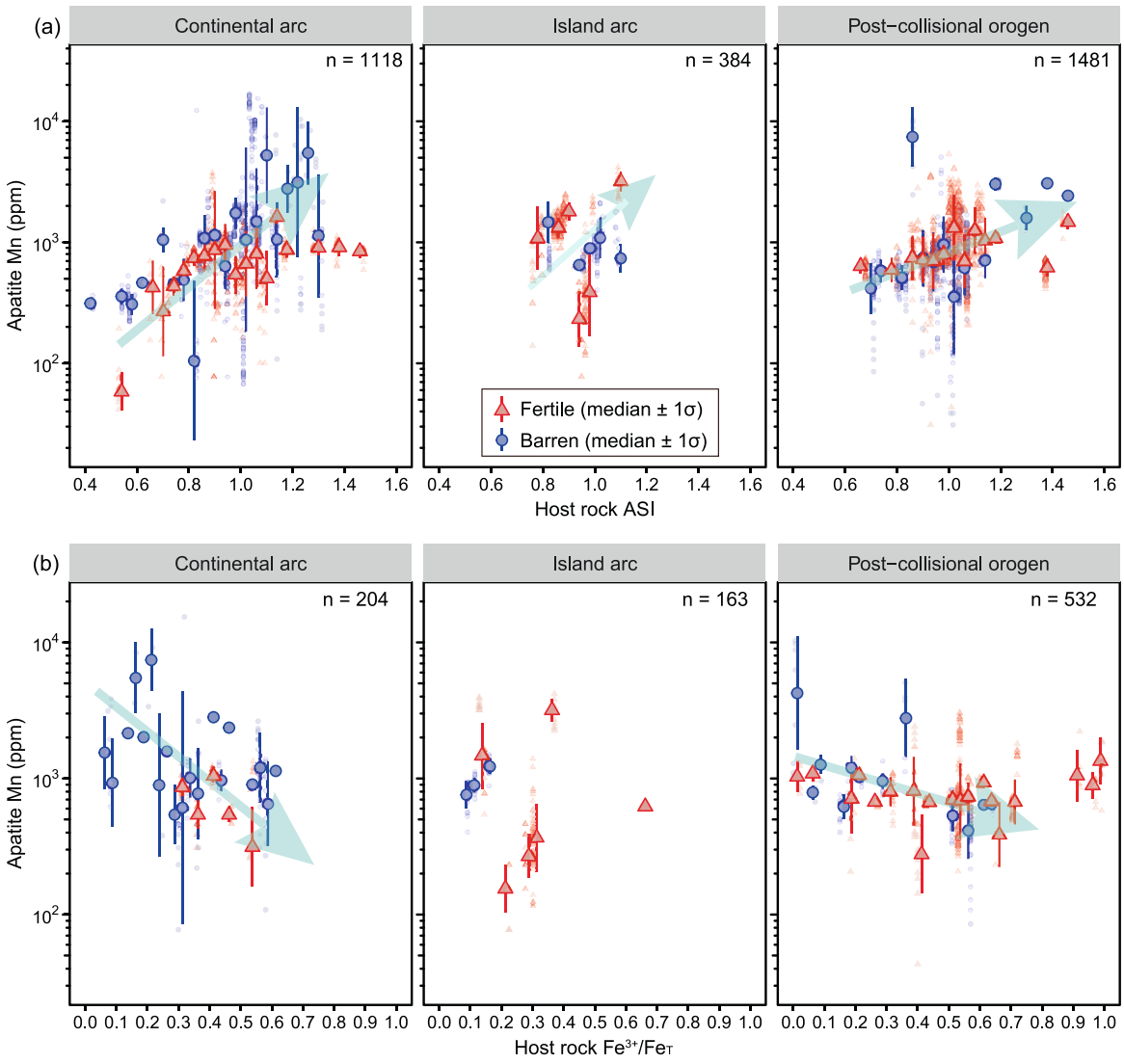


Fig. 7. (a) Apatite Mn vs host rock ASI, and (b) Apatite Mn vs host rock Fe^{3+}/Fe_T plot. ASI, Aluminium Saturation Index = molecular $Al_2O_3/(Na_2O + K_2O + CaO)$ ratio. Data are grouped by fertility of the host rock. Large symbols represent binned median values of host rock ASI (0.04 intervals) and Fe^{3+}/Fe_T (0.025 intervals) with error bars indicating one standard deviation (1σ). Small translucent symbols correspond to individual analyses. See Table S2 for data sources.

agreement suggests very limited overfitting. Upon evaluation on the testing set with optimal hyperparameters derived from grid-search-optimized cross-validation, XGBoost sustained its edge with an accuracy of 0.94 compared to RF's 0.89, alongside F1-scores of 0.94 and 0.91, respectively (Fig. 4). The two new models exhibit a superior ability to identify both fertile and barren apatites, demonstrating enhanced capability in distinguishing them compared to traditional models (Fig. 3).

The machine learning model's enhanced performance likely stems from its quantitative capabilities and ability to process high-dimensional data. By integrating a broader range of elements and their interactions, the model addresses the shortcomings of the traditional approach, providing a more nuanced and accurate discrimination of fertility and mineralization potential.

4.3. *Feature importance for classifying magma fertility*

Fertility classification depends solely on the input features provided. Our dataset includes 3 major elements (volatiles), 17 trace elements, and 13 elemental combinations of apatite as classification features. The relative abundance of trace elements within igneous apatite is controlled by melt composition, the presence of coexisting accessory phases and crystallization conditions such as temperature, redox state, H_2O content, and differentiation degree (Piccoli and Candela 2002; Miles et al. 2014), with the latter three being important parameters in determining mineralization potential (Belousova et al., 2002; Brugge et al., 2017). Understanding which of these features most effectively distinguish apatite from barren and fertile rocks is therefore crucial.

For feature importance in the optimal models (Fig. 4), higher scores indicate stronger contributions to discriminating apatite in fertile versus barren systems. In the XGBoost model, the top 15 features ranked by descending importance for mineralization identification are: LREE, Eu/Eu*, REE, F/Cl, Ce/Ce*, Yb, Sr, Mn, HREE, Sm, Cl, La, Y, Tb, and Dy (Fig. 4a). By contrast, the top 15 feature importance scores for the RF model, listed from highest to lowest, include: Sr, Eu/Eu*, Ce/Ce*, Cl, Eu, F/Cl, Ce, La, Sr/Y, F, Mn, LREE, LREE/HREE, La/Yb_N, and Gd/Lu_N (Fig. 4b).

XGBoost and RF models demonstrated different performances in classifying fertile and barren samples. According to the confusion matrix comparison, XGBoost achieved higher accuracy (94 % vs. 89 %) and had fewer misclassifications (51 out of 814), indicating its superior overall performance. Both models excelled at identifying barren samples, but XGBoost model had a slight edge in accuracy, making it more suitable when high precision is required. Both models ranked Mn, Sr, Cl, La, REE,

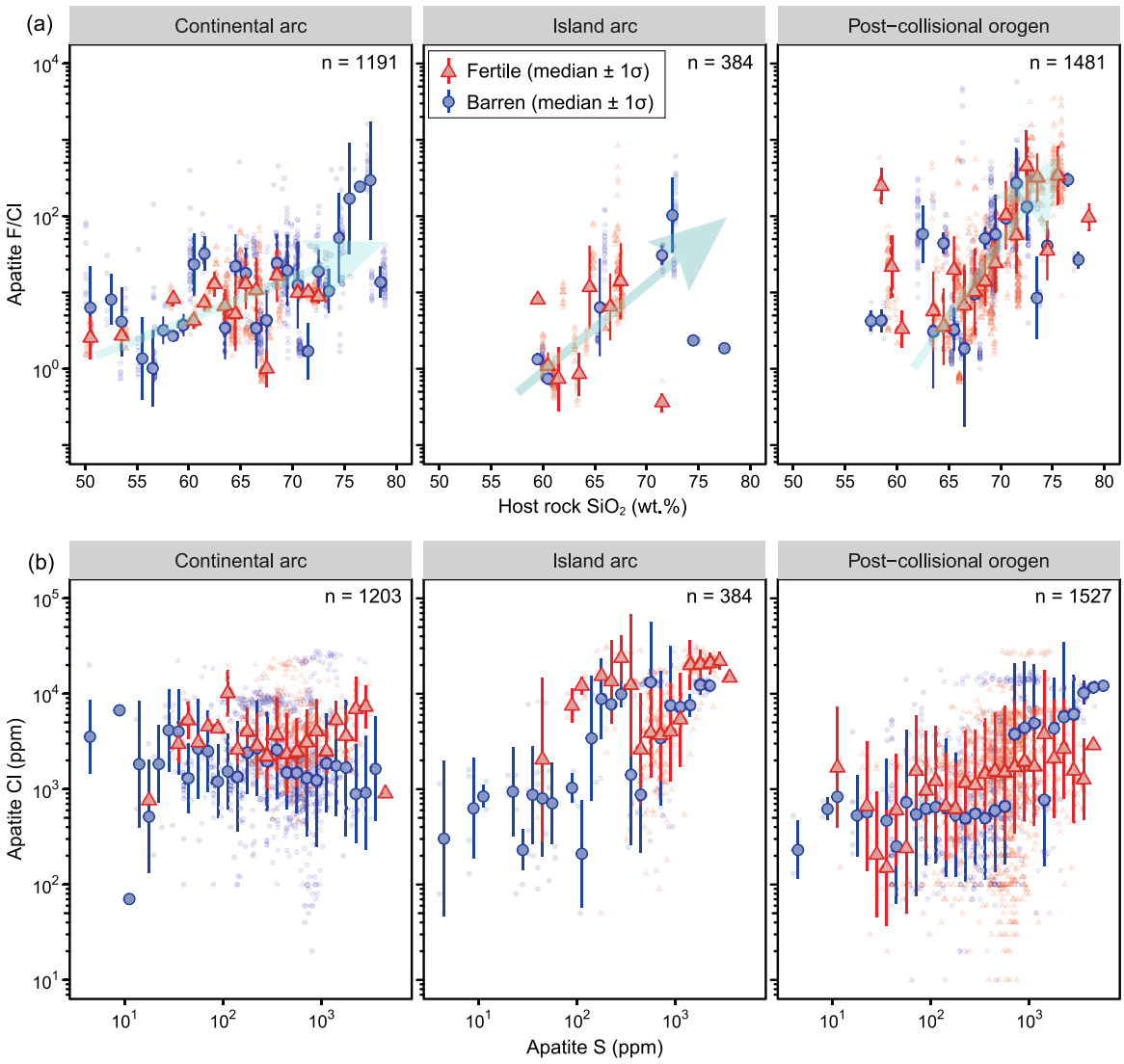


Fig. 8. (a) Apatite F/Cl vs host rock SiO_2 and (b) apatite Cl vs apatite S plots. Data are grouped by fertility of the host rock. Large symbols represent binned median values of host rock SiO_2 (1 wt% intervals) and S ($10^{0.1}$ intervals) with error bars indicating one standard deviation (1σ). Small translucent symbols correspond to individual analyses. See Table S2 for data sources.

F/Cl, Ce/Ce*, and Eu/Eu* as the most important features, underscoring their role as key indicators of magma fertility.

Additionally, the Shapley Additive Explanations (SHAP) method was used to evaluate the contribution of specific features to classifying a given sample in machine learning models (Fig. 5). The eight key features, highlighted in red (Fig. 4), also rank prominently at the top of both methods based on SHAP values (Fig. 5). Analysis of the scores reveals that elevated ratios of Eu/Eu*, Ce/Ce*, LREE/HREE, La/Yb_N, and Sr/Y, along with high concentrations of Cl, F, Mn, Eu, and Sr, positively influence the identification of fertile apatites, while Y, Nd, Pr, and Er concentrations are notably lower in such apatites. The critical role of high Eu/Eu* and Ce/Ce* ratios in identifying fertile rocks aligns with prior studies, which indicate that magmas with high water content and elevated oxidation states are essential for forming porphyry-skarn deposits (e.g., Rukhlov et al., 2016). Previous research also suggests that hydrous, fertile melts likely undergo early fractionation of amphibole ± garnet, which preferentially incorporate MREE and HREE, depleting these elements in the residual melt (Davidson et al., 2007; Richards, 2011; Chiaradia et al., 2012; Tang et al., 2020). This process likely accounts for the low Y and Er concentrations observed in fertile apatites.

Apatite F contents exhibit a positive correlation with the degree of magma fractionation, whereas Cl contents are predominantly influenced by the timing of fluid saturation and exsolution, as Cl preferentially partitions into the fluid phase compared to F (Webster, 1990; Piccoli and Candela, 2002; Kusebauch et al., 2015). Apatite S contents are mainly controlled by extent of magmatic sulfide saturation and fO_2 (Peng et al., 1997), with less sulfide saturation and high fO_2 favoring high apatite S contents. Apatite Sr content is primarily controlled by the Sr concentration in the parent magma and the degree of plagioclase fractionation (Sha and Chappell, 1999; Belousova et al., 2002; Piccoli and Candela, 2002). In high-pressure, hydrous magmatic systems where feldspar crystallization is delayed or absent but porphyry mineralization is favored (Loucks et al., 2024), apatite may incorporate higher Sr levels due to reduced competition.

Cerium (Ce) and europium (Eu) stand out among REE as the only ones exhibiting two oxidation states: Ce^{3+} and Ce^{4+} for cerium, and Eu^{2+} and Eu^{3+} for europium. Plagioclase tend to concentrate Eu^{2+}, zircon favors Ce^{4+}, and monazite preferentially incorporates Ce^{3+} (Sha and Chappell, 1999; Belousova et al., 2002; Piccoli and Candela, 2002; Miles et al., 2014). Within the apatite structure, Ce^{3+} and Eu^{2+} are preferentially incorporated (Prowatke and Klemme, 2006; Cao et al., 2012). As a REE-compatible mineral, apatite Ce and Eu contents are largely controlled by magma differentiation and water content instead of oxygen fugacity. Higher water contents not only enhance Ce^{3+}

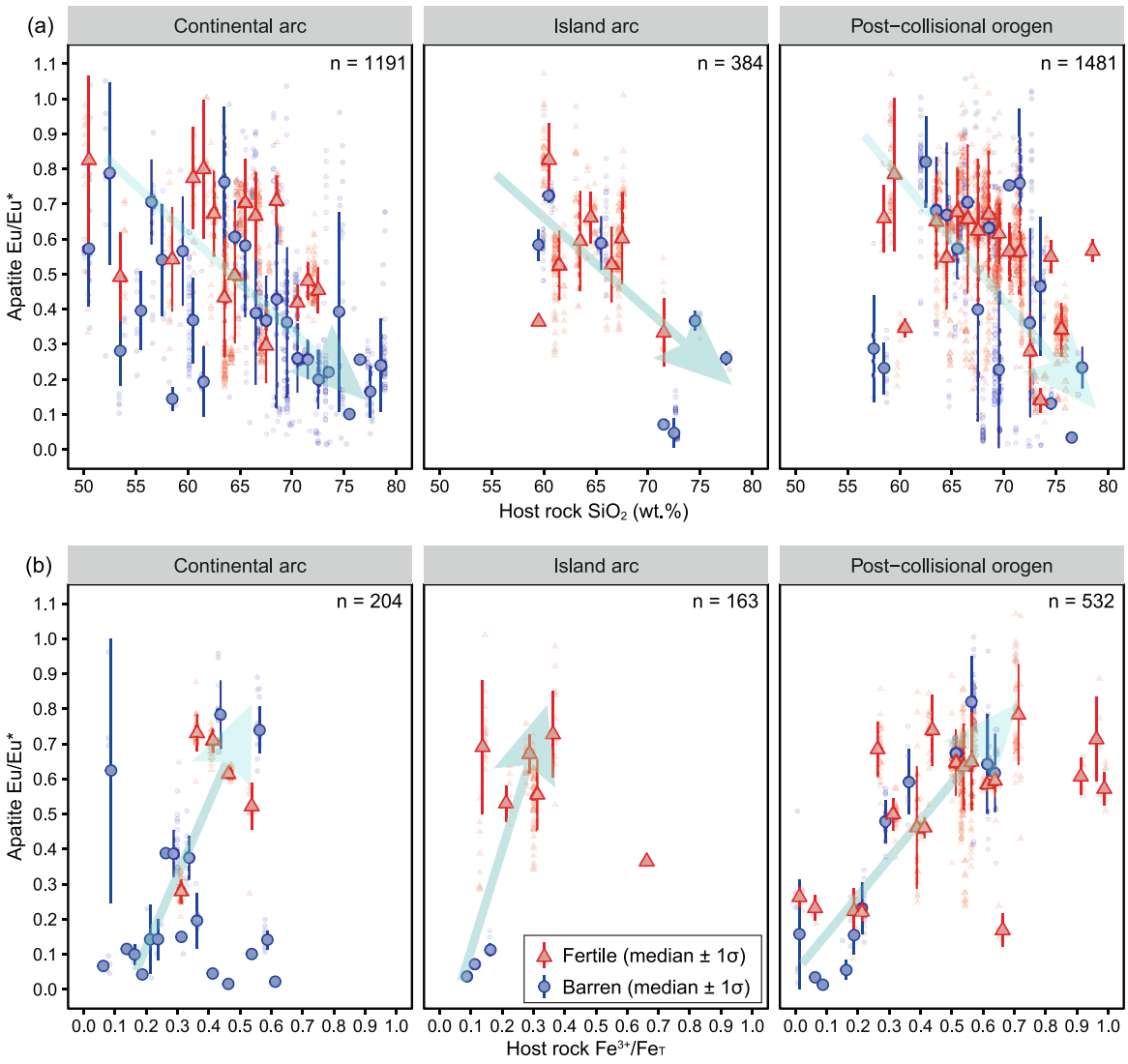


Fig. 9. (a) Apatite Eu/Eu* vs host rock SiO_2 and (b) apatite Eu/Eu* vs host rock Fe^{3+}/Fe_T plots. Data are grouped by fertility of the host rock. Large symbols represent binned median values of host rock SiO_2 (1 wt% intervals) and Fe^{3+}/Fe_T (0.025 intervals) with error bars indicating one standard deviation (1σ). Small translucent symbols correspond to individual analyses. See Table S2 for data sources.

stability (Smythe and Brenan, 2015), but also suppress crystallization of plagioclase, leading to high Ce^{3+} and Eu^{2+} concentrations in melt, and hence apatite Ce/Ce* and Eu/Eu*.

4.4. *Prediction of mineralization potential for the Güzelyayla porphyry Cu-Mo deposit*

The Güzelyayla porphyry Cu-Mo deposit is located in the Eastern Sakarya Zone (ESZ), northern Türkiye, within the broader Tethyan orogenic belt (Moghadam et al., 2025). The ESZ features a well-preserved north–south trending magmatic arc and back-arc system, extending ~500 km in length and ~100 km in width. This region hosts Late Mesozoic to Early Cenozoic magmatic arcs, including Late Cretaceous subduction-related (~120–70 Ma) and Eocene-Oligocene post-collisional magmatism (< 55 Ma), linked to the northward subduction of the northern Neotethys Ocean beneath Eurasia or alternative southward subduction models. Located south of Trabzon in the Trabzon-Gümüşhane district of the central Eastern Pontides Zone, the Güzelyayla deposit is hosted by Late Cretaceous volcano-sedimentary sequences and associated granitoids. The deposit has probable and proven reserves at 186.2 million tons with average grades of 0.3 % Cu and 0.014 % Mo. Magmatic and mineralization ages are Eocene: Re-Os molybdenite dating yields ~48 Ma for Cu-Mo mineralization (Delibaş et al., 2016), and zircon U-Pb ages range from 48.9 ± 0.3 Ma to 49.2 ± 0.2 Ma for dacite porphyries (Moghadam et al., 2025). Host Late Cretaceous rhyolites and granitoids date to 78.9 ± 0.2 Ma to 80.2 ± 0.4 Ma (zircon U-Pb; Moghadam et al., 2025). The deposit reflects post-collisional magmatism derived from an enriched sub-continental lithospheric mantle (SCLM), with adakitic signatures indicating deep crustal amphibole fractionation under hydrous, oxidized conditions (Moghadam et al., 2025).

Our XGBoost model for apatite fertility classification achieved an accuracy of 91.5 % and an F1-score of 90.4 % in predicting the mineralization potential of apatites from the Güzelyayla deposit (Fig. 6). The RF model performs better with 98.3 % accuracy and 98.2 % F1-score. Notably, for corrected predictions, XGBoost demonstrates superior performance compared to the RF model. It assigns higher mineralization probabilities to true fertile observations (typically > 0.8, compared to mostly < 0.8 for RF) and lower probabilities to true barren observations (typically < 0.2, compared to mostly > 0.2 for RF).

The models' performance (Fig. 6) effectively illustrates the robustness and generalizability of both models when applied to an independent dataset, with the RF model outperforming XGBoost in achieving balanced performance metrics, likely attributable to its ensemble

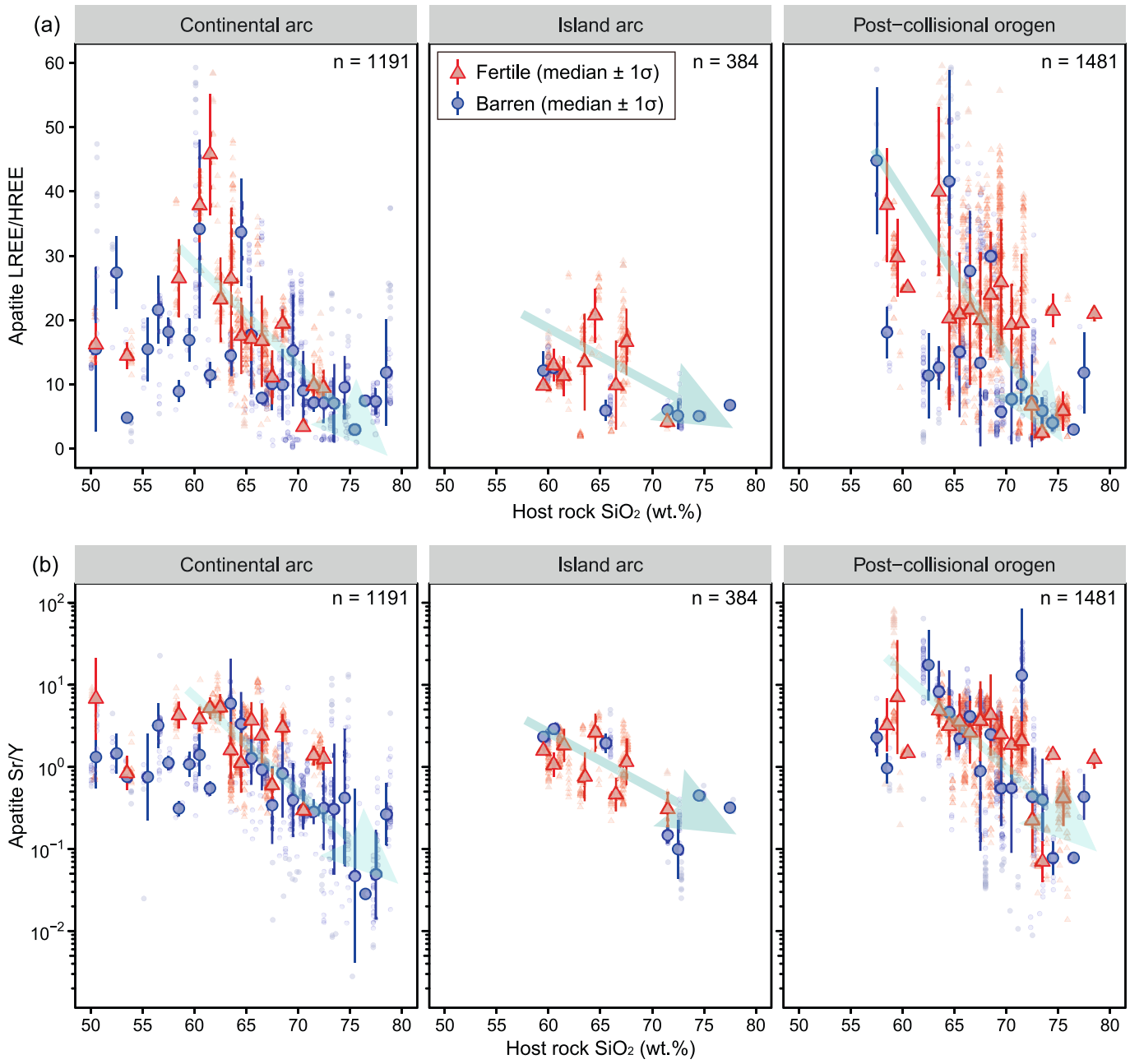


Fig. 10. (a) Apatite LREE/HREE vs host rock SiO_2 and (b) apatite Sr/Y vs host rock SiO_2 plots. Data are grouped by fertility of the host rock. Large symbols represent binned median values of host rock SiO_2 (1 wt% intervals) with error bars indicating one standard deviation (1σ). Small translucent symbols correspond to individual analyses. See Table S2 for data sources.

averaging, which minimizes variance. These findings underscore the reliability of both models for assessing porphyry fertility, although XGBoost's elevated rate of false negatives suggests a potential sensitivity to feature interactions in borderline cases.

4.5. Linking apatite geochemistry with magma fertility and ore-forming processes

Apatite Mn concentrations increase with higher aluminum saturation index (ASI) in host rocks, particularly in continental arcs and post-collisional orogens (Fig. 7a). This trend suggests that peraluminous, aluminum-rich host rocks promote Mn incorporation into apatite, largely driven by magma differentiation processes (Chu et al., 2009). Additionally, a broad negative correlation exists between apatite Mn content and host rock Fe^{3+}/Fe_T ratios in these settings (Fig. 7b), highlighting the significant influence of oxygen fugacity (fO_2) on Mn incorporation (Miles et al., 2014, 2016). In continental arcs, fertile apatites typically exhibit lower Mn contents compared to barren apatites with equivalent host rock ASI values (Fig. 7a). This indicates that fertile magmas are generally more oxidized than their barren counterparts at similar degrees of differentiation, underscoring the role of redox conditions in controlling apatite chemistry and magma fertility in corresponding setting.

Strong positive correlations between apatite F/Cl ratios and host rock SiO_2 content (Fig. 8a) indicate that the F/Cl ratios in apatite are primarily governed by the degree of magma fractionation. In continental arcs, fertile apatites generally exhibit lower F/Cl ratios compared to barren apatites with equivalent host rock SiO_2 contents, suggesting higher Cl concentrations in fertile magmas at similar stages of differentiation. Notably, across all three geological settings, most fertile

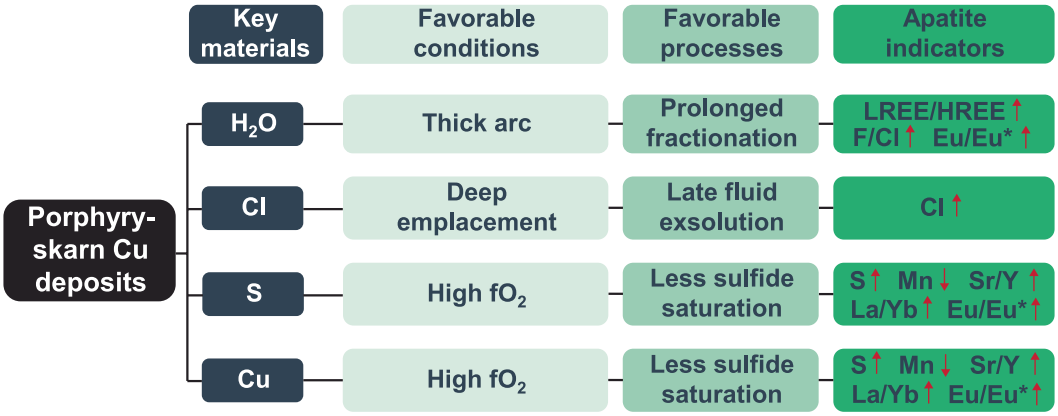


Fig. 11. Diagram illustrating the key materials, favorable geological conditions, associated processes, and diagnostic apatite indicators for porphyry-skarn Cu deposits.

apatites have higher Cl contents than barren apatites with constant sulfur (S) contents (Fig. 8b). This implies that fertile magmas are more Cl-enriched than barren magmas under similar sulfide saturation levels and oxygen fugacity (fO_2, as inferred from apatite S contents; Peng et al., 1997). Collectively, these findings highlight that Cl-rich magmas are more likely to be fertile than barren ones, assuming other parameters remain constant.

Apatite Eu/Eu* values generally decline with increasing SiO_2 content in continental arcs, island arcs, and post-collisional orogens (Fig. 9a), reflecting Eu depletion in more evolved felsic magmas due to plagioclase fractionation, which preferentially incorporates Eu^{2+}. A broad positive correlation exists between apatite Eu/Eu* and host rock Fe^{3+}/Fe_T ratios (Fig. 9b), likely because low Eu/Eu* results from extensive plagioclase fractionation under water-poor, low-pressure, and reduced conditions, while high Eu/Eu* values are associated with suppressed plagioclase crystallization in water-rich, high-pressure, oxidized environments (Loucks et al., 2024). In continental arcs, fertile magmas generally exhibit higher Eu/Eu* values (Fig. 9a), indicating their distinct hydrous and high-pressure fractionation characteristics.

Apatite LREE/HREE and Sr/Y ratios decrease with increasing SiO_2 content across continental arcs, island arcs, and post-collisional orogens (Fig. 10), indicating fractionation as a key control on these ratios. Fertile apatites broadly exhibit higher LREE/HREE and Sr/Y ratios compared to barren apatites in all three settings. These trends suggest that extensive fractionation promotes porphyry-skarn Cu mineralization (e.g., Yuan et al., 2025).

Establishing links between apatite geochemistry and the source materials as well as ore-forming processes of porphyry deposits is essential for advancing our understanding of their genesis. Notably, H_2O, Cl, S, and Cu are critical constituents of porphyry-skarn Cu deposits (Meinert et al., 2005; Sillitoe, 2010; Fig. 11). Magmas in thick continental arcs exhibit high water solubility, facilitating protracted fractional crystallization and water accumulation in shallow magma chambers, conditions favorable for Cu mineralization (Loucks et al., 2024). Such hydrous, high-pressure fractionation in thick arcs typically involves amphibole and garnet crystallization with suppressed plagioclase formation (Richards and Kerrich, 2007; Richards, 2011; Chiaradia et al., 2012; Loucks, 2014; Loucks et al., 2024), leading to elevated Sr/Y, La/Yb_N, LREE/HREE, and Eu/Eu* ratios in apatite. Chlorine, a significant anion ligand for Cu, is enriched in apatite due to late-stage fluid exsolution associated with deep porphyry emplacement, common in subduction settings and scarce in post-collisional settings (Huang et al., 2023). A primary barrier to transporting substantial quantities of Cu/Mo/Au and S to mineralization levels is sulfide saturation, which can be mitigated by high oxygen fugacity. Fractionation of garnet and amphibole elevates magma fO_2 (Tang et al., 2018; Zhang et al., 2022; Wang et al., 2025), enhancing Sr/Y and La/Yb_N ratios, stabilizing oxidized sulfur (S^{6+}, which readily substitutes for P^{5+} in apatite), and destabilizing Mn^{2+} (which substitutes for Ca^{2+} in apatite). Consequently, apatite from fertile magmas exhibits high Sr/Y, La/Yb_N, and S contents but low Mn content (Fig. 11).

5. Conclusions

Our study demonstrates that XGBoost and RF models significantly outperform traditional methods in distinguishing fertile from barren magmas, attaining accuracies of 94 % and 89 %, respectively. These two models exhibit robust and generalizable performance in identifying fertile and barren rocks from the Güzelyayla deposit as a blind test, achieving accuracies of 92 % (XGBoost) and 98 % (RF), respectively. The interpretability of key features in the XGBoost and Random Forest models—such as Eu/Eu*, F/Cl, and Mn—aligns with established geochemical models.

Our analysis of apatite geochemistry reveals clear patterns linked to fertile magmas capable of forming porphyry deposits. Key indicators include lower Mn content, low F/Cl ratios, high Cl content, elevated Eu/Eu* values, and higher LREE/HREE and La/Yb_N ratios. Collectively, these markers underscore the pivotal roles of oxidation state, volatile contents, and magma differentiation in mineralization processes, and ultimately highlighting the significance of water, Cl, S, and metals in porphyry-skarn Cu/Mo/Au deposits. These findings enhance our understanding of the geochemical conditions and processes that favor porphyry mineralization. In summary, machine learning provides a robust, quantitative framework for mineralization and genetic studies, surpassing traditional methods in both reliability and depth of insight. Moving forward, incorporating additional factors—such as structural favorability, erosion levels, and alteration intensity—could further refine these models, boosting exploration success in greenfield areas and supporting sustainable resource development.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Le Wang acknowledges support from the National Natural Science Foundation of China (NSFC, 42202085) and China Scholarship Council (202104910161). Ben Qin acknowledges support from the Inner Mongolia Autonomous Region Natural Science Foundation (2024QN04004) and Research Special Program for First-Class Disciplines (YLXKZX-NGD-052). Ming-Jian Cao acknowledges support from

NSFC (42122013). Le Wang thanks Xiao-Yu Ma for beneficial discussions. We thank Editor-in-Chief Hua-Yong Chen and the associate editor for their handling. Reviews by two anonymous reviewers greatly improved the quality of this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.oregeorev.2025.106926.

Data availability

Data will be made available on request.

References

Azadbakht, Z., Lentz, D.R., McFarlane, C.R., 2018. Apatite chemical compositions from Acadian-related granitoids of New Brunswick, Canada: Implications for petrogenesis and metallogenesis. Minerals 8, 598.

Baldwin, J.A., Pearce, J.A., 1982. Discrimination of productive and nonproductive porphyritic intrusions in the Chilean Andes. Econ. Geol. 77 (3), 664–674.

Belousova, E.A., Griffin, W.L., O'Reilly, S.Y., Fisher, N.I., 2002. Apatite as an indicator mineral for mineral exploration: Trace-element compositions and their relationship to host rock type. J. Geochem. Explor. 76, 45–69.

Bouzari, F., Hart, C.J.R., Barker, S., Bissig, T., 2011. Porphyry indicator minerals (PIMs): Exploration for concealed deposits in central British Columbia. Geoscience BC Report 2011-17, 31 p.

Bouzari, F., Hart, C.J.R., Bissig, T., Barker, S., 2016. Hydrothermal alteration revealed by apatite luminescence and chemistry: a potential indicator mineral for exploring covered porphyry copper deposits. Econ. Geol. 111, 1397–1410.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Brugge, E., 2021. Apatite in porphyry systems and its applications in mineral exploration. Ph.D. thesis, Imperial College London, London, 289 p.

Brugge, E., Wilkinson, J.J., Miles, A., 2017. Habit and chemistry of apatite from Chuquicamata. Proceedings, Chile. In, pp. 287–290.

Cao, K., Yang, Z.M., Hou, Z.Q., White, N.C., Yu, C., 2021a. Contrasting porphyry Cu fertilities in the Yidun Arc, eastern Tibet: Insights from zircon and apatite compositions and implications for exploration. SEG Spec. Publ. 24, 231–255.

Cao, M., Li, G., Qin, K., Seitmuratova, E.Y., Liu, Y., 2012. Major trace element characteristics of apatites in granitoids from Central Kazakhstan: Implications for petrogenesis and mineralization. Resour. Geol. 62, 63–83.

Cao, M.J., Evans, N.J., Hollings, P., Cooke, D.R., McInnes, B.I.A., Qin, K.Z., 2021b. Apatite texture, composition, and O-Sr-Nd isotope signatures record magmatic and hydrothermal fluid characteristics at the Black Mountain porphyry deposit. Philippines. Econ. Geol. 116, 1189–1207.

Chen, G.X., Cheng, Q.M., Puetz, S., 2023. Special issue: Data driven discovery in geosciences: opportunities and challenges. Math. Geosci. 55 (3), 287–293.

Chen, L., Yan, Z., Wang, Z.Q., Wang, K.M., 2017. Characteristics of apatite from 160~140 Ma Cu (Mo) and Mo (W) deposits in East Qinling. Acta Geol. Sin. 91, 1925–1941 in Chinese with English abstract.

Chen, T., Guestrin, C., 2016. XGBoost: a scalable tree boosting system. In: In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794.

Chiaradia, M., 2021. Magmatic controls on metal endowments of porphyry Cu-Au deposits. SEG Spec. Publ. 24, 1–16.

Chiaradia, M., Bellver-Baca, M.T., Ulyanov, A., 2025. Progressive build-up of a transcrustal system beneath an adakite-like volcanic complex (Chachimbiro, Ecuador): an example of an embryonic porphyry Cu system. J. Petrol. 66, egaf009.

Chiaradia, M., Merino, D., Spikings, R., 2009. Rapid transition to long-lived deep crustal magmatic maturation and the formation of giant porphyry-related mineralization (Yanacocha, Peru). Earth Planet. Sci. Lett. 288, 505–515.

Chiaradia, M., Ulianov, A., Kouzmanov, K., Beate, B., 2012. Why large porphyry Cu deposits like high Sr/Y magmas? Sci. Rep. 2, 685.

Chu, M.F., Wang, K.L., Griffin, W.L., Chung, S.L., O'Reilly, S.Y., Pearson, N.J., Iizuka, Y., 2009. Apatite composition: Tracing petrogenetic processes in Transhimalayan granitoids. J. Petrol. 50, 1829–1855.

Cooke, D.R., Agnew, P., Hollings, P., Baker, M., Chang, Z., Wilkinson, J.J., Ahmed, A., White, N.C., Zhang, L., Thompson, J., Gemmell, J.B., 2020. Recent advances in the application of mineral chemistry to exploration for porphyry copper–gold–molybdenum deposits: Detecting the geochemical fingerprints and footprints of hypogene mineralization and alteration. Geochem. Explor. Environ. Anal. 20, 176–188.

Davidson, J., Turner, S., Handley, H., Macpherson, C., Dosseto, A., 2007. Amphibole "sponge" in arc crust? Geology 35, 787–790.

Delibaş, O., Moritz, R., Ulianov, A., Chiaradia, M., Saraç, C., Revan, K.M., Göç, D., 2016. Cretaceous subduction-related magmatism and associated porphyry-type Cu–Mo prospects in the Eastern Pontides, Turkey: new constraints from geochronology and geochemistry. Lithos 248, 119–137.

Duan, D.F., Jiang, S.Y., 2018. Using apatite to discriminate synchronous ore-associated and barren granitoid rocks: a case study from the Edong metallogenic district, South China. Lithos 310, 369–380.

Elliot, J.C., 1994. Structure and chemistry of the apatites and other calcium orthophosphates. Elsevier, Amsterdam, p. 389.

Frei, R., 1996. Sulfur in bulk rock and igneous apatite; tracing mineralized and barren trends in intrusions. Schweiz. Mineral. Petrogr. Mitt. 76, 57–73.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29, 1189–1232.

Ge, Y.Z., Zhang, Z.J., Zhou, Y.Z., Li, Q., Zhang, F., 2025. Explainable machine learning reveals apatite fertility and porphyry copper mineralization processes in the syn- and post-subduction settings. Ore Geol. Rev. 169, 106679.

Grondahl, C., Zajacz, Z., 2022. Sulfur and chlorine budgets control the ore fertility of arc magmas. Nat. Commun. 13, 1–11.

Grunsky, E.C., de Caritat, P., 2020. State-of-the-art analysis of geochemical data for mineral exploration. Geochem. Explor. Environ. Anal. 20, 217–232.

Harlov, D.E., 2015. Apatite: a fingerprint for metasomatic processes. Elements 11, 171–176.

Holliday, J.R., Cooke, D.R., 2007. Advances in geological models and exploration methods for copper ± gold porphyry deposits. In: In: Proceedings of Exploration, pp. 791–809.

Hoskin, P.W., Kinny, P.D., Wyborn, D., Chappell, B.W., 2000. Identifying accessory mineral saturation during differentiation in granitoid magmas: an integrated approach. J. Petrol. 41, 1365–1396.

Huang, M.L., Zhu, J.J., Bi, X.W., Xu, L.L., Xu, Y., 2022. Low magmatic Cl contents in giant porphyry Cu deposits caused by early fluid exsolution: a case study of the Yulong belt and implication for exploration. Ore Geol. Rev. 141, 104664.

Huang, M.L., Zhu, J.J., Chiaradia, M., Hu, R.Z., Xu, L.L., Bi, X.W., 2023. Apatite volatile contents of porphyry Cu deposits controlled by depth-related fluid exsolution processes. Econ. Geol. 118 (5), 1201–1217.

Hughes, J.M., Rakovan, J.F., 2015. Structurally robust, chemically diverse: Apatite and apatite supergroup minerals. Elements 11, 165–170.

Imai, A., 2004. Variation of Cl and SO3 contents of microphenocrystic apatite in intermediate to silicic igneous rocks of Cenozoic Japanese island arcs: Implications for porphyry Cu metallogenesis in the Western Pacific Island arcs. Resour. Geol. 54, 357–372.

Jennings, E.S., Marschall, H.R., Hawkesworth, C.J., Storey, C.D., 2011. Characterization of magma from inclusions in zircon: Apatite and biotite work well, feldspar less so. Geology 39, 863–866.

Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. Proceed. Fourteenth Int. Joint Conference on Artificial Intelligence 2, 1137–1145.

Kubat, M., 2021. An introduction to machine learning (3rd ed.). Springer Nature Switzerland AG, Cham, Switzerland, 458 p.

Kusebauch, C., John, T., Whitehouse, M.J., Engvik, A.K., 2015. Apatite as probe for the halogen composition of metamorphic fluids (Bamble Sector, SE Norway). Contrib. Miner. Petrol. 170, 1–20.

Li, Q., Sun, X., Lu, Y., Wang, F., Hao, J., 2021. Apatite and zircon compositions for Miocene mineralizing and barren intrusions in the Gangdese porphyry copper belt of southern Tibet: Implication for ore control. Ore Geol. Rev. 139, 104474.

Liang, Q., Chen, G., Luo, L., Huang, X., Hu, H., 2025. Appraising the porphyry Cu fertility using apatite trace elements: a machine learning method. J. Geochem. Explor. 270, 107664.

Loader, M.A., 2017. Mineral indicators of porphyry Cu fertility. Ph.D. thesis, Imperial College London, London, 436 p.

Loucks, R.R., 2014. Distinctive composition of copper-ore-forming arc magmas. Aust. J. Earth Sci. 61, 5–16.

Loucks, R.R., Henríquez, G.J., Fiorentini, M.L., 2024. Zircon and whole-rock trace element indicators of magmatic hydration state and oxidation state discriminate copper ore-forming from barren arc magmas. Econ. Geol. 119, 511–523.

Lu, Y.J., Loucks, R.R., Fiorentini, M.L., McCuaig, T.C., Evans, N.J., Yang, Z.M., et al., 2016. Zircon compositions as a pathfinder for porphyry Cu ± Mo ± Au deposits. Soc. Econ. Geol. Spec. Publ. 19, 329–347.

Mao, M., Rukhlov, A.S., Rowins, S.M., Spence, J., Coogan, L.A., 2016. Apatite trace element compositions: a robust new tool for mineral exploration. Econ. Geol. 111, 1187–1222.

Meinert, L.A., Dipple, G.M., Nicolescu, S., 2005. World skarn deposits. Econ. Geol. 100, 299–336.

Miles, A., Graham, C., Hawkesworth, C., Gillespie, M., Hinton, R., Bromiley, G., et al., 2016. Reply to comment by Marks, on "Apatite: a new redox proxy for silicic magmas?". Geochim. Cosmochim. Acta 183, 271–273.

Miles, A.J., Graham, C.M., Hawkesworth, C.J., Gillespie, M.R., Hinton, R.W., Bromiley, G.D., Emmac,, 2014. Apatite: a new redox proxy for silicic magmas? Geochim. Cosmochim. Acta 132, 101–119.

Moghadam, H.S., Li, Q.L., Griffin, W.L., Stern, R.J., Santos, J.F., Ducea, M.N., O'Reilly, S. Y., 2022. Temporal changes in subduction-to-collision-related magmatism in the Neotethyan orogen: the Southeast Iran example. Earth Sci. Rev. 226, 103930.

Moghadam, H.S., Xiao, W., Karsli, O., Aydin, F., He, Y., Chiaradia, M., Griffin, W.L., 2025. Copper endowment of the magmatic rocks from Eastern Sakarya Zone (Türkiye): insights from zircon and apatite geochemical evolution. Contrib. Mineral. Petrol. 180 (6), 38.

Nathwani, C.L., Blundy, J., Large, S.J., Wilkinson, J.J., Buret, Y., Loader, M.A., Tavazzani, L., Chelle-Michou, C., 2024. A zircon case for super-wet arc magmas. Nat. Commun. 15, 8982.

Nathwani, C.L., Wilkinson, J.J., Fry, G., Armstrong, R.N., Smith, D.J., Ihlenfeld, C., 2022. Machine learning for geochemical exploration: Classifying metallogenic fertility in arc magmas and insights into porphyry copper deposit formation. Mineral. Deposita 57, 1143–1166.

Pan, Y.M., Fleet, M.E., 2002. Compositions of the apatite-group minerals: Substitution mechanisms and controlling factors. Rev. Mineral. Geochem. 48, 13–49.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.
Peng, G.Y., Luhr, J.F., McGee, J.J., 1997. Factors controlling sulfur concentrations in volcanic apatite. Am. Mineral. 82, 1210–1224.
Petrelli, M., Perugini, D., 2016. Solving petrological problems through machine learning: the study case of tectonic discrimination using geochemical and isotopic data. Contrib. Miner. Petrol. 171, 81.
Piccoli, P.M., Candela, P.A., 2002. Apatite in igneous systems. Rev. Mineral. Geochem. 48, 255–292.
Prowatke, S., Klemme, S., 2006. Trace element partitioning between apatite and silicate melts. Geochim. Cosmochim. Acta 70, 4513–4527.
Qin, B., Huang, F., Huang, S., Python, A., Chen, Y., ZhangZhou, J., 2022. Machine learning investigation of clinopyroxene compositions to evaluate and predict mantle metasomatism worldwide. J. Geophys. Res. Solid Earth 127, e2021JB023614.
Qiu, K.F., Zhou, T., Chew, D., Hou, Z.L., Müller, A., Yu, H.C., Lee, R.G., Chen, H., Deng, J., 2024. Apatite trace element composition as an indicator of ore deposit types: a machine learning approach. Am. Mineral. 109, 303–314.
Qu, P., Yang, W., Niu, H., Li, N., Wu, D., 2022. Apatite fingerprints on the magmatic-hydrothermal evolution of the Daheishan giant porphyry Mo deposit, NE China. Geol. Soc. Am. Bull. 134, 1863–1876.
Richards, J.P., 2011. High Sr/Y arc magmas and porphyry Cu±Mo±Au deposits: just add water. Econ. Geol. 106, 1075–1081.
Richards, J.P., 2022. Porphyry copper deposit formation in arcs: what are the odds? Geosphere 18, 130–155.
Richards, J.P., Kerrich, R., 2007. Adakite-like rocks: their diverse origins and questionable role in metallogenesis. Econ. Geol. 102, 537–576.
Rukhlov, A.S., Plouffe, A., Ferbey, T., Mao, M., Spence, J., 2016. Application of trace-element compositions of detrital apatite to explore for porphyry deposits in central British Columbia. In: Geological Fieldwork 2015, British Columbia Ministry of Energy and Mines, British Columbia Geological Survey Paper 2016-1, 145–179.
Schodde, R., 2018. Where, what, when and who? Highlighting key global exploration opportunities, trends and a perspective on the cycle of mineral exploration. Conference Presentation, International Mining and Resource Conference (IMARC).
Schodde, R., 2019. Role of technology and innovation for identifying and growing economic resources. Conference presentation, AMIRA International's 12th Biennial Exploration Managers Conference, Hunter Valley, Australia, March 26–29, 2019.
Sha, L.K., Chappell, B.W., 1999. Apatite chemical composition, determined by electron microprobe and laser-ablation inductively coupled plasma mass spectrometry, as a probe into granite petrogenesis. Geochim. Cosmochim. Acta 63, 3861–3881.
Shu, Q., Chang, Z., Lai, Y., Hu, X., Wu, H., Zhang, Y., et al., 2019. Zircon trace elements and magma fertility: Insights from porphyry (-skarn) Mo deposits in NE China. Mineral. Deposita 54, 645–656.
Sillitoe, R.H., 2010. Porphyry copper systems. Econ. Geol. 105, 3–41.
Smythe, D.J., Brenan, J.M., 2015. Cerium oxidation state in silicate melts: combined fO2, temperature and compositional effects. Geochim. Cosmochim. Acta 170, 173–187.
Tang, M., Erdman, M., Eldridge, G., Lee, C.T.A., 2018. The redox "filter" beneath magmatic orogens and the formation of continental crust. Sci. Adv. 4, eaar4444.
Tang, M., Lee, C.T.A., Ji, W.Q., Wang, R., Costin, G., 2020. Crustal thickening and endogenic oxidation of magmatic sulfur. Sci. Adv. 6, eaba6342.
Wang, Y., Qiu, K.F., Müller, A., Hou, Z.L., Zhu, Z.H., Yu, H.C., 2021. Machine learning prediction of quartz forming-environments. J. Geophys. Res. Solid Earth 126, e2021JB021925.
Wang, L., Su, B.X., Chiaradia, M., Mao, Y.J., Qin, K.Z., Cao, M.J., 2025. The oxidation of arc magmas: A critical review. Earth Sci. Rev. 271, 105287.
Webster, J.D., 1990. Partitioning of F between H2O and CO2 fluids and topaz rhyolite melt: Implications for mineralizing magmatic-hydrothermal fluids in F-rich granitic systems. Contrib. Miner. Petrol. 104, 424–438.
Wilkinson, J.J., 2013. Triggers for the formation of porphyry ore deposits in magmatic arcs. Nat. Geosci. 6, 917–925.
Williams, S.A., Cesbron, F.P., 1977. Rutile and apatite: Useful prospecting guides for porphyry copper deposits. Mineral. Mag. 41, 288–292.
Xing, K., Shu, Q., Lentz, D.R., 2021. Constraints on the formation of the giant Daheishan porphyry Mo deposit (NE China) from whole-rock and accessory mineral geochemistry. J. Petrol. 62, egab018.
Xu, B., Hou, Z.Q., Griffin, W.L., Lu, Y., Belousova, E., Xu, J.F., O'Reilly, S.Y., 2021. Recycled volatiles determine fertility of porphyry deposits in collisional settings. Am. Mineral. 106, 656–661.
Yuan, S., Williams-Jones, A.E., Bodnar, R.J., Zhao, P., Zajacz, Z., Chou, I.-M., Mao, J., 2025. The role of magma differentiation in optimizing the fluid-assisted extraction of copper to generate large porphyry-type deposits. Sci. Adv. 11, eadr8464.
Zhang, J., Wang, R., Hong, J., 2022. Amphibole fractionation and its potential redox effect on arc crust: evidence from the Kohistan arc cumulates. Am. Mineral. 107, 1779–1788.
Zhang, P., Zhang, Z., Yang, J., Cheng, Q., 2023. Machine learning prediction of ore deposit genetic type using magnetite geochemistry. Nat. Resour. Res. 32, 99–116.
ZhangZhou, J., Li, Y., Chowdhury, P., Sen, S., Ghosh, U., Xu, Z., Liu, J., Wang, Z., Day, J. M.D., 2024. Predicting sulfide precipitation in magma oceans on Earth, Mars, and the Moon using machine learning. Geochim. Cosmochim. Acta 366, 237–249.
Zhao, Y., Zhang, Y., Geng, M., Jiang, J., Zou, X., 2019. Involvement of slab-derived fluid in the generation of Cenozoic basalts in Northeast China inferred from machine learning. Geophys. Res. Lett. 46, 5234–5242.
Zheng, Y.Y., Xu, B., Lentz, D.R., Yu, X.Y., Hou, Z.Q., Wang, T., 2024. Machine learning applied to apatite compositions for determining mineralization potential. Am. Mineral. 109, 1394–1405.
Zhong, S.H., Feng, C.Y., Seltmann, R., Li, D.X., Dai, Z.H., 2018. Geochemical contrasts between late Triassic ore-bearing and barren intrusions in the Weibao Cu–Pb–Zn deposit, East Kunlun Mountains, NW China: Constraints from accessory minerals (zircon and apatite). Mineral. Deposita 53, 855–870.
Zhou, R.J., Wen, G., Li, J.W., Jiang, S.Y., Hu, H., Deng, X.D., Zhao, X.F., Yan, D.R., Wei, K.T., Cai, H.A., Shang, S.C., Li, B.C., Dai, X.K., 2022a. Apatite chemistry as a petrogenetic–metallogenic indicator for skarn ore-related granitoids: an example from the Daye Fe–Cu–(Au–Mo–W) district, Eastern China. Contrib. Miner. Petrol. 177, 1–21.
Zhou, Y., Zhang, Z., Yang, J., Ge, Y., Cheng, Q., 2022b. Machine learning and singularity analysis reveal zircon fertility and magmatic intensity: Implications for porphyry copper potential. Nat. Resour. Res. 31, 1–18.
Zou, S.H., Chen, X.L., Brzozowski, M.J., Leng, C.B., Xu, D.R., 2022. Application of machine learning to characterizing magma fertility in porphyry Cu deposits. J. Geophys. Res. Solid Earth 127, e2022JB024584.