Performance of SVM model

Prior to opting for the XGBoost method, we explored the utilization of the common-used machine learning technique, Support Vector Machine (SVM), which yielded unsatisfactory results. The SVM model, trained on the "Major & Trace" subset, employed a radial basis function (RBF) as the kernel function. Its workflow mirrored that of the XGBoost model presented in the main text (Figure 3). Similarly, a 5-fold cross-validation strategy and grid search techniques were employed for parameter optimization and overfitting avoidance. The optimal SVM model achieved an accuracy of 0.690, an F1 score of 0.781, and an AUC value of 0.629 on the test set. It is noteworthy that a significant portion of "Unmineralized" data were erroneously classified as "Mineralized," as evident from the confusion matrix (Figure S1). In comparison to the performance of the XGBoost model M-T-1 (Accuracy = 0.992, F1 score = 0.993, AUC = 0.991), the disparity is substantial.

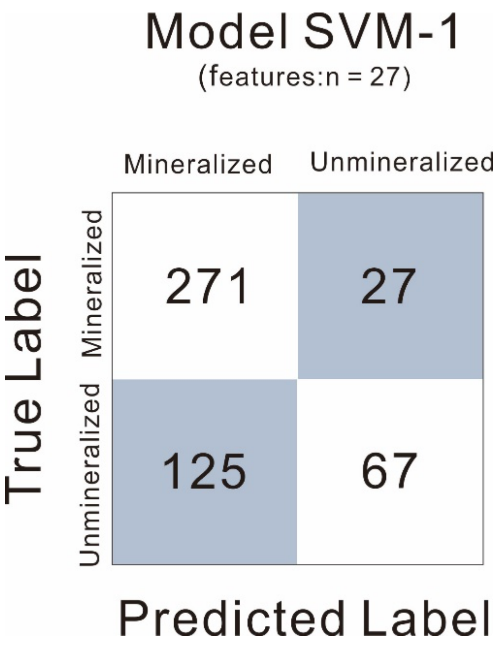

Figure S1. The confusion matrix of model SVM-1.

XGBoost model for deposit type classification

In order to provide a more comprehensive response to the query raised by Reviewer 1 regarding the compositional disparities of apatites from different deposit types, we have preliminarily trained a classification model to differentiate between apatite from various deposit types. Initially, we employed the "Major & Trace" subset, where each data point was assigned a corresponding deposit type label, encompassing the following categories: "Unmineralized," "Porphyry," "IOA," "Skarn," "Carbonatite," "IOCG," "Orogenic Au," "Orogenic Ni-Cu," and "Epithermal Au-Ag." It is noteworthy that due to limited data availability, apatite from epithermal Au-Ag deposit were excluded from this analysis.

Following the standardized workflow outlined in Figure 3, we identified the optimal model, denoted as M-T-DC-1, and applied it to the test set. The result is shown by the confusion matrix (Figure S2). revealing an impressive accuracy of 0.955 and an F1 score of 0.913. These metrics underscore the commendable performance of this model in distinguishing between different deposit types of apatites. Furthermore, an analysis of the relative importance rank yielded a list of the top ten influential features, namely Sm, Pr, CaO, Th, Zr, V/Y, Sr/Y, Rb, Cl, and F. It is noteworthy that Sm and Pr cannot be ruled out as potentially influenced by data accuracy; however, the remaining features bear a striking resemblance to the crucial determinants found in the mineralization discrimination model.

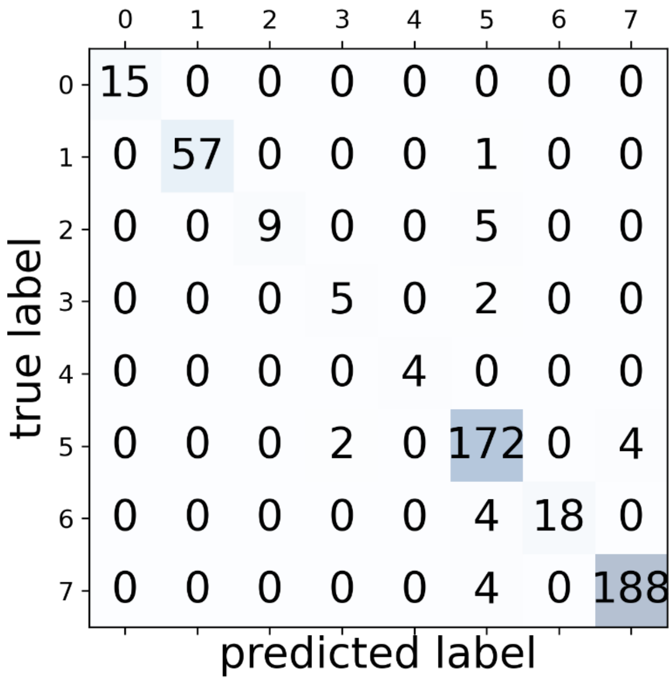

Figure S2. The confusion matrix of model M-T-DC-1; 0-Carbonatite; 1- IOA; 2-IOCG; 3- Orogenic Au; 4-Orogenic Ni-Cu; 5- Porphyry; 6-Skarn; 7-Unmineralized.

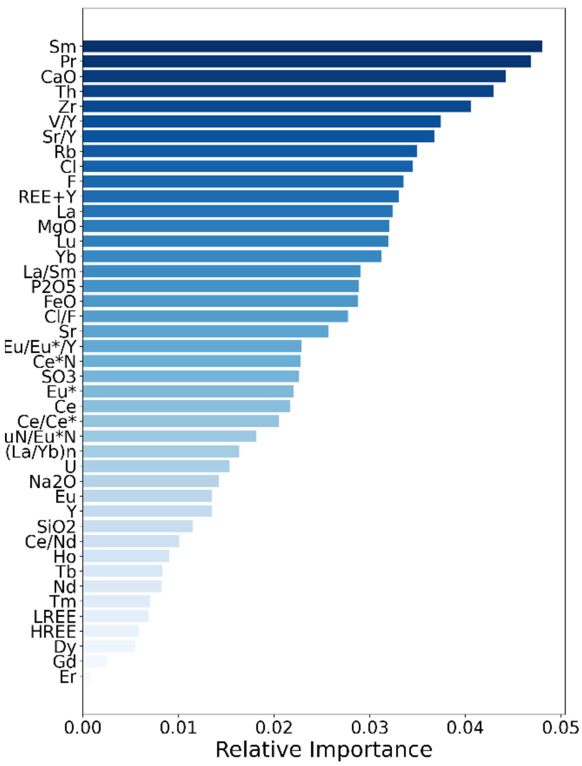

Figure S3. Relative Importance rank of model M-T-DC-1.

XGBoost model for identifying fertile magmatic and barren apatite

he primary objective of our study is to distinguish between fertile and barren apatite; therefore, we do not make a specific differentiation between fertile magmatic apatite and fertile hydrothermal apatite, although hydrothermal and magmatic apatites often exhibit significant geochemical differences. However, in response to the comments of Reviewer 2, in order to illustrate the differences of these two types of apatite in their contributions to determine mineralization potential, we excluded the data points of fertile hydrothermal apatite from the "Major & Trace" subset and retrained the model, denoted as M-T-MB-1. This model continued to demonstrate robust predictive performance (Accuracy = 0.978, F1-score = 0.976, Figure S4a). The feature importance rank (Figure S4b) revealed a striking similarity to model M-T-1, with Rb, Cl, Eu, Sr/Y, and Ce/Ce* appearing in the top 10 ranked features in both models, indicating the importance of these features for discriminating mineralization in both types of apatite. On the other hand, some features such as SO_3, Cl/F, V/Y, and certain REE elements displayed significant differences in their importance weights between the two models. This discrepancy may signify distinctions in the ability of magmatic and hydrothermal apatites to discriminate mineralization, aligning with observations by Bouzari et al. (2016) that altered apatites tend to exhibit depletion in S, Cl, F, and REE content.

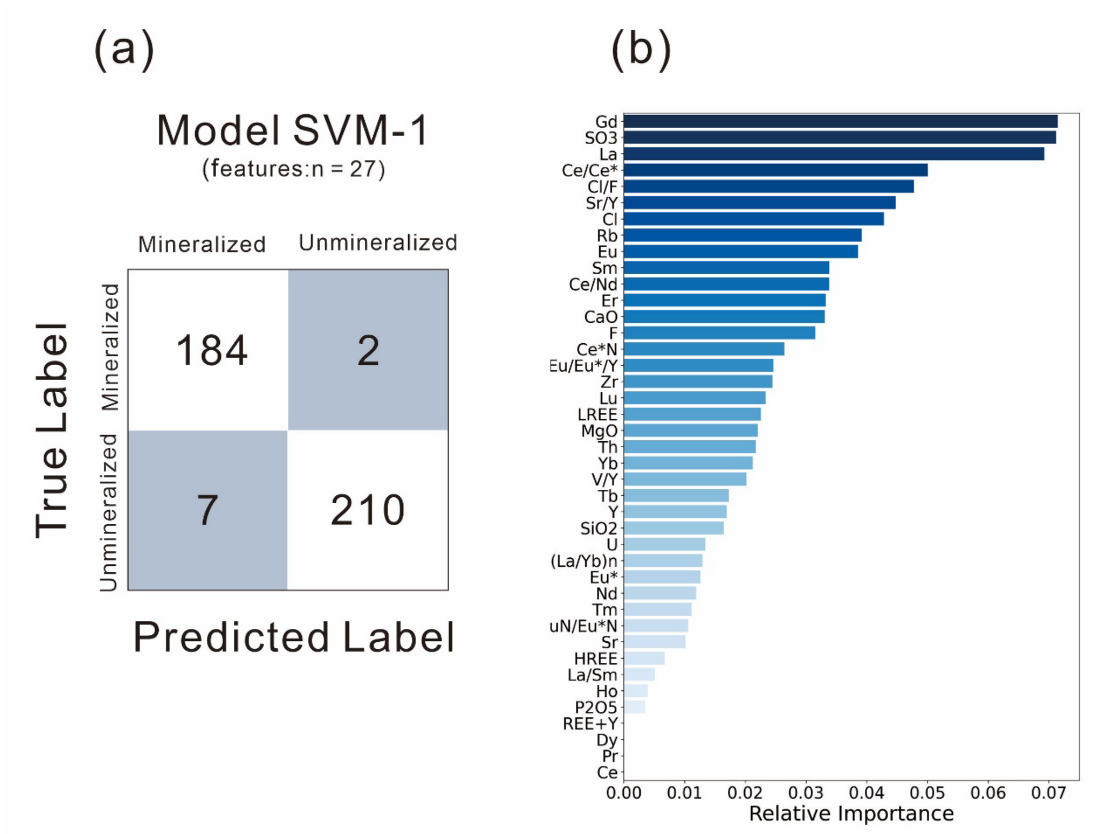

Figure S4. (a) The confusion matrix of model M-T-MB-1; (b) Relative Importance rank of model M-T-MB-1.