Revision 3

Geochemical discrimination of pyrite in diverse ore deposit types through statistical analysis and machine learning techniques

Hengxu Li [a], Zhaochong Zhang [a,*], Ruixuan Zhang [b], Qiuhong Xie [c], Lei Zhang [a], M. Santosh [a,d]

[a] State Key Laboratory of Geological Processes and Mineral Resources, China University of Geosciences, Beijing 100083, China

[b] School of Civil and Environmental Engineering, Queensland University of Technology, Brisbane, QLD, Australia

[c] Laboratory of Isotope Geology, Institute of Geology, Chinese Academy of Geological Sciences, Beijing 100037, China

[d] Department of Earth Sciences, University of Adelaide, Adelaide, SA, Australia

[*] Corresponding author: Zhaochong Zhang;

E-mail address: zczhang@cugb.edu.cn;

Telephone: 86-10-13910168892;

Fax: 86-10-82322195.

Abstract

Pyrite is a ubiquitous mineral in many ore deposits and sediments, and its trace element composition is mainly controlled by temperature, oxygen fugacity, pH, compositions of fluids and host rock composition. A Discriminant Analysis (DA), based on multi-element compositions of pyrite, distinguish iron oxide-apatite (IOA), iron oxide copper-gold (IOCG), skarn Cu-(Fe), porphyry Cu-Mo, orogenic Au, volcanic-hosted massive sulfide (VMS), sedimentary exhalative (SEDEX) deposits and barren sedimentary pyrite. Testing of the DA classifier yield an accuracy of 98% for IOA, 96% for IOCG, 91% for skarn Cu-(Fe), 94% for porphyry Cu-Mo, 87% for orogenic Au, 84 % for VMS, 96% for SEDEX and 85% for barren sedimentary pyrite. Furthermore, Neural Network, Support Vector Machine and Random Forest, were performed for selecting the optimum classifier more accurately. In these three techniques, the Support Vector Machine yield the highest overall accuracy (98% for IOA, IOCG, skarn Cu-Fe and porphyry Cu-Mo, and 97% for orogenic Au, VMS, SEDEX and barren sedimentary pyrite), and thus is an appropriate technique in predicting pyrite types.

Keywords: Pyrite; trace elements; discrimination diagrams; machine learning

1. Introduction

Metallogenetic models for different ore deposit types exert a major influence on ore exploration especially when the surface geology or geochemistry fails to reveal details about the deposit at depth. For example, minor disseminated pyrite from a sericite alteration zone in a drill hole could be related to a porphyry Cu-Mo deposit, a VMS system, an epithermal Au zone, or barren pyrite unrelated to an ore system

(Revan et al. 2014; Gregory et al. 2019; Chugaev et al. 2022 and references therein). Each of the mineralization type demands a different approach to exploration. Thus, a reliable method of distinguishing the type of ore deposits enhances the efficiency of mineral exploration. For newly discovered deposits with unclear deposit geology, the prediction of ore deposit type is helpful in understanding the process of the mineralization.

The global tectonics controls the formation of various ore deposits is the basis for their classification (e.g., Groves et al. 2005; Lydon 2007; Santosh and Groves 2022). Most of the deposit types considered in this study formed at destructive plate margins and are related to magmatic and/or associated hydrothermal systems. These deposits include porphyry Cu-Mo, skarn Fe-(Cu), orogenic Au, and IOA systems. Mineral deposits that form in constructive plate margins include some styles of IOCG systems. After classification based on the tectonic environment, the major mineral deposit types are further characterized by ore mineralogy, alteration, and host-rock associations (e.g., Hedenquist et al. 2000; Goodfellow 2007).

The IOCG comprise a diverse group of deposits viewed as iron oxide-associated deposits (Groves et al. 2010). IOA are sometimes classified as the Cu-poor end member of the IOCG system, although their genetic association remains controversial (Knipping et al. 2015a, b). IOCG deposits have abundant low-Ti iron oxides and have a close temporal relationship with the related intrusions (Groves et al. 2010). IOCG deposits have Cu ± Au as economic metals, which are formed by magmatic-hydrothermal processes. In contrast to IOCG system, IOA deposits typically lack economic Cu ± Au, and are associated with calc-alkaline magmatism (Knipping et al. 2015a, b; Mao et al. 2016).

Skarn Cu-(Fe) deposits are characterized by pervasive calc-silicate alteration (typically garnet and pyroxene) through magmatic-hydrothermal fluids at the margins of felsic intrusions (e.g., Einaudi et al. 1981; Meinert et al. 2005). Skarn deposits are commonly polymetallic with a wide range of grades and tonnages. Among the seven major skarn ore types (Fe, Au, Cu, Pb-Zn, W, Mo, and Sn), many types are parts of larger porphyry systems (Meinert et al. 2005; Ray 2013). Porphyry Cu-Mo deposits are large magmatic-hydrothermal deposits associated with intermediate to felsic porphyritic intrusions (Seedorff et al. 2005; Sillitoe 2010). The deposits typically contain hundreds of millions of tons of ore with low grades (generally <1% Cu and <0.1% Mo). Porphyry Cu-Mo are derived from I-type granites (Dilles et al. 2014) that possess variable degree of alkalinity (e.g., Barr et al. 1976) and states of oxidation (Cao et al. 2014).

Orogenic Au deposits encompass all epigenetic, structurally hosted, gold vein systems in metamorphic terranes (Groves et al. 1998; Mao et al. 2016). The deposits normally contain between 20 and 40 million tons of ore with average Au grade of 7.6 g/t (Dubé and Gosselin 2007). Most gold orebodies form at crustal depths (5–10 km), although deeper (~20 km) and shallower (~5 km) deposits are recognized (Groves 1993).

VMS and SEDEX deposits are significant sources of base metals such as copper, zinc, and lead, as well as precious metals like gold and silver (Galley et al. 2007).

VMS deposits are formed as a result of submarine volcanic activity, which occur in clusters or following the tectonic plate boundaries (Galley et al. 2007; Piercey 2009). VMS deposits are typically hosted in volcanic or volcaniclastic rocks, including basalt, andesite, rhyolite, and volcanic breccias (Praveen et al. 2020). SEDEX deposits are typically associated with extensional tectonic settings, such as rift basins, foreland basins, or passive margins (Large et al. 2005). They are typically hosted in fine-grained clastic sedimentary rocks, such as shale, siltstone, and mudstone and associated with the interaction of hydrothermal fluids in marine basins (Cooke et al. 2000; Chen et al. 2003; Large et al. 2005).

Pyrite is a ubiquitous mineral in various rocks and an essential constituent in many ore deposits (e.g., Huston et al. 1995b; Large et al. 2009; Reich et al. 2013, 2016; Gregory et al. 2015; Tardani et al. 2017). In spite of its relatively chemical composition, pyrite contains trace elements as solid solution or micro-particles (Reich et al. 2005, 2013; Deditius et al. 2009, 2011). The trace element composition of pyrite is mainly controlled by temperature, oxygen fugacity, pH and compositions of fluids and wall rocks, as well as the metal sources and depositional mechanisms (Huston et al. 1995a; Abraitis et al. 2004; Tardani et al. 2017). Therefore, pyrite trace elements discriminate different types of ore deposits (e.g., Huston et al. 1995b; Large et al. 2009; Reich et al. 2013, 2016; Gregory et al. 2015; Tardani et al. 2017). Binary element scatter plots of pyrite chemistry have been used to distinguish pyrite from different ore deposits (Loftus-Hills and Solomon 1967; Bajwah et al. 1987; Bralia et al. 1979). However, these diagrams have significant compositional overlaps.

Therefore, multi-element discrimination diagrams involving several different elements may help in discriminating the ore deposit type.

In this paper, we process a total of 3287 trace element spot measurement of pyrite from IOA (iron oxide-apatite), IOCG (iron oxide copper-gold), skarn Cu-(Fe), porphyry Cu-Mo, orogenic Au, VMS (volcanic-hosted massive sulfide), SEDEX (sedimentary exhalative) deposits and barren sedimentary pyrite with Discriminant analysis (DA), and build discrimination diagrams to identify and visualize the distinctions among different types of deposits. Despite the pyrite composition is affected by various factors, data from different analysis spots could record their complicated crystallization environment. Therefore, the discrimination model constructed investigate the various factors that controlling the pyrite composition, and distinguish different ore deposit types. Furthermore, we discuss the performance of three machine learning algorithms, i.e., Artificial Neural Network, Support Vector Machine and Random Forest, for selecting the optimum classifier to distinguish different types of deposits more accurately.

2. Methods

2.1 Data collection

The data set was collected from publications covering different mineralization systems (Fig. 1; Table 1). Most of the pyrite data are from laser ablation-inductively coupled plasma mass spectrometry (LA-ICP-MS) although the interlaboratory variations in analytical methods, standards, and instruments cause varying detection

DOI: https://doi.org/10.2138/am-2023-8976. http://www.minsocam.org/

limits obtained from different laboratories (Supp Table 1). Data on pyrite from the Los Colorados IOA deposit is from secondary ion mass spectrometry (SIMS) and electron probe microanalysis (EPMA).

2.2 Pretreatment of data

Concentrations of 23 elements in the pyrite were collected. Since some elements are below the detection limits or not reported in some articles, we mainly focus on the residual 12 elements, including Co, Ni, Cu, Zn, As, Se, Ag, Sb, Te, Au, Pb and Bi. The geochemical data of pyrite from different deposit types is shown in Supp Table 2. The residual 12 elements are sometimes absent in the reported data of pyrite and hardly contribute to indicator characterization. Previous researchers provide multiple approaches and offer various replacement options to deal with such values (e.g., Filzmoser et al. 2009; Hron et al. 2010; Grunsky et al. 2013). In order to avoid the potential effects of outliers and respect the compositional nature of the database, the IRMI (iterative model-based imputation) algorithm developed by Templ et al. (2011) was used to impute the null data. Initial guesses of the missing values were taken to be equal to the median of the corresponding data column, and random noise was added to the final best estimate of each value to preserve randomness. Data for each deposit type was treated separately. To deal with data below the detection limit, we use the random imputation of the raw values with a normal distribution with both mean and standard deviation equal to the detection limit (van den Boogaart and Tolosana-Delgado 2013; Frenzel et al. 2016). Negative values resulting from this imputation procedure may be adjusted simply by changing their sign. Te and Au were not analyzed (or reported) in all of the SEDEX deposit. Therefore, medians reported by Gregory et al. (2019) for SEDEX deposit were used for all the SEDEX analyses. This has probably overestimated the ability of the classifier to identify SEDEX analyses, because the same value for Te and Au was used by all the SEDEX samples. However, Gregory et al. (2019) also used the medians to deal with Te content which is not analyzed. They suggest that SEDEX pyrite has distinctly higher Cu, Sb and Pb concentrations compared to most other deposits, so it is thought that Te is not particularly important for SEDEX classification. To minimize the influence of micro-inclusions on trace elements of pyrite, data with higher than 1% Zn, 1% Cu, 1% Ni, 1% Pb, 2% As, and 2% Co were screened.

Finally, a total of 3287 analyses from 7 ore deposit types were used as follows: IOA (126 analyses, 18 samples from 2 localities), IOCG (279 analyses, 21 samples from 4 localities), skarn Cu-(Fe) (203 analyses, 19 samples from 3 localities), porphyry Cu-Mo (419 analyses, 32 samples from 4 localities), orogenic Au (386 analyses, 36 samples from 10 localities), VMS (353 analyses, 46 samples from 5 localities), SEDEX (693 analyses, 43 samples from 2 localities) deposits and barren sedimentary pyrite (828 analyses, 55 samples from 13 localities) (Table 1; Supp Table 3).

2.3 Classification methods

2.3.1 Discriminant Analysis

To find the optimum criteria for discriminating deposits using pyrite trace element compositions, we employed the Discriminant Analysis (DA) in SPSS®

software. DA is a multivariate statistical technique that projects multivariate data into a lower dimensional space to achieve the best group separation (Flury 1997; Makvandi et al. 2016; Chen et al. 2019). The Discriminant Analysis calculates a set of linear discriminant functions that are combinations of the original variables (i.e., element concentrations) that maximize the differences between the predefined groups, which allow the samples to be plotted in the discriminant space so that group separation can be visualized and investigated. Before data processing, the data set was logarithmic transformed, by simply taking the logarithm of the variables. Although the logarithmic transformed vectors are still constrained in a sub-space (Filzmoser et al. 2018; Buccianti and Grunsky 2014), no variable will be sacrificed during the transformation and the original geochemical sense of the variables could be possessed (Wang et al. 2014). Many researchers have also used logarithmic transformation to normalized the data set (Mao et al. 2016; Hu et al. 2022).

2.3.2 Machine learning algorithm

The multilayer perceptron (MLP): MLP is used in this study and therefore the term “artificial neural network (ANN)” here refers to MLP. An MLP consists of several layers, an input layer, an output layer and one or more hidden layers. The nodes in each layer are called neurons, and the neurons between each layer are connected by adjustable weights and biases. Activate functions are used in hidden neurons for non-linear mapping. The MLP optimizes the objective problem by adjusting the weights and biases iteratively.

Support Vector Machine (SVM): Different from traditional methods that minimize the empirical training error, SVM attempts to find the optimum separating hyperplane by maximizing the margin between the hyperplane and the training data (Kuo et al. 2013). For nonlinear classification task, SVM introduces the kernel functions for nonlinear mapping. SVM was initially designed for binary classification problem but can be further extended to multiclass classification task using "one-versus-all" or "one-versus-one" approach. In this study, the SVM classifier developed for pyrite type recognition uses "one-versus-one" approach. For K class classification task, the "one-versus-one" approach developed an SVM between any two classes, resulting in a total K(K-1)/2 number of SVM. Given an unlabeled sample, the class with the most votes is identified as the class of this sample.

Random forest (RF): RF is an ensemble machine learning algorithm introduced by Breiman (2001), which solve classification and regression problem. RF is a collection of decision trees. The decision tree is a hierarchical model constructed by recursively partitioning the feature space of a dataset into single class subspaces (Myles et al. 2004).

RF uses bootstrap sampling to construct n different decision trees based on n subsets generated from original dataset. Furthermore, the decision trees in RF only select a fixed number of features randomly to increase the difference between decision trees. The final decision of RF is the majority predictions of trees.

2.3.3 Machine-learning model development

Standardization of the data set is a common requirement for many machine learning estimators. To eliminate the spurious relationships between the compositions, the centered log-ratio transformation (clr) is used (Filzmoser et al. 2018; Buccianti and Grunsky 2014) to normalize the compositional data. The normalization method could be seen in Buccianti and Grunsky (2014) and Wang et al. (2014). And after the clr transformation, the dataset is normalized into [0,1] following the equation presented as: $x_i = \frac{x_i - x_{min}}{x_{max} - x_{min}}$ before the machine learning training process.

It is necessary to evaluate the optimum hyperparameters of the machine learning models, for which a 5-fold cross validation process was used with Bayesian optimization (BO) for hyperparameters optimization purpose, where each fold of validation dataset consisted of a randomly selected 20% of training dataset and the objective function of BO used the average overall error rate obtained on the 5 validation sets. Bayesian optimization technique is provided by MATLAB. The splitting process still followed the aforementioned standard. For developing RF model, the number of trees (ntree) and the number of features selected in each split (mtry) need to be determined. The optimizable hyperparameters of SVM includes the regularization coefficient (C) and the kernel size (σ). As for MLP, designing its topology is essential, and therefore the number of hidden layer (numL) and the number of hidden neurons in each layer (numH) were considered to be optimized. The searching spaces of these hyperparameters were presented in Table 3. Table 4 shows the BO optimized hyperparameters for each category.

In addition, the misclassification cost matrix of SVM classifier was also optimized to address the issue of imbalanced dataset. The misclassification cost matrix is given as:

DOI: https://doi.org/10.2138/am-2023-8976. http://www.minsocam.org/

$$\begin{bmatrix} 0 & c_{12} & c_{13} & c_{14} \\ c_{21} & 0 & c_{23} & c_{24} \\ c_{31} & c_{32} & 0 & c_{34} \\ c_{41} & c_{42} & c_{43} & 0 \end{bmatrix} \quad (1)$$

where c_{ij} represents the cost (or penalty) of misclassifying type i into type j (false negative of type i). Since the SVM classifier developed in this study uses "one versus one" approach, where 6 sub-SVMs ($6 = \frac{4*3}{2}$) were developed for each category. Therefore, following the method suggested by Savu-Krohn et al. (2011), the false negative penalty c_{ij} and false positive penalty c_{ji} obeys the following relationship:

$$\begin{cases} c_{ji} = 1 - c_{ij} \\ 0 < c_{ij} < 1 \end{cases} \quad (2)$$

As a result, the optimization of the misclassification cost matrix presented in equation (1) is identical to the optimization of 6 false negative penalties in the upper triangle of the matrix, $\{c_{12}, c_{13}, c_{14}, c_{23}, c_{24}, c_{34}\}$. Considering the distribution of datasets, the searching range of these penalty values were presented in Supp Table 4. Since we attempt to assign those classes with less training samples with larger weights, the cost of misclassifying a class with less training samples into a class with more samples were designed to be at least 0.5. As for two classes with similar sample sizes, the misclassification cost between them should be around 0.5, and therefore was searched within [0.4,0.6].

Furthermore, during the penalty optimization process, the previously obtained hyperparameters were used and remained fixed. The BO optimized penalty values for category A and B were presented in Table 5 and Table 6.

After the optimization processes, all of the classifiers were trained again on the whole training dataset (including the validation dataset) using the optimized hyperparameters and subsequently evaluated on testing dataset. All the models were developed and tested on MATLAB.

3. Results

Co and Ni concentrations for pyrite from IOA and IOCG deposits are considerably higher than those in other metallogenic systems and barren sedimentary rocks (Fig. 2; Table 2). Large variations of Zn and Pb concentrations exist between the different pyrite groups (Fig. 2). Pyrite from the IOA deposit and SEDEX deposit have the highest Pb contents, whereas pyrite from porphyry Cu-Mo deposit has the lowest Pb contents. Pyrites from porphyry Cu-Mo deposits have the lowest As contents.

Pyrites in IOA contain Se similar to that of IOCG, porphyry Cu-Mo, orogenic Au, SEDEX, VMS deposit and sedimentary rocks, whereas pyrite from the skarn Fe-Cu deposit has the lowest Se contents. Sb contents of the pyrite for the porphyry Cu-Mo deposits has the lowest Sb concentrations. Pyrite is generally poor in Te, although pyrites from the VMS deposits contain Te contents of up to 2925 ppm. All other pyrites contain Te < 278.5 ppm. The highest values of gold are observed in pyrites from IOCG and orogenic Au deposits while Au contents in pyrite from other deposits are low or below the detection limit. Bi content in pyrite varies from sub-detection limit to 2077 ppm, with the highest values in pyrite from VMS deposits.

To sum up, Co and Ni in pyrite from IOA and IOCG deposits are almost an order of magnitude higher than those in pyrite from the other deposits (Fig. 2). VMS pyrite is more enriched in silver, Se, Te and Bi while porphyry Cu-Mo deposits display distinctly lower Arsenic, Pb and Sb values. Barren sedimentary pyrite is commonly enriched in Cu, Zn and Pb compared with ore-related pyrite. Hence, trace elements of pyrite have the potential to discriminate the various deposit types.

4. Discussion

4.1 Co-Ni discrimination diagrams

Co/Ni ratio >1 in pyrite has been attributed to high-temperature, magmatic-hydrothermal systems (e.g., Bralia et al. 1979; Bajwah et al. 1987; Koglin et al. 2010). In contrast, a low Co/Ni ratio <1 indicates felsic magmatic or a sedimentary origin for pyrite (Loftus-Hills and Solomon 1967; Koglin et al. 2010). Previous studies have defined overlapping compositional fields for pyrite from different deposits based on their Co/Ni ratios (Fig. 3; Reich et al. 2016), indicating that Co/Ni ratio is not an effective tool to discriminate deposit types.

4.2 Discrimination analysis of pyrite through DA

The ore deposit types and barren sedimentary rocks are first divided into two categories according to their geological environments and each category consists of 4 types of pyrites.

4.2.1 Magmatic-hydrothermal deposits (category A)

The IOA, IOCG, skarn Cu-(Fe) and porphyry Cu-Mo deposit are associated with volcanic and intrusive rocks. The skarn and porphyry deposits have similar ore elements, mineral assemblages, alteration zones and fluid sources (Meinert et al. 2005 and references therein; Reich et al. 2013; Keith et al. 2022). IOA and IOCG deposits sometimes occur in the same metallogenic belt, such as in the case of the Chilean iron belt. Although the genesis of IOA deposits, whether they are magmatic or hydrothermal in origin is a long-standing controversy, the magmatic-hydrothermal fluid is considered to play an important role in the mineralization process (Knipping et al. 2015a, 2015b). The close spatial association among IOA, IOCG and iron skarn ores sometimes leads to misjudgments of these ore types (Allen et al. 1996; Mao et al. 2011; Nold et al. 2014; Bilenker et al. 2016; Harlov et al. 2016; Jonsson et al. 2016; Hu et al. 2020 and references therein).

The discrimination diagram (Fig. 4; Supp Table 5) based on discriminant functions effectively separate the IOA and IOCG deposits from the skarn Cu-(Fe) and porphyry Cu-Mo deposits. There is overlap between skarn Cu-(Fe) and porphyry Cu-Mo deposit (Fig. 4). Sillitoe (2010) considered that the skarn Cu-(Fe) and porphyry Cu-Mo deposits could form through the evolution of a single magmatic-hydrothermal system. The overlap in the discrimination diagram also demonstrates the similarities in their fluid and metal source.

4.2.2 Orogenic Au, VMS, SEDEX deposits and barren sedimentary rocks (category B)

SEDEX and barren sedimentary pyrite form in the sedimentary environment and share similarities of trace element characteristics (Fig. 2). There is a controversy on the genesis of orogenic Au deposit related to the fluid and metal source (Bath et al. 2013; Lawrence et al. 2013; Goldfarb and Groves 2015; Spence-Jones et al. 2018). Therefore, distinguishing pyrite from the orogenic Au, VMS, SEDEX deposits and barren sedimentary rocks is not only important for the construction of metallogenetic model, but may also be helpful in mineral exploration.

As shown in Fig. 4, discriminant functions separate VMS and orogenic Au deposits from the SEDEX deposit and barren sedimentary rocks. However, the VMS shows overlap with orogenic Au deposit. It should be noted that VMS deposits could be divided into different subtypes (Urals-, Kuroko-, Cyprus-, Besshi-), which formed in different environment, but only Cyprus- and Urals-type deposits represent the VMS deposit.

4.2.3 Evaluation of the discrimination analysis

To evaluate the Discriminant Analysis, a stratified sampling construct the training data set by sampling four-fifths of the data from each deposit type, and the remaining one-fifth data of the same ratios of deposit types acted as a test data set (Supp Table 3). The training data set was used to build the classifier, and the test data set was used to evaluation.

Using the classification functions (Supp Table 6), most of the literature data are classified with high accuracy, see Supp Table 7. Apart from the test dataset, five replicated sampling 20% from the whole data set was used to evaluation. The sampling is done randomly. The literature data are correctly classified with high accuracy (98% for IOA, 96% for IOCG, 91% for skarn Cu-(Fe), 94% for porphyry Cu-Mo, 87% for orogenic Au, 84 % for VMS, 96% for SEDEX and 85% for barren sedimentary pyrite).

The DA builds discrimination diagrams used to visualize the distinctions among different types of deposits evaluate the contribution of each variable to identify the type of ore deposits. However, the relatively low accuracy for category B warrants more accurate methods to predict ore deposit types.

4.3 Discrimination of pyrite based on machine learning

Prior to the development of classification model, both category A and B were randomly split into training and testing datasets with a proportion of 80% and 20%, that is, 818 and 209 sets of data were used for training and testing for category A and 1806 and 454 sets of data were used for the same purpose for category B. To ensure the fairness, each subtype of pyrite is split according to the aforementioned ratio. For instance, for the total 126 IOA samples, 99 samples (nearly 80%) were used as training purpose, and the remaining 25 samples were split into testing dataset. The splitting of training and test dataset is done by random, respecting the hierarchical data structure. The number of the training and test data from each deposit could be seen in Supp Table 3.

4.3.1 Evaluation of models

The performance of the model was evaluated using confusion matrix, ROC curve and accuracy. Accuracy of the model is obtained by calculating:

$$Accuracy = \frac{No.of\ correct\ predictions}{No.of\ total\ samples}$$

The confusion matrix shows the predicted and actual classification. The ROC curve shows the tradeoff between the true positive rate (TPR) and false positive rate (FPR) of a classification model, given by:

$$TPR = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$

where TP refers to No. of true positive samples, FN denotes the No. of false negative samples, FP represents false positives and TN means true negatives.

4.3.2 Results and Discussion

The classification accuracies of the three machine learning models were summarized in Table 7. The ROC curves and confusion matrix on each general category were presented in Fig. 5 and Fig. 6.

The confusion matrix presented in Fig. 5 illustrated that SVM can perfectly recognize the IOA and porphyry Cu-Mo type of pyrite. All IOA and porphyry Cu-Mo testing samples were correctly identified with no false positive, while the identified skarn Cu-(Fe) and IOCG samples both had 2 false positive. Furthermore, according to the results presented in Table 7, although all these three models achieved relatively high classification accuracy, SVM still has the strongest classification ability by achieving an overall accuracy of 98%. This can also be proved from the ROC curve given in Fig. 6, where SVM has the largest area under curve (AUC).

The pyrites in general category B are a little more difficult to be distinguished. According to the confusion matrix presented in Fig. 5, the VMS type of pyrite was the most difficult to be identified for all the three models, since the obtained classification accuracies of which were relatively low compared to other types. RF demonstrated a perfect classification performance on SEDEX pyrite, with no false negative or false positive. However, the ROC curves in Fig. 6 demonstrated that SVM is the most suitable model in estimating pyrite type. The summarized results presented in Table 7 also illustrates the superiority of SVM in estimating the pyrite types in category B, which ranked the 1st among all models with an overall accuracy of 97%, although all these three models achieved relatively high classification accuracy.

As a result, the machine learning classifier, especially SVM is useful for identifying ore deposit type. The higher overall accuracy of SVM illustrates the superiority of machine learning models in estimating the pyrite types compared to Discriminant analysis. However, Discriminant analysis provides clear and easily interpretable results, which typically requires less computational resources and training time compared to RF, especially when dealing with large datasets.

5. Implications

This study adopted data on pyrite from IOA, IOCG, skarn Cu-(Fe), porphyry Cu-Mo, orogenic Au, VMS, SEDEX deposits and barren sedimentary pyrite, and through the application of statistical technique, we constructed discrimination diagrams from Discriminant analysis. The calculated discriminant functions highlight distinct multi-element differences among pyrite from various types of deposit. Furthermore, three machine learning algorithms, i.e., Artificial Neural Network, Support Vector Machine and Random Forest, are performed for the purpose of selecting the optimum classifier to distinguish different types of deposits more accurately. The accuracy demonstrates that pyrite trace element data combined with Support Vector Machine is a useful tool to discriminate ore types.

A weakness of the four current discriminator is the variability in the number of deposits and the amount of data from each site. IOA and SEDEX deposit are based on samples from two localities. This may suggest that pyrite trace element concentrations for IOA and SEDEX deposit are not fully representative of the ranges for the ore systems. Tl, Sn and W may be useful element in discriminating the deposit types. These elements were not included in the classifier because of a general lack of data in most data sources.

Acknowledgments

This study was supported by National Key Research and Development Program (2022YFC2903705). We appreciate two anonymous reviewers and associate editor Dr. Andreas Ertl for their thoughtful and constructive comments.

References

Abraitis, P.K., Pattrick, R.A.D., and Vaughan, D.J. (2004) Variations in the compositional, textural and electrical properties of natural pyrite: A review. International Journal of Mineral Processing, 74, 41–59.

Allen, R.L., Lundstrom, I., Ripa, M., and Christofferson, H. (1996) Facies analysis of a 1.9 Ga, continental margin, back-arc, felsic caldera province with diverse Zn-Pb-Ag-(Cu-Au) sulfide and Fe oxide deposits, Bergslagen region, Sweden. Economic Geology, 91, 979–1008.

Augustin, J., and Gaboury, D. (2019) Multi-stage and multi-sourced fluid and gold in the formation of orogenic gold deposits in the world-class Mana district of

Burkina Faso–Revealed by LA-ICP-MS analysis of pyrites and arsenopyrites. Ore Geology Reviews, 104, 495–521.

Bajwah, Z.U., Seccombe, P.K., and Offler, R. (1987) Trace element distribution, Co: Ni ratios and genesis of the Big Cadia iron-copper deposit, New South Wales, Australia. Mineralium Deposita, 22, 292–300.

Barr, D.A., Fox, P.E., Northcote, K.E., and Preto, V.A. (1976) The alkaline suite porphyry deposits: A summary, in Sutherland Brown, A.E., ed., Porphyry deposits of the Canadian Cordillera. Canadian Institute of Mining and Metallurgy Special Volume 15, 359–367.

Bath, A.B., Walshe, J.L., Cloutier, J., Verrall, M., Cleverley, J.S., Pownceby, M.I., Macrae, C.M., Wilson, N.C., Tunjic, J., Nortje, G.S., Robinson, P. (2013) Biotite and Apatite as Tools for Tracking Pathways of Oxidized Fluids in the Archean East Repulse Gold Deposit, Australia. Economic Geology, 108, 667–690.

Basori, M.B.I., Gilbert, S., Large, R.R., and Zaw, K. (2018) Textures and trace element composition of pyrite from the Bukit Botol volcanic-hosted massive sulphide deposit, Peninsular Malaysia. Journal of Asian Earth Sciences, 158, 173–185.

Belousov, I., Large, R., Meffre, S., Danyushevsky, L., Steadman, J., and Beardsmore, T. (2016) Pyrite compositions from VHMS and orogenic Au deposits in the Yilgarn craton, Western Australia: Implications for gold and copper exploration. Ore Geology Reviews, 79, 474–499.

Bilenker, L.D., Simon, A.C., Reich, M., Lundstrom, C.C., Gajos, N., Bindeman, I., Barra, F., and Munizaga, R. (2016) Fe-O stable isotope pairs elucidate a high-temperature origin of Chilean iron oxide-apatite deposits. Geochimica et Cosmochimica Acta, 177, 94–104.

Buccianti, A., and Grunsky, E. (2014) Compositional data analysis in geochemistry: Are we sure to see what really occurs during natural processes? Journal of Geochemical Exploration, 141, 1–5.

Bosori, M.B.I., Gilbert, S., Large, R.R., and Zaw, K. (2018) Textures and trace element composition of pyrite from the Bukit Botol volcanic-hosted massive sulphide deposit, Peninsular Malaysia. Journal of Asian Earth Sciences, 158, 173–185.

Bralia, A., Sabatini, G., and Troja, F. (1979) A revaluation of the Co/Ni ratio in pyrite as geochemical tool in ore genesis problems. Mineralium Deposita, 14, 353–374.

Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

Cao, M., Qin, K., Li, G., Jin, L., Evans, N., and Yang, X. (2014) Baogutu: An example of a reduced porphyry Cu deposit in western Junggar. Ore Geology Reviews, 56, 159–180.

Che, Y.D., (2014) A study on the characteristics of leucocratic alteration zone of Nihe iron deposit, Lujiang, Anhui. Hefei University of Technology. Master Thesis (in Chinese with English abstract).

Chen, J., Walter, M.R., Logan, G.A., Hinman, M.C., and Summons, R.E. (2003) The Paleoproterozoic McArthur River (HYC) Pb/Zn/Ag deposit of northern Australia: organic geochemistry and ore genesis. Earth and Planetary Science Letters, 210, 467–479.

Chen, W., Pradhan, B., Li, S.J., Shahabi, H., Rizeei, H.M., Hou, E.K., and Wang, S.Q. (2019) Novel Hybrid Integration Approach of Bagging-Based Fisher's Linear Discriminant Function for Groundwater Potential Analysis. Natural Resources Research, 28, 1239–1258.

Chugaev, A.V., Plotinskaya, O.Y., Dubinina, E.O., Stepanov, S.Y., Gareev, B.I., Batalin, G.A., Rassokhina, I.V., Chazhova, J.N., Bondar, D., and Abramova, V.D. (2022) Mixed crustal-mantle source of porphyry Cu-Mo deposits of the Urals: Pyrite trace element geochemistry and Pb-S isotope data. Journal of Geochemical Exploration, 242, 107075.

Cooke, D.R., Bull, S.W., Large, R.R., and McGoldrick, P.J. (2000) The importance of oxidized brines for the formation of Australian Proterozoic stratiform, sediment hosted Pb-Zn (SEDEX) deposits. Economic Geology, 95, 1–18.

Cooke, D.R., Wilkinson, J.J., Baker, M., Agnew, P., Phillips, J., Chang, Z.S., Chen, H.Y., Wilkinson, C.C., Inglis, S., Hollings, P., Zhang, L.J., Gemmell, J.B., White, N.C., Danyushevsky, L., and Martin, H. (2020) Using Mineral Chemistry to Aid Exploration: A Case Study from the Resolution Porphyry Cu-Mo Deposit, Arizona. Economic Geology, 115, 813–840.

Deditius, A.P., Kesler, S.E., Ewing, R.C., and Walshe, J. (2011) Trace metal nanoparticles in pyrite. Ore Geology Reviews, 42, 32–46.

Deditius, A.P., Utsunomiya, S., Ewig, R.C., Chryssoulis, S.L., Venter, D., and Kesler, S.E. (2009) Decoupled geochemical behavior of As and Cu in hydrothermal systems. Geology, 37, 707–710.

del Real, I., Thompson, J.F.H., Simon, A.C., and Reich, M. (2020) Geochemical and Isotopic Signature of Pyrite as a Proxy for Fluid Source and Evolution in the Candelaria-Punta del Cobre Iron Oxide Copper-Gold District, Chile. Economic Geology, 115, 1493–1517.

Dilles, J.H., Kent, J.R., Wooden, J.L., Tosdal, R.M., Koleszar, A., Lee, R.G., and Farmer, L.P. (2014) Zircon compositional evidence for sulfur-degassing from ore-forming arc magmas. Economic Geology, 110, 1–11.

Dubé, B., and Gosselin, P. (2007) Greenstone-hosted quartz-carbonate vein deposits, in Goodfellow, W.D., eds., Mineral deposits of Canada: A synthesis of major deposit types, district metallogeny, the evolution of geological provinces, and exploration methods. Geological Association of Canada, Mineral Deposits Division Special Publication, 5, 49–73.

Einaudi, M.T., Meinert, L.D., and Newberry, R.J. (1981) Skarn deposits: Economic Geology 75th Anniversary Volume, 317–391.

Filzmoser, P., Hron, K., Reimann, C. (2009) Principal component analysis for compositional data with outliers. Environmetrics, 20, 621–632.

Filzmoser, P., Hron, K., and Templ, M. (2018) Applied compositional data analysis. Springer.

Flury, B. (1997) A first course in multivariate statistics. 715 p. Springer Verlag, New York.

Frenzel, M., Hirsch, T., and Gutzmer, J. (2016) Gallium, germanium, indium, and other trace and minor elements in sphalerite as a function of deposit type—a meta-analysis. Ore Geology Reviews, 76, 52–78.

Gadd, M.G., Layton-Matthews, D., Peter, J.M., and Paradis, S.J. (2016) The world-class Howard's Pass SEDEX Zn-Pb district, Selwyn Basin, Yukon. Part I: trace element compositions of pyrite record input of hydrothermal, diagenetic, and metamorphic fluids to mineralization. Mineralium Deposita, 51, 319–342.

Galley, A.G., Hannington, M.D., and Jonasson, I.R. (2007) Volcanogenic massive sulffde deposits, in Goodfellow, W.D., ed., Mineral Deposits of Canada: A Synthesis of Major Deposit-Types, District Metallogeny, the Evolution of Geological Provinces, and Exploration Methods. Geological Association of Canada, Mineral Deposits Division, Special Publication, 5, 141–161.

Genna, D., and Gaboury, D. (2015) Deciphering the Hydrothermal Evolution of a VMS System by LA-ICP-MS Using Trace Elements in Pyrite: An Example from the Bracemac-McLeod Deposits, Abitibi, Canada, and Implications for Exploration. Economic Geology, 110, 2087–2108.

Goldfarb, R.J., and Groves, D.I. (2015) Orogenic gold: common or evolving fluid and metal sources through time. Lithos, 233, 2–26.

Goodfellow, W.D. (2007) Mineral deposits of Canada: A synthesis of major deposit types, district metallogeny, the evolution of geological provinces, and exploration methods. Geological Association of Canada, Mineral Deposits Division, Special Publication, 5, 1068 p.

Gregory, D.D., Cracknell, M.J., Large, R.R., McGoldrick, P., Kuhn, S., Maslennikov, V.V., Baker, M.J., Fox, N., Belousov, I., Figueroa, M.C., Steadman, J.A., Fabris, A.J., and Lyons, T.W. (2019) Distinguishing Ore Deposit Type and Barren Sedimentary Pyrite Using Laser Ablation-Inductively Coupled Plasma-Mass Spectrometry Trace Element Data and Statistical Analysis of Large Data Sets. Economic Geology, 114, 771–786.

Gregory, D.D., Large, R.R., Halpin, J.A., Baturina, E.L., Lyons, T.W., Wu, S., Danyushevsky, L., Sack, P.J., Chappaz, A., Maslennikov, V.V., and Bull, S.W. (2015) Trace element content of sedimentary pyrite in black shales. Economic Geology, 110, 1389–1410.

Groves, D.I., Condie, K.C., Goldfarb, R.J., Hronsky, J.M.A., and Veilreicher, R.M. (2005) Secular changes in global tectonic processes and their influence on the temporal distribution of gold-bearing mineral deposits. Economic Geology, 100, 203–224.

Groves, D.I. (1993) The crustal continuum model for late Archean lode-gold deposits of the Yilgarn Block, Western Australia. Mineralium Deposita, 28, 366–374.

Groves, D.I., Goldfarb, R.J., Gebre-Mariam, M., Hagemann, S.G., and Robert, F. (1998) Orogenic gold deposits: A proposed classification in the context of their crustal distribution and relationship to other gold deposit types. Ore Geology Reviews, 13, 7–27.

Groves, D.I., Bierlein, F.P., Meinert, L.D., and Hitzman, M.W. (2010) Iron oxide copper-gold (IOCG) deposits through Earth History: implications for origin, lithospheric setting, and distribution from other epigenetic iron oxide deposits.

Economic Geology, 105, 641–654.

Grunsky, E.C., Drew, L.J., Woodruff, L.G., Friske, P.W.B., Sutphin, D.M. (2013) Statistical variability of the geochemistry and mineralogy of soils in the maritime provinces of Canada and part of the Northeast United States. Geochemistry Exploration, Environment, Analysis, 13, 249–266.

Harlov, D.E., Meighan, C.J., Kerr, I.D., and Samson, I.M. (2016) Mineralogy, chemistry, and fluid-aided evolution of the Pea Ridge Fe oxide-(Y+ REE) deposit, southeast Missouri, USA. Economic Geology, 111, 1963–1984.

Hedenquist, J.W., Arribas, A.R., and Gonzalez-Urien, E. (2000) Exploration for epithermal gold deposits. Reviews in Economic Geology, 13, 245–277.

Hron, K., Templ, M., Filzmoser, P. (2010) Imputation of missing values for compositional data using classical and robust methods. Computational Statistics and Data Analysis, 5412, 3095–3107.

Hu, B., Zeng, L.P., Liao, W., Wen, G., Hu, H., Li, M.Y.H., and Zhao, X.F. (2022) The Origin and Discrimination of High-Ti Magnetite in Magmatic-Hydrothermal Systems: Insight from Machine Learning Analysis. Economic Geology, 117, 1613–1627.

Hu, H., Li, J.W., Harlov, D.E., Lentz, D.R., McFarlane, C.R.M., and Yang, Y.H. (2020) A genetic link between iron oxide-apatite and iron skarn mineralization in the Jinniu volcanic basin, Daye district, eastern China: Evidence from magnetite geochemistry and multi-mineral U-Pb geochronology. Geological Society of America Bulletin, 132, 899–917.

Huston, D.L., Sie, S.H., and Suter, G.F. (1995b) Selenium and its importance to the study of ore genesis: The theoretical basis and its application to volcanic-hosted massive sulfide deposits using pixeprobe analysis. Nuclear Instruments and Methods in Physics Research, Section B, 104, 476–480.

Huston, D.L., Sie, S.H., Suter, G.F., Cooke, D.R., and Both, R.A. (1995a) Trace elements in sulfide minerals from eastern Australian volcanic-hosted massive sulfide deposits: Part I. Proton microprobe analyses of pyrite, chalcopyrite, and sphalerite, and Part II. Selenium levels in pyrite; comparison with δ34S values and implications for the source of sulfure in volcanogenic hydrothermal systems. Economic Geology, 90, 1167–1196.

Jonsson, E., Harlov, D.E., MaJka, J., Högdahl, K., and Persson-Nilsson, K. (2016) Fluorapatite-monazite-allanite relations in the Grängesberg apatite-iron oxide ore district, Bergslagen, Sweden. American Mineralogist, 101, 1769–1782.

Keith, M., Haase, K.M., Chivas, A.R., and Klemd, R. (2022) Phase separation and fluid mixing revealed by trace element signatures in pyrite from porphyry systems. Geochimica et Cosmochimica Acta, 329, 185–205.

Knipping, J.L., Bilenker, L.D., Simon, A.C., Reich, M., Barra, F., Deditius, A.P., Lundstrom, C., Bindeman, I., and Munizaga, R. (2015b) Giant Kiruna-type deposits form by efficient flotation of magmatic magnetite suspensions. Geology 43, 591–594.

Knipping, J.L., Bilenker, L.D., Simon, A.C., Reich, M., Barra, F., Deditius, A.P., Wälle, M., Heinrich, C.A., Holtz, F., and Munizaga, R. (2015a). Trace elements in magnetite from massive iron oxide-apatite deposits indicate a combined formation by igneous and magmatic-hydrothermal processes. Geochimica et Cosmochimica Acta ,171, 15–38.

Koglin, N., Frimmel, H.E., Minter, W.E.L., and Brätz, H. (2010) Trace-element characteristics of different pyrite types in Mesoarchean to Palaeoproterozoic placer deposits. Mineralium Deposita, 45, 259–280.

Kuo, B.C., Ho, H.H., Li, C.H., Hung, C.C., and Taur, J.S. (2013). A kernel-based feature selection method for SVM with RBF kernel for hyperspectral image classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7(1), 317-326.

Large, R.R., Bull, S.W., McGoldrick, P.J., and Walters, S.G. (2005) Stratiform and Strata-Bound Zn-Pb-Ag Deposits in Proterozoic Sedimentary Basins, Northern Australia. Economic Geology 100th, 931–963.

Large, R.R., Danyushevsky, L., Hollit, C., Maslennikov, V., Meffre, S., Gilbert, S., Bull, S., Scott, R., Emsbo, P., Thomas, H., Singh, B., and Foster, J. (2009) Gold and trace element zonation in pyrite using a laser imaging technique: Implications for the timing of gold in orogenic and Carlin-style sediment-hosted deposits. Economic Geology, 104, 635–668.

Lawrence, D.M., Treloar, P.J., Rankin, A.H., Harbidge, P., and Holliday, J. (2013) The geology and mineralogy of the Loulo Mining District, Mali, West Africa: evidence for two distinct styles of orogenic gold mineralization. Economic Geology, 108, 199–227.

Li. R.C., Chen, H.Y., Xia, X.P., Yang, Q., Li, L., Danyushevsky, L.V., and Lai, C. (2018) Using integrated in-situ sulfide trace element geochemistry and sulfur isotopes to trace ore-forming fluids: Example from the Mina Justa IOCG deposit (southern Perú). Ore Geology Reviews, 101, 165–179.

Li. R.C., Chen, H.Y., Xia, X.P., Yang, Q., Li, L., Xu, J., Huang, C., and Danyushevsky, L.V. (2017) Ore fluid evolution in the giant Marcona Fe-(Cu) deposit, Perú: Evidence from in-situ sulfur isotope and trace element geochemistry of sulfides. Ore Geology Reviews, 86, 624–638.

Liang, P., Chen, L., Li, R.C., Xie, Y.L., Wu, C., and Lai, C. (2021) In-situ element geochemical and sulfur isotope signature of pyrite and chalcopyrite: Constraints on ore-forming processes of the Laoshankou iron oxide-copper (-gold) deposit, northern East Junggar. Ore Geology Reviews, 139, 104510.

Liu, Y.N. (2019) Metallogenic system and model of Fe deposits in Luzong volcanic basin, Anhui, eastern China. Hefei University of Technology. Doctor Thesis (in Chinese with English abstract).

Loftus-Hills, G., and Solomon, M. (1967) Cobalt, nickel and selenium in sulphides as indicators of ore genesis: Mineralium Deposita, 2, 228–242.

Lydon, J.W. (2007) An overview of the economic and geological contexts of Canada's major mineral deposit types, in Goodfellow, W.D., eds., Mineral deposits of Canada: a synthesis of major deposit types, district metallogeny, the evolution of geological provinces, and exploration methods. Geological Association of Canada, Mineral Deposits Division Special Publication, 5, 3–48.

Makvandi, S., Ghasemzadeh-Barvarz, M., Beaudoin, G., Grunsky, E.C., McClenaghan, M.B., Duchesne, C., and Boutroy, E. (2016) Partial least squares-discriminant analysis of trace element compositions of magnetite from various VMS deposit subtypes: Application to mineral exploration. Ore Geology Reviews, 78, 388–408.

Mao, J.W., Xie, G.Q., Duan, C., Pirajno, F., Ishiyama, D., and Chen, Y.C. (2011) A tectono-genetic model for porphyry-skarn-stratabound Cu-Au-Mo-Fe and magnetite-apatite deposits along the Middle-Lower Yangtze River Valley, Eastern China. Ore Geology Reviews, 43, 294–314.

Mao, M., Jody, S., Laurence, A.C., Stephen, M.R., and Alexei, S.R. (2016) Apatite Trace Element Compositions: A Robust New Tool for Mineral Exploration. Economic Geology, 111, 1187–1222.

Martin, A.J., McDonald, I., Jamieson, J.W., Jenkin, G.R.T., McFall, K.A., Piercey, G., MacLeod, C.J., and Layne, G.D. (2022) Mineral-scale variation in the trace metal and sulfur isotope composition of pyrite: implications for metal and sulfur sources in mafic VMS deposits. Mineralium Deposita, 57, 911–933.

Meinert, L.D., Dipple, G.M., and Nicolescu, S. (2005) World Skarn Deposits. Economic Geology 100th Anniversary Volume, 299–336.

Meng, L., Lan, C.Y., Zhan, Q., Wu, Q., and Zhao, T.P. (2022) Origin of the Shanggong gold deposit, the southern margin of the North China Craton: Constraints from Rb-Sr ages of sericite, and trace elements and sulfur isotope of pyrite. Ore Geology Reviews, 142, 104728.

Mukherjee, I., and Large, R. (2017) Application of pyrite trace element chemistry to exploration for SEDEX style Zn-Pb deposits: McArthur Basin, Northern Territory, Australia. Ore Geology Reviews, 81, 1249–1270.

Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A., and Brown, S. D. (2004). An introduction to decision tree modeling. Journal of Chemometrics: A Journal of the Chemometrics Society, 18(6), 275-285.

Nold, J.L., Dudley, M.A., and Davidson, P. (2014) The Southeast Missouri (USA) Proterozoic iron metallogenic province: Types of deposits and genetic relationships to magnetite-apatite and iron oxide-copper-gold deposits. Ore Geology Reviews, 57, 154–171.

Piercey, S.J. (2009) An overview of petrochemistry in the regional exploration for volcanogenic massive sulfide (VMS) deposits. Geochemistry Exploration, Environment, Analysis, 10, 1–18.

Praveen, M.N., Nambiar, C.G., and Huston, D.L. (2020) Geochemistry and petrogenesis of rhyolite hosted zinc-rich metamorphosed volcanogenic massive sulfide deposits in the eastern Betul Belt, central India. Ore Geology Reviews, 103918.

Ray, G.E. (2013) Skarns of the Canadian Cordillera. British Columbia Ministry of Energy and Mines, British Columbia Geological Survey Paper 2013–11, 112 p.

Reich, M., Deditius, A., Chryssoulis, S., Li, J.W., Ma, C.Q., Parada, M.A., Barra, F., and Mittermayr, F. (2013) Pyrite as a record of hydrothermal fluid evolution in a porphyry copper system: A SIMS/EMPA trace element study. Geochimica et

Cosmochimica Acta, 104, 42–62.

Reich, M., Kesler, S.E., Utsunomiya, S., Palenik, C.S., Chryssoulis, S.L., and Ewing, R.C. (2005) Solubility of gold in arsenian pyrite. Geochimica et Cosmochimica Acta, 69, 2781–2796.

Reich, M., Simon, A., Deditius, A., Barra, F., Chryssoulis, S., Lagas, G., Tardani, D., Knipping, J., Bilenker, L., Sánchez-Alfaro, P., Roberts, M.P., and Munizaga, R. (2016) Trace element signature of pyrite from the Los Colorados iron oxide-apatite (IOA) deposit, Chile: A missing link between Andean IOA and iron oxide copper-gold systems. Economic Geology, 11, 743–761.

Revan, M.K., Genç, Y., Maslennikov, V.V., Maslennikova, S.P., Large, R.R., and Danyushevsky, L.V. (2014) Mineralogy and trace-element geochemistry of sulfide minerals in hydrothermal chimneys from the Upper-Cretaceous VMS deposits of the eastern Pontide orogenic belt (NE Turkey). Ore Geology Reviews, 63, 129–149.

Santosh, M. and Groves, D.I. (2022) Global metallogeny in relation to secular evolution of the Earth and supercontinent cycles. Gondwana Research, 107, 395-422.

Savu-Krohn, C., Rantitsch, G., Auer, P., Melcher, F., and Graupner, T. (2011) Geochemical fingerprinting of coltan ores by machine learning on uneven datasets. Natural Resources Research, 20, 177-191.

Seedorff, E., Dilles, J.H., Proffett, J.M. Jr., Einaudi, M.T., Zurcher, L., and Stavast, W.J.A. (2005) Porphyry deposits: Characteristics and origin of hypogene features.

Economic Geology 100th Anniversary Volume, 251–298.

Sillitoe, R.H. (2010) Porphyry copper systems. Economic Geology, 105, 3–41.

Soltani, D.A., Lentz, D.R., and McFarlane, C.R.M. (2015) LA-ICP-MS analysis of volatile trace elements in massive sulphides and host rocks of selected VMS deposits of the Bathurst Mining Camp, New Brunswick: methodology and application to exploration, In: Targeted Geoscience Initiative 4: Contributions to the Understanding of Volcanogenic Massive Sulphide Deposit Genesis and Exploration Methods Development, (ed.) J.M. Peter and P. Mercier-Langevin; Geological Survey of Canada, Open File 7853, p. 59–80.

Spence-Jones, C.P., Jenkin, G.R.T., Boyce, A.J., Hill, N.J., and Sangster, C.J.S. (2018) Tellurium, magmatic fluids and orogenic gold: an early magmatic fluid pulse at Cononish gold deposit, Scotland. Ore Geology Reviews, 102, 894–905.

Tan, H.J., Shao, Y.J., Liu, Q.Q., Zhang, Y., Feng, Y.Z., Zhang, Y., and Shah, S.A. (2022) Textures, trace element geochemistry and in-situ sulfur isotopes of pyrite from the Xiaojiashan gold deposit, Jiangnan Orogen: Implications for ore genesis. Ore Geology Reviews, 144, 104843.

Templ, M., Kowarik, A., and Filzmoser, P. (2011) Iterative regression imputation using standard and robust methods. Comput. Stat. Data Anal., 55, 2793–2806.

Tardani, D., Reich, M., Deditius, A.P., Chryssoulis, S., Sánchez-Alfaro, P., Wrage, J., and Roberts, M.P. (2017) Copper-arsenic decoupling in an active geothermal system: A link between pyrite and fluid composition. Geochimica et Cosmochimica Acta, 204, 179–204.

van den Boogaart, K.G., and Tolosana-Delgado, R. (2013) Analyzing compositional data with R. first ed. Springer, Heidelberg-New York-Dodrecht-London, 258 p.

Wang, W.L., Zhao, J., and Cheng, Q.M. (2014) Mapping of Fe mineralization-associated geochemical signatures using logratio transformed stream sediment geochemical data in eastern Tianshan, China. Journal of Geochemical Exploration, 141, 6–14.

Wang, Y., Yang, X.D., Wang, C.T., Zhang, H.S., Liu, S.S., and Sun, C. (2022) Comparative studies on stratabound and skarn-type deposits, Tongling region, lower Yangtze River metallogenic Belt: Constraints from pyrite geochemistry. Ore Geology Reviews, 146, 104919.

Wei, Y.J., Yang, L.Q., Qiu, K.F., Wang, S.R., Ren, F., Dai, Z.H., Li, D.P., Shan, W., Li, Z.S., Wang, J.H., and Tang, L. (2022) Geology, mineralogy and pyrite trace elements constraints on gold mineralization mechanism at the giant Dayingezhuang gold deposit, Jiaodong Peninsula, China. Ore Geology Reviews, 148, 104992.

Xie, J.C., Tang, D.W., Qian L., Wang, Y., and Sun, W.D. (2020) Geochemistry of sulfide minerals from skarn Cu (Au) deposits in the Fenghuangshan ore field, Tongling, eastern China: Insights into ore-forming process. Ore Geology Reviews, 122, 103537.

Zhang, Y., Shao, Y.J., Chen, H.Y., Liu, Z.F., and Li, D.F. (2017) A hydrothermal origin for the large Xinqiao Cu-S-Fe deposit, Eastern China: Evidence from sulfide geochemistry and sulfur isotopes. Ore Geology Reviews, 88, 534–549.

Zhao, X.Y., Yang, Z.S., Zhang, X., Pei, Y.R., Liu, Y.C., and Han, D. (2020) Geochemical compositions of ore-related pyrite, quartz and fluid inclusions of Bangbu gold deposit in the Indus–Yurlung Zangbo suture zone, Tibet. Ore Geology Reviews, 120, 103441.

Figure captions

Figure 1 World map showing the location of samples in this study.

Figure 2 Box plots showing trace element contents of pyrite for each ore deposit. Line = median value; box = 25th–75th percentile; open circles = outliers (within 3 box length). Whiskers are drawn to the last data point that extends 1.5 times the length of the box toward the maximum and minimum.

Figure 3 Scatter plot of Co versus Ni in pyrite from different ore deposit types.

Figure 4 Discrimination diagrams for pyrite from category A (a) and category B (b).

Figure 5 Confusion Matrix for Category A (the top three, 1 = IOA, 2 = IOCG, 3 = Skarn Cu-(Fe) deposit, 4 = Porphyry Cu-Mo deposit) and Category B (the bottom three, 1 = SEDEX, 2 = Barren sedimentary rocks, 3 = Orogenic Au deposit, 4 =

DOI: https://doi.org/10.2138/am-2023-8976. http://www.minsocam.org/

VMS).

Figure 6 ROC curves for Category A (BLUE = RF, GREEN = ANN, RED = SVM) and Category B (BLUE = RF, GREEN = ANN, RED = SVM).

Table captions

Table 1 Summary of sample locations from different ore deposit types. IOA: iron oxide-apatite, IOCG: iron oxide copper-gold, VMS: volcanic-hosted massive sulfide, SEDEX: sedimentary exhalative.

Table 2 Summary of trace element data for pyrite from different ore deposit types. SD = standard deviation.

Table 3 Searching Spaces of Hyperparameters.

Table 4 Bayesian Optimization Results of Hyperparameters.

Table 5 Optimized Misclassification Matrix for Category A.

Table 6 Optimized Misclassification Matrix for Category B.

Table 7 Summary of Model Performances.

Supplementary Materials

Supp Table 1 Summary of LA-ICP-MS analytical methods for pyrites from different ore deposit types.

Supp Table 2 The original geochemical data of pyrite from different ore deposit types.

Supp Table 3 Summary of measurement spot numbers in each sample, deposit, ore clusters and deposit types.

Supp Table 4 Searching range of misclassification cost.

Supp Table 5 Coefficients and variables for discriminant functions. The trace element concentrations were logarithmic transformed.

Supp Table 6 Coefficients and variables for classification functions. The trace element concentrations were logarithmic transformed.

Supp Table 7 The classification results of the test data set.

DOI: https://doi.org/10.2138/am-2023-8976. http://www.minsocam.org/

Supplementary Materials The code for the performance of three machine learning algorithms.

This is the peer-reviewed, final accepted version for American Mineralogist, published by the Mineralogical Society of America.
The published version is subject to change. Cite as Authors (Year) Title. American Mineralogist, in press.
DOI: https://doi.org/10.2138/am-2023-8976. http://www.minsocam.org/

Table 1 Summary of training and test analyses from different ore deposit types used for Discrimination Analysis.

Deposit type	Deposit/locality name	Reference	Analyses	Number of training analyses	Number of test analyses	Total analyses
IOA	Los Colorados, Chile	Reich et al. 2016	48	37	11	126
IOA	Nihe, China	Che 2014; Liu 2019	78	62	16	
IOCG	Candelaria-Punta del Cobre district, Chile	del Real et al. 2020	162	130	32	279
IOCG	Marcona, Peru	Li et al. 2017	27	21	6	
IOCG	Mina Justa, Peru	Li et al. 2018	38	30	8	
IOCG	Laoshankou, China	Liang et al. 2021	52	41	11	
Skarn Cu-(Fe)	Xinqiao, China	Zhang et al. 2017	33	25	8	203
Skarn Cu-(Fe)	Fenghuangshan, China	Xie et al. 2020	118	94	24	
Skarn Cu-(Fe)	Xinqiao, Tongling region, China	Wang et al. 2022	25	21	4	
Skarn Cu-(Fe)	Baoshantao, Tongling region, China	Wang et al. 2022	27	22	5	
Porphyry Cu-Mo	Koloula, Solomon Island	Keith et al. 2022	264	212	52	419
Porphyry Cu-Mo	Verkhneuralskoe, Urals	Chugaev et al. 2022	85	68	17	
Porphyry Cu-Mo	Talitsa, Urals	Chugaev et al. 2022	21	16	5	
Porphyry Cu-Mo	Resolution, Arizona	Cooke et al. 2020	49	39	10	
Orogenic Au	Bangbu, Tibet, China	Zhao et al. 2020	6	5	1	386
Orogenic Au	Fofina, Africa	Augustin and Gaboury 2019	39	31	8	
Orogenic Au	Nyafé, Africa	Augustin and Gaboury 2019	18	15	3	
Orogenic Au	Siou, Africa	Augustin and Gaboury 2019	28	22	6	
Orogenic Au	Wona-Kona, Africa	Augustin and Gaboury 2019	85	68	17	
Orogenic Au	Yaho, Africa	Augustin and Gaboury 2019	14	11	3	
Orogenic Au	Yilgarn, Australia	Belousov et al. 2016	36	29	7	
Orogenic Au	Shanggong, China	Meng et al. 2022	51	40	11	
Orogenic Au	Xiaojiashan, China	Tan et al. 2022	89	72	17	
Orogenic Au	Dayingezhuang, China	Wei et al. 2022	20	16	4	

VMS	Bukit Botol	Basori et al. 2018	30	23	7	
VMS	Mala, Cyprus	Martin et al. 2022	123	99	24	
VMS	Yilgarn, Australia	Belousov et al. 2016	8	6	2	353
VMS	Bathurst mining camp, Canada	Soltani et al. 2015	30	24	6	
VMS	Bracemac-McLeod, Canada	Genna et al. 2015	162	131	31	
Sedimentary pyrite	Canning Basin, Australia	Gregory et al. 2015	134	107	27	
Sedimentary pyrite	Hamersly Basin, Australia	Gregory et al. 2015	165	130	35	
Sedimentary pyrite	Kalgoolie, Australia	Gregory et al. 2015	38	30	8	
Sedimentary pyrite	Pibara, Australia	Gregory et al. 2015	30	24	6	
Sedimentary pyrite	Perth Basin, Australia	Gregory et al. 2015	47	37	10	
Sedimentary pyrite	Arthur River, Australia	Gregory et al. 2015	29	23	6	
Sedimentary pyrite	Simithton, Australia	Gregory et al. 2015	52	41	11	828
Sedimentary pyrite	Amadeus Basin, Australia	Gregory et al. 2015	37	30	7	
Sedimentary pyrite	Yeneena Basin, Australia	Gregory et al. 2015	15	12	3	
Sedimentary pyrite	Yukon, Canada	Gregory et al. 2015	187	152	35	
Sedimentary pyrite	McArthur Basin, Australia	Gregory et al. 2015	30	24	6	
Sedimentary pyrite	Gummon, Australia	Gregory et al. 2015	38	31	7	
Sedimentary pyrite	Tanami, Australia	Gregory et al. 2015	26	21	5	
SEDEX	Howard'S Pass, Yukon	Gadd et al. 2016	474	379	95	693
SEDEX	Barney Creek, Australia	Mukherjee and Large 2017	219	173	46	

Table 2 Summary of trace element data for pyrite from different ore deposit types. SD = standard deviation.

Deposit	Statistic	Co	Ni	Cu	Zn	As	Se	Ag	Sb	Te	Au	Pb	Bi
IOA	n	126	126	126	126	126	126	126	126	126	126	126	126
	Min	90	20	0.11	100	1.1	3	0.1	0.04	0.04	0.04	100	0.1
	25%ile	622.5	160	20	100	60	54.1	1.85	5.52	3.105	0.08	200	9.46
	Average	3201.8	1407.4	788.0	2557.3	381.6	143.4	19.3	26.3	284.0	44.6	449.7	17.2
	Median	905	335	200	1370	129.5	110	4.45	18.67	240	0.14	514	18.58
	75%ile	1527.5	1300	532.5	4290	220	210	7.9	35.5	420	0.19	637	23.83
	Max	20000	9700	9700	9620	6430	770	200	100	1050	1200	757	38.77
	SD	5653.8	2407.6	1632.8	2749.3	1134.3	116.4	51.1	26.3	278.5	164.7	228.0	9.6
IOCG	n	279	279	279	279	279	279	279	279	279	279	279	279
	Min	1.29	1.80	0.10	0.25	0.17	5.00	0.01	0.01	0.07	0.01	0.01	0.01
	25%ile	957	61.38	4.04	1.9	69.79	25.77	0.028	0.032	0.137	0.08	0.11	0.051
	Average	6085.5	1189.7	277.6	11.8	917.3	55.0	4.0	3.8	0.9	61.2	42.8	5.1
	Median	3300	499.3	10.48	3.81	140.6	50.33	0.15	0.12	0.24	1.86	1.8	0.72
	75%ile	9223	1613	46.58	6.96	545.8	77.65	1.54	0.93	0.41	12	16.19	4.98
	Max	19910	8910	7933	604.1	9223.4	181.6	232.5	87.7	7.8	3237	1940	91.1
	SD	5991.9	1714.2	874.6	48.7	1857.0	35.1	16.7	11.6	1.8	285.6	176.9	11.4
Skarn Cu-(Fe)	n	203	203	203	203	203	203	203	203	203	203	203	203
	Min	0.05	0.097	0.03	0.144	0.168	0.014	0.009	0.006	0.03	0.005	0.005	0.007
	25%ile	1.31	0.74	0.42	0.46	7.10	0.14	0.16	0.07	0.11	0.029	0.72	0.22
	Average	168.34	58.69	164.65	7.57	954.15	4.74	6.30	46.54	7.43	0.33	235.32	33.18
	Median	11.03	3.17	1.614	0.76	51.71	1.14	0.57	0.43	0.58	0.05	3.742	1.015
	75%ile	43.08	37.06	29.3	1.98	793.02	3.92	6.23	6.66	3.96	0.29	69.5	7.94
	Max	9237	1135	8872.7	171.26	9223.4	86.1	68.5	2556.5	126.9	4.5	5685.3	1018

	SD	703.9	145.7	690.0	24.7	1824.8	11.7	13.9	226.5	19.0	0.7	688.7	108.9
Porphyry Cu-Mo	n	419	419	419	419	419	419	419	419	419	419	419	419
	Min	0.015	0.09	0.12	0.21	0.192	0.38	0.01	0.01	0.055	0.005	0.01	0.005
	25%ile	3.37	1.33	1.9	2.15	1.9	6.98	0.13	0.03	0.5	0.04	0.41	0.01
	Average	163.35	60.83	249.01	5.42	33.68	48.05	3.82	0.11	6.79	0.07	17.66	2.58
	Median	18.4	6.1	9.5	3.96	6.8	19.09	1.9	0.044	0.88	0.065	1.27	0.14
	75%ile	104.69	31.99	21.9	5.8	9.5	50	5.82	0.08	3.8	0.1	4.31	1.21
	Max	5094.9	1113.2	7990.2	128.0	2603.5	618.0	31.1	4.0	282.0	0.6	980.0	376.5
	SD	458.9	169.1	1141.7	9.8	202.3	89.8	5.3	0.3	21.8	0.1	72.7	19.0
Orogenic Au	n	386	386	386	386	386	386	386	386	386	386	386	386
	Min	0.1	0.102	0.044	0.059	0.115	0.553	0.010	0.013	0.142	0.008	0.033	0.006
	25%ile	27.17	41.28	7.69	2.20	495.26	7.95	0.20	1.13	1.62	0.20	4.59	0.30
	Average	213.66	368.89	196.14	105.03	3104.64	29.27	5.55	32.65	21.57	11.29	161.96	9.36
	Median	124.58	172.50	23.67	5.42	1728.61	16.70	0.68	7.15	3.80	0.75	24.91	1.40
	75%ile	260.25	379.30	134	27.7	3283.33	35.12	3.33	42.52	9.75	2.74	180.08	8.11
	Max	2581.5	5706.8	7000.0	7642.1	9223.4	674.2	263.0	1449.3	1320.0	1433.0	8694.1	202.7
	SD	310.8	623.9	570.5	522.9	4056.1	45.5	22.7	86.6	83.2	103.4	508.8	21.2
VMS	n	353	353	353	353	353	353	353	353	353	353	353	353
	Min	0.460	0.007	0.005	0.446	0.901	1.770	0.032	0.009	0.076	0.005	0.100	0.026
	25%ile	67.93	90.65	19.76	13.14	37.49	52.91	1.21	0.43	1.99	0.06	11.11	1.76
	Average	875.74	310.83	847.30	858.13	1003.22	171.65	75.29	35.67	54.05	0.76	505.25	89.06
	Median	243.2	193.5	222.2	105.4	424	106	6.4	1.2	6.7	0.14	51.28	10.79
	75%ile	650.1	362.15	1005.56	767.75	1320	209.34	24.85	11.79	39.73	0.46	325.7	56.07
	Max	9223.4	5890.0	8082.0	9223.4	9223.4	1898.5	1841.9	1744.0	2925.0	36.0	9223.4	2077.0

DOI: https://doi.org/10.2138/am-2023-8976. http://www.minsocam.org/

	SD	2174.2	525.2	1457.1	1692.5	1628.5	232.1	240.3	148.4	184.4	2.9	1322.9	247.2
	n	693	693	693	693	693	693	693	693			693	693
	Min	0.28	13.89	7.42	0.02	3.94	0.36	0.1	0.88			2.86	0.02
	25%ile	25.15	245.02	166.33	26.67	244.06	6.76	3.22	22.37			321.28	0.60
SEDEX	Average	226.16	1274.21	1300.37	290.60	891.77	56.27	26.97	131.49			1463.93	8.02
	Median	81.39	708.52	655.82	62.85	529.5	18.02	10.74	63.89			748.11	2.52
	75%ile	189.54	1710.88	1500.25	142.44	979.69	66.36	34.89	168.13			2014.61	7.25
	Max	4589.4	9039.3	9776.7	7905.6	16358.3	436.5	418.6	884.9			9223.4	170.9
	SD	473.4	1539.8	1794.3	762.3	1284.6	87.0	42.8	166.8			1752.4	18.7
	n	828	828	828	828	828	828	828	828	828	828	828	828
	Min	0.01	0.7	0.04	0.03	0.3	0.5	0.01	0.04	0.03	0.01	0.03	0.01
	25%ile	10.04	112.62	45.48	4.7	84.25	4.23	0.42	5.15	0.18	0.01	34.15	0.08
Sedimentary pyrite	Average	154.32	616.68	262.32	216.10	853.70	117.35	9.17	67.28	1.19	0.05	273.27	6.12
	Median	56.39	258.85	124.06	21.14	279.7	29.7	1.4	18.68	0.43	0.02	133.64	0.56
	75%ile	205.17	613.47	323.81	136.09	679.43	108.43	4.73	47.19	1.20	0.04	305.39	2.19
	Max	2859	8577	3237	8772	17850	4948	423	2878	29	1	5042	245
	SD	249.9	1018.3	376.7	753.5	1948.5	361.3	34.6	205.3	2.4	0.1	429.8	23.5

Table 3 Searching Spaces of Hyperparameters

	Hyperparameters	Type	Range
RF	$ntree$	Integer	$[100, 1000]$
	$mtry$	Integer	$[6, 12]$
SVM	C	Constant	$[1,1000]$
	σ	Constant	$[0.01, 2]$
MLP	$numH$	Integer	$[1,24]$
	$numL$	Integer	$[1,3]$

Table 4 Bayesian Optimization Results of Hyperparameters

	Category A	Category B	Min Objective (Mean overall error)
RF	$ntree = 714$	$ntree = 101$	$errA = 2.54\%$
	$mtry = 7$	$mtry = 6$	$errB = 3.85\%$
SVM	$C = 978.54$	$C = 284.42$	$errA = 10.97\%$
	$\sigma = 0.55468$	$\sigma = 0.33861$	$errB = 14.76\%$
MLP	$numL = 3$	$numL = 3$	$errA = 7.88\%$
	$numH = \{24, 24, 23\}$	$numH = \{23, 24, 20\}$	$errB = 18.51\%$

Table 5 Optimized Misclassification Matrix for Category A

	IOA	IOCG	Skarn Cu-(Fe)	Porphyry Cu-Mo
IOA	0	0.54	0.77	0.59
IOCG	0.46	0	0.68	0.50
Skarn Cu-(Fe)	0.23	0.32	0	0.23
Porphyry Cu-Mo	0.41	0.50	0.77	0

Table 6 Optimized Misclassification Matrix for Category B

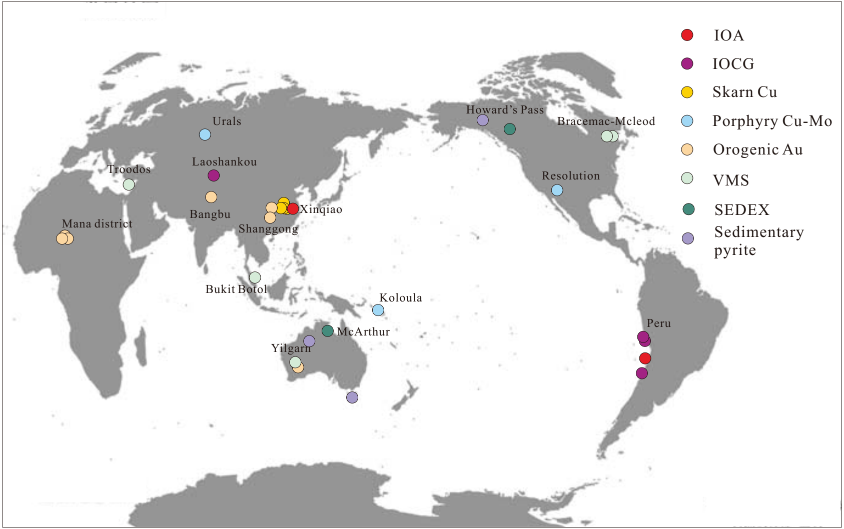
	SEDEX	Barren sedimentary pyrite	Orogenic Au	VMS
SEDEX	0	0.50	0.27	0.42
Barren sedimentary pyrite	0.50	0	0.46	0.24
Orogenic Au	0.73	0.53	0	0.58
VMS	0.58	0.76	0.42	0

DOI: https://doi.org/10.2138/am-2023-8976. http://www.minsocam.org/

Table 7. Summary of Model Performances

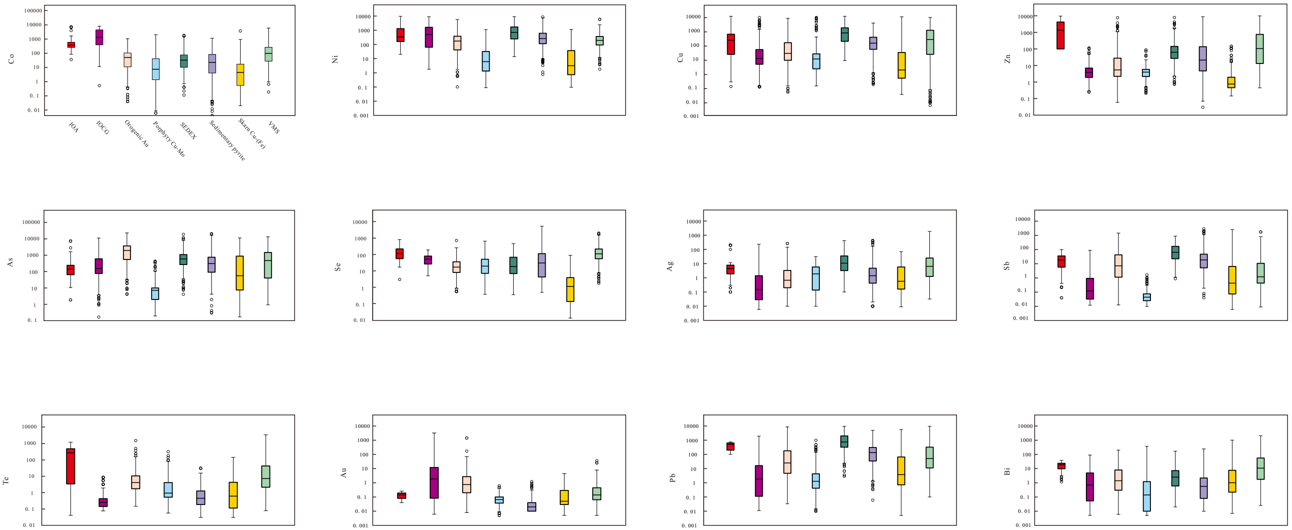
Category A	IOA	IOCG		Skarn Cu-(Fe)	Porphyry Cu-Mo	Overall Accuracy
RF	1	0.964		0.964	0.951	0.966
SVM	1	0.964		0.976	1	0.981
MLP	0.962	0.982		0.976	0.976	0.976
Category B	SEDEX	Barren rocks	sedimentary	Orogenic Au	VMS	Overall Accuracy
RF	1	0.982		0.962	0.873	0.967
SVM	0.986	0.994		0.962	0.916	0.974
MLP	0.978	0.994		0.923	0.916	0.965

Figure 1



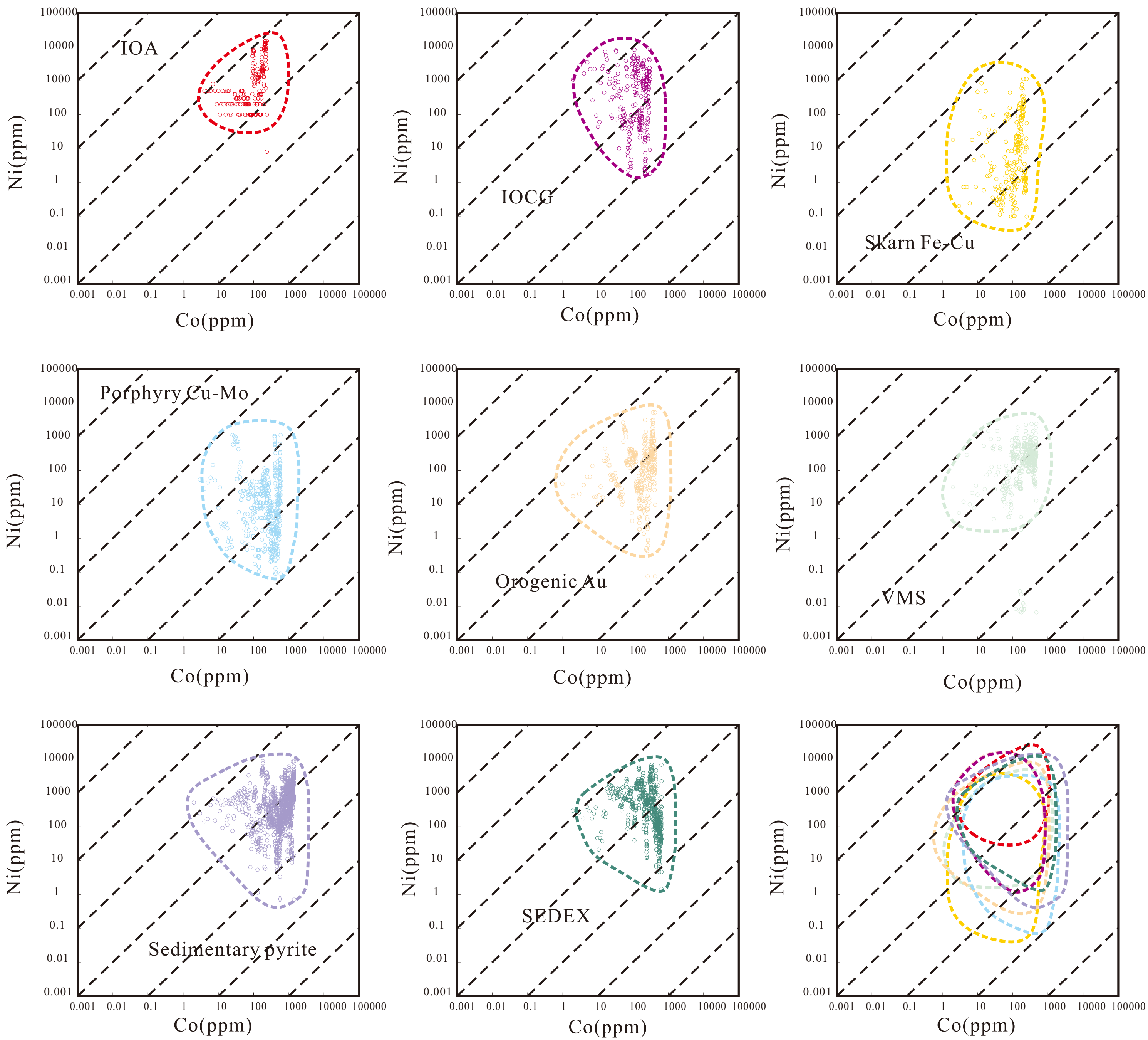
Urals
Laoshankou
Troodos
Bangbu
Xinqiao
Shanggong
Mana district
Bukit Botol
Koloula
McArthur
Yilgarn
Howard's Pass
Bracemac-Mcleod
Resolution
Peru
IOA
IOCG
Skarn Cu
Porphyry Cu-Mo
Orogenic Au
VMS
SEDEX
Sedimentary pyrite

Figure 2



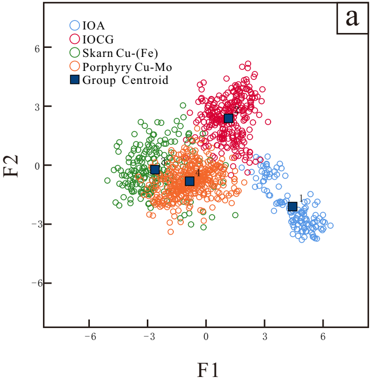
Co
Ni
Cu
Zn
As
Se
Ag
Sb
Te
Au
Pb
Bi
IOA
IOCG
Orogenic Au
Porphyry Cu-Mo
SEDEX
Sedimentary pyrite
Skarn Cu-(Fe)
VMS

Figure 3

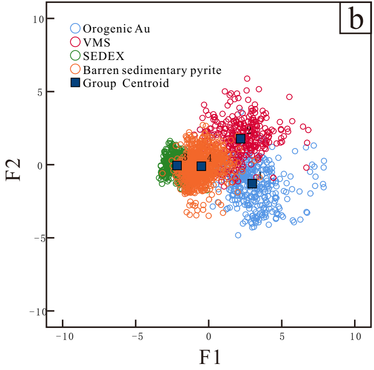


IOA
IOCG
Skarn Fe-Cu
Porphyry Cu-Mo
Orogenic Au
VMS
Sedimentary pyrite
SEDEX
Ni(ppm)
Co(ppm)
0.001
0.01
0.1
1
10
100
1000
10000
100000

Figure 4



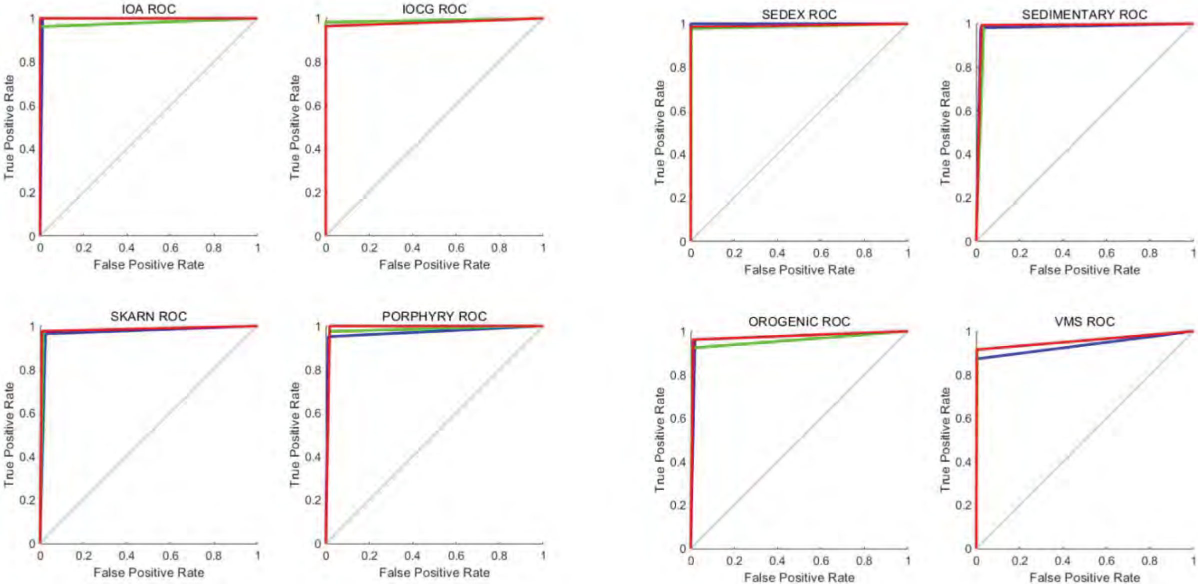
a
IOA
IOCG
Skarn Cu-(Fe)
Porphyry Cu-Mo
Group Centroid
F2
F1
6
3
0
-3
-6
-6
-3
0
3
6
1
2
3
4



b
Orogenic Au
VMS
SEDEX
Barren sedimentary pyrite
Group Centroid
F2
F1
10
5
0
-5
-10
-10
-5
0
5
10
1
2
3
4

Figure 5

RF Confusion Matrix
ANN Confusion Matrix
SVM Confusion Matrix
Output Class
Target Class
RF Confusion Matrix
ANN Confusion Matrix
SVM Confusion Matrix
Output Class
Target Class


IOA ROC
IOCG ROC
SEDEX ROC
SEDIMENTARY ROC
SKARN ROC
PORPHYRY ROC
OROGENIC ROC
VMS ROC
True Positive Rate
False Positive Rate
0
0.2
0.4
0.6
0.8
1