JGR Solid Earth

RESEARCH ARTICLE
10.1029/2022JB024584

Special Section:
Machine learning for Solid Earth observation, modeling and understanding

Application of Machine Learning to Characterizing Magma Fertility in Porphyry Cu Deposits

Shaohao Zou[1], Xilian Chen[1], Matthew J. Brzozowski[2], Cheng-Biao Leng[1], and Deru Xu[1]

[1]State Key Laboratory of Nuclear Resources and Environment, East China University of Technology, Nanchang, China, [2]Department of Geology, Lakehead University, Thunder Bay, ON, Canada

Key Points:

- A global data set of magmatic zircon trace-element chemistry is used to train two separate machine learning algorithms
- The Eu/Eu*, Eu/Eu*/Y, Ce/Nd, and Ce/Ce* ratios, and Dy, Hf, and Ti contents are important for distinguishing fertile and barren zircons
- The performance of the machine learning models is not affected by regional differences and geological setting

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

S. Zou and D. Xu,
shaohaozou@hotmail.com;
xuderu@gig.ac.cn

Citation:

Zou, S., Chen, X., Brzozowski, M. J., Leng, C.-B., & Xu, D. (2022). Application of machine learning to characterizing magma fertility in porphyry Cu deposits. *Journal of Geophysical Research: Solid Earth*, *127*, e2022JB024584. https://doi.org/10.1029/2022JB024584

Received 15 APR 2022
Accepted 21 JUL 2022

Shaohao Zou and Xilian Chen contributed equally to this work.

Author Contributions:

Conceptualization: Shaohao Zou, Xilian Chen
Funding acquisition: Shaohao Zou, Cheng-Biao Leng, Deru Xu
Investigation: Shaohao Zou, Xilian Chen
Methodology: Shaohao Zou
Supervision: Deru Xu
Visualization: Shaohao Zou

Abstract Large and easily accessible porphyry Cu deposits have already been identified, exploited, and gradually exhausted. Novel strategies, therefore, are required to identify new, deeply buried deposits. Previous studies have proposed several lithogeochemical and mineralogical approaches for identifying porphyry Cu systems. Most of these methods, however, require significant a priori knowledge of the exploration region and are, generally, of low effectiveness. In this study, machine learning models using Random Forest and Deep Neural Network algorithms are utilized to characterize magma fertility. The two models have first been trained using a large trace-element data set of magmatic zircon and then validated on unseen data set during the training process. The performance of both models was evaluated using a fivefold cross-validation technique, which demonstrates that the models provide consistent results and yield good classification accuracy (~90% on average) with low false positive rates. Feature importance analysis of the models suggests that Eu/Eu*, Eu/Eu*/Y, Ce/Nd, Ce/Ce*, Dy, Hf, and Ti are the important parameters that distinguish fertile and barren zircons. The real-world applicability of the validated models was evaluated using two well-characterized porphyry Cu deposits in subduction and postcollisional settings—the Highland Valley porphyry Cu district (south-central British Columbia, Canada) and the southern Gangdese belt (Tibet, China), respectively. The results demonstrate that our generalized models can discriminate zircon from igneous rocks associated with porphyry Cu deposits from those in nonmineralized systems with high accuracy and independent of geological setting, suggesting that this new approach can be used effectively in greenfield and brownfield exploration.

Plain Language Summary Mineral resources, which are important to the development of human society, are increasingly being consumed as our need for technological advancement increases. With this consumption comes the need to identify additional resources, many of which may be invisible at the Earth's surface. This challenges current exploration methods, requiring novel techniques to be developed to identify variably mineralized rock below Earth's surface. In this study, machine learning methods are trained using a global data set of zircon chemistry to evaluate the prospectivity of porphyry Cu mineralization in magmatic districts. This work demonstrates that machine learning methods provide a robust, accurate, and effective approach to identifying porphyry Cu mineral resources.

1. Introduction

With the global community moving toward a low-carbon future, the demand for minerals and metals has surged in the past decade (Ali et al., 2017). Despite the increased investment in mineral exploration, however, the rate of discovery of mineral deposits has steadily decreased (Lusty & Gunn, 2015). One of the main reasons for this is that large, shallow, and high-grade deposits have largely been identified and exploited, requiring exploration geologists to identify variably mineralized "blind" systems that are overlain by tens to hundreds of meters of barren cover (Brugger et al., 2010). To circumvent these challenges, new exploration methods and technologies are urgently needed to assist geologists in discovering new mineral deposits (Holliday & Cooke, 2007).

Porphyry Cu deposits provide most of world's copper, as well as significant quantities of gold, molybdenum, and rhenium (Sillitoe, 2010), making them key exploration targets for the mining industry (Richards, 2016). These deposits form as a result of fluid exsolution derived from hydrous and oxidized magmas (Richards, 2015; Sillitoe, 2010) that are characterized by high Sr/Y, V/Sc, Sr/MnO, Ta/Nb, and Eu/Eu* ratios (Baldwin & Pearce, 1982; Lang & Titley, 1998; Loucks, 2014; Richards, 2011). These whole-rock geochemical signatures are, therefore, often used as indicators of magma fertility to assess metallogenic potential in magmatic districts (Loucks, 2014; Y.-J. Lu et al., 2015; Richards & Kerrich, 2007; Wells et al., 2020). However, these signatures can

Writing – original draft: Shaohao Zou, Matthew J. Brzozowski
Writing – review & editing: Shaohao Zou, Xilian Chen, Matthew J. Brzozowski, Cheng-Biao Leng

easily be affected by weathering, hydrothermal alteration, and metamorphism, leading to ambiguous conclusions when they are utilized as prospectivity indicators during mineral exploration (e.g., Chen et al., 2019; R. Wang et al., 2017), especially when the geochemical data scatter close to classification boundaries.

Zircon is a ubiquitous and geochemically robust accessory mineral in intermediate to felsic igneous rocks (Cherniak & Watson, 2003). It contains high concentrations of rare earth elements (REEs) and other elements, such as Th, U, Hf, Ti, and Y (Hoskin & Schaltegger, 2003), recording the compositional characteristics and physicochemical conditions (e.g., temperature, oxidation state, and water content) of the parent magma (Ballard et al., 2002; Claiborne et al., 2010; Dilles et al., 2015). Consequently, trace-element chemistry and trace-element ratios (e.g., Eu/Eu*, Ce/Ce*, Ce/Nd, and Eu/Eu*/Y) in zircon are also commonly utilized as magma fertility indicators to diagnose the metallogenic fertility of the source magma (e.g., Dilles et al., 2015; Y.-J. Lu et al., 2015; Shu et al., 2019). Several studies, however, have demonstrated that these one-dimensional zircon trace-element ratios or binary classification diagrams for porphyry Cu deposits in different regions do not match well either with each other or with preexisting classification ranges (e.g., Chen et al., 2019; Pizarro et al., 2020; S. Zhong et al., 2019); this may be because some other zircon trace-element signals that record information on magma fertility are not included in the classification, leading to false positives. Mathematically, zircon chemistry can be regarded as high-dimensional vectors, which is challenging to fully represent and understand using traditional data visualization techniques.

Machine learning is the science of using computational algorithms to identify patterns in data and applying them to make predictions; it provides a powerful toolset for decoding hidden information in high-dimensional data. Over the past decade, machine learning algorithms have been widely applied in the Earth sciences to solve complex problems, including determining the temperature and pressure of magma crystallization (e.g., Petrelli et al., 2020), estimating crustal thickness (e.g., Zou et al., 2021), identifying the tectonic setting in which rocks were generated (e.g., Kuwatani et al., 2015; Petrelli & Perugini, 2016; R. Zhong et al., 2021), predicting geothermal heat flux (e.g., Lösing & Ebbing, 2021; Rezvanbehbahani et al., 2017), and aiding in geochemical exploration (e.g., Nathwani et al., 2022). In this contribution, two popular, supervised machine learning algorithms, Random Forest and Deep Neural Network, were utilized to evaluate the fertility of magmas related to porphyry Cu deposits. The Random Forest and Deep Neural Network models were first trained using the trace-element chemistry of zircon grains from barren and porphyry Cu-mineralized rocks worldwide. After performance evaluation, the fertility of the Highland Valley porphyry Cu district (south-central British Columbia, Canada) and the southern Gangdese belt (Tibet, China) was assessed using these trained models. A fertility index based on the outputs of the machine learning models was then designed to quantify magma fertility. The potential for mineralization predicted by the fertility index is then compared with previous lithogeochemical and mineralogical information. The comparison demonstrates that the machine learning models can improve the efficiency and accuracy of distinguishing barren and porphyry-fertile rocks and highlights the potential of machine learning methods to improving the success rate of identifying deeply buried deposits.

2. Data Sources and Filtration

Fertile magmas in this study refer to magmas that can generate porphyry deposits, while barren magmas represent the opposite (Loader et al., 2022; Wilkinson, 2013). All of the zircon trace-element data utilized for model training and validation were collated from the GEOROC database (http://georoc.mpch-mainz.gwdg.de/georoc/) and preexisting references. Zircon trace-element data collected from representative mineralized intrusions from 27 economic porphyry deposits around the world were labeled as "fertile zircon," whereas data from intrusions without any mineralization were labeled as "barren zircon." The sources of these data sets are summarized in Table 1 and the sample locations are illustrated in Figure 1. The full data set can be found in Table S1 in Supporting Information S1.

The following steps were applied to filter the collected data. First, zircon analyses with La and Ti contents greater than 1 and 50 ppm, respectively, were excluded as these analyses may have been contaminated by apatite and/or Fe–Ti oxides (Y. Lu et al., 2016). Second, zircons with Th/U ratios less than 0.1 were discarded as they may be of metamorphic origin (Hoskin & Schaltegger, 2003). Third, analyses with missing trace-element values were rejected in order to maintain a cohesive and complete data set. The data set resulting from this filtering process comprises a total of 1,016 zircon grains from fertile rocks and 1,037 zircon grains from barren rocks and includes

Table 1
Zircons From Barren and Fertile Rocks From Literature Complied in This Study

No.	Deposit name	Country	Deposit type	Complied zircon numbers	Source
1	Borly	Kazakhstan	Porphyry Cu	5	Shen et al. (2015)
2	Baogutu	China	Porphyry Cu	6	Shen et al. (2015)
3	Shujiadian	China	Porphyry Cu	14	S. Wang et al. (2016)
4	Opache	Chile	Porphyry Cu	20	Ballard et al. (2002) and Campbell et al. (2006)
5	Tuwu-Yandong	China	Porphyry Cu	29	Shen et al. (2015) and Xiao et al. (2017)
6	Koksai	Kazakhstan	Porphyry Cu	5	Shen et al. (2015)
7	Nurkazghan	Kazakhstan	Porphyry Cu–Au	10	Shen et al. (2015)
8	Bozshakol	Kazakhstan	Porphyry Cu	17	Shen et al. (2015)
9	Jiama	China	Porphyry Cu–Mo–Au	42	Y. Lu et al. (2016) and Yang et al. (2016)
10	Kounrad	Kazakhstan	Porphyry Cu	11	Shen et al. (2015)
11	Sungun	Iran	Porphyry Cu–Mo	45	Hezarkhani (2006) and Y. Lu et al. (2016)
12	Dexing	China	Porphyry Cu–Au	48	G.-G. Wang et al. (2015) and Y. Lu et al. (2016)
13	Yulong	China	Porphyry Cu–Mo	90	Jiang et al. (2006) and Li et al. (2012)
14	Batu Hijau	Indonesia	Porphyry Cu–Au	63	Fiorentini and Garwin (2010) and Y. Lu et al. (2016)
15	Tampakan	Philippines	Porphyry Cu–Au	51	Y. Lu et al. (2016)
16	El Abra	Chile	Porphyry Cu	9	Ballard et al. (2002) and Correa et al. (2016)
17	Qulong	China	Porphyry Cu–Mo	30	Y. Lu et al. (2016) and Yang et al. (2016)
18	Erdenet	Mongolia	Porphyry Cu	8	Shen et al. (2015)
19	Aktogai	Kazakhstan	Porphyry Cu–Mo	8	Shen et al. (2015)
20	Tintaya	Peru	Porphyry Cu–Au	25	Chelle-Michou et al. (2014, 2015)
21	Sar Cheshmeh	Iran	Porphyry Cu–Mo	50	Asadi et al. (2014) and Y. Lu et al. (2016)
22	Radomiro Tomic	Chile	Porphyry Cu	20	Ballard et al. (2002) and Cooke et al. (2005)
23	Oyu Tolgoi	Mongolia	Porphyry Cu–Au–Mo	526	Wainwright et al. (2011), Porter (2016), and Loader et al. (2017)
24	Chuquicamata	Chile	Porphyry Cu–Mo–Au	30	Ballard et al. (2002) and Cooke et al. (2005)
25	El Teniente	Chile	Porphyry Cu–Mo	37	Cooke et al. (2005) and Munoz et al. (2012)
26	EI Salvador	Chile	Porphyry Cu	172	Lee et al. (2017)
27	Red Chris	Canada	Porphyry Cu–Au	40	J.-J. Zhu et al. (2018)
28	GEOROC	Worldwide	Barren	1,278	GEOROC
29	Red Chris	Canada	Not-ore associated	40	J.-J. Zhu et al. (2018)
30	Bishop	United States	Barren	19	Reid et al. (2011)
31	Yellowstone	United States	Barren	18	Wotzlaw et al. (2015)
32	Younger Toba Tuff	Indonesia	Barren	20	Smythe and Brenan (2016)

15 trace elements (Ti, Ce, Nd, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu, Y, and Hf). The REE contents of zircon are highly correlated (linearly dependent) and, hence, have the same impact on the machine learning algorithm, which may mask interactions between different elements and result in overfitting (Lösing & Ebbing, 2021). To reduce the probability of misinterpretation of the machine learning outputs as a result of this similar behavior of the REE and to evaluate previous mineralization indicators, the revised Ce/Ce* and Eu/Eu* (regression with

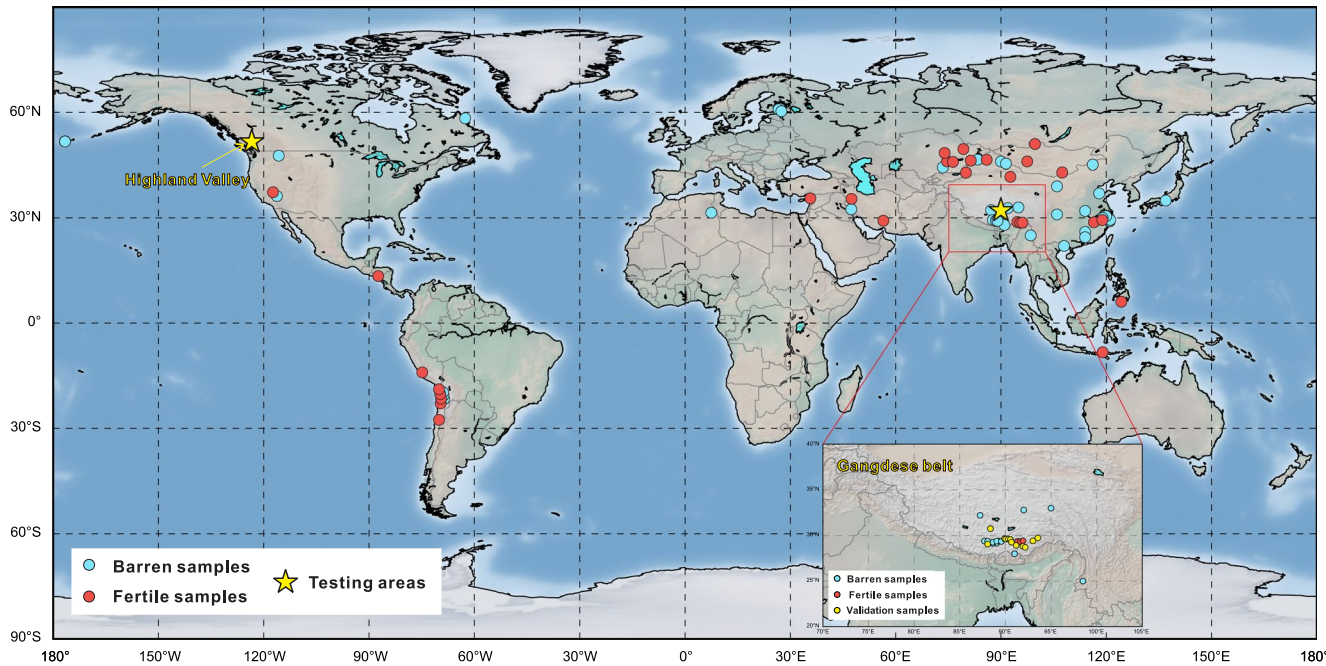

Figure 1. Global map showing the distribution of fertile and barren samples and testing areas in this study. The insert map shows the location of complied samples in the Gangdese belt. Validation samples in the Gangdese belt were used to validate the constructed machine learning models.

an exponential power function described in S. Zhong et al., 2019), Eu/Eu*/Y, Ce/Nd, and (Ce/Nd)/Y (Y. Lu et al., 2016) were also calculated and added to the data set that was used to train and test the models.

In addition to the global zircon data set, zircon trace-element data of intrusions in Highland Valley porphyry Cu district (British Columbia, Canada; Table S2 in Supporting Information S1) and the Cenozoic igneous rocks in the southern Gangdese belt (Tibet, China; Table S3 in Supporting Information S1) were also aggregated. These regional data sets were filtered using the same criteria as in the global data set and were utilized to validate the constructed machine learning methods.

3. Methods

3.1. Machine Learning Algorithms

The Random Forest algorithm (Breiman, 2001) has been widely applied to solve classification and regression problems, such as predicting the origin of basalts (cf., Zhao et al., 2019) and the fertility of igneous rocks with respect to porphyry Cu mineralization (Nathwani et al., 2022). It is an ensemble learning technique that utilizes a series of decision trees to identify the most popular classification (the final output; i.e., the classification, i.e., selected by the majority of the decision tree votes; Breiman, 2001). Compared with a single decision tree, the Random Forest algorithm reduces the variation in the final result when doing repeated calculations, thus greatly reducing the error rate. In the Random Forest algorithm, each decision tree is constructed by a bootstrap resampling of the training data set, thereby reducing the variance of the decision tree (Breiman, 2001). In this study, the Random Forest algorithm was used initially to classify the zircon grains as fertile or barren, and then to extract the importance of features during the classification process. The Scikit-Learn toolkit was used to construct the Random Forest model (Pedregosa et al., 2011).

Deep Neural Networks are a special kind of artificial neural network with multiple layers hidden between the input and output layers. They are designed to solve supervised and reinforcement learning problems (LeCun et al., 2015) and benefit from the fact that mathematical equations are not required to describe the mapping relationships between the input and output data sets. Accordingly, this method has been widely applied to processing nonlinear classification and prediction problems in the Earth sciences (e.g., Reichstein et al., 2019; Zhao et al., 2019; R. Zhong et al., 2021). In this study, trace elements in zircon were regarded as input parameters and

presented to the neural network, and zircon classification (i.e., from fertile vs. barren rock) was regarded as the target output. During the training process, both the input parameters and the target data were passed to the neutral network to regress their relationships, and misclassifications were evaluated by comparing the difference between the target data and the model output results. Here, the Deep Neural Network model comprises four hidden layers, with 20, 50, 25, and 2 neurons. The cross-entropy method (Shore & Johnson, 1980) and Adaptive Moment Estimation (ADAM; Kingma & Ba, 2014) were implemented to ensure that the model solution achieves the best possible match between the input data and the target data and minimizes the mean loss (i.e., how far the output is from the target); this was done for both the training and testing data sets. The contribution of individual features to the prediction of the Deep Neural Network model was calculated using algorithms in the SHAP (SHapley Additive exPlanations) package (Lundberg & Lee, 2017); the SHAP value of a feature indicates how much it impacts the model prediction (i.e., features with large absolute SHAP values indicate that they are important). In this study, Deep Neural Network model building and training were accomplished using the Python deep learning API, Keras (https://github.com/fchollet/keras).

3.2. Model Construction Processes

3.2.1. Data Preprocessing

The concentration of trace elements in zircon is expressed in parts per million that sum to a constant (Liu et al., 2008), not an absolute value. A constant sum may result in the concentration of most trace elements not being normally distributed and lead to spurious high correlations between elements (e.g., Chayes, 1960). To reduce the probability of misinterpretation and overfitting, it is, therefore, necessary to transform the original data set. In this study, the centered logratio (CLR) transformation (Aitchison, 1986) was applied before constructing the Random Forest and Deep Neural Network models; the pyrolite package (Williams et al., 2020) in python was used to do the CLR-transformation. The filtered and transformed data set was then randomly subdivided into a training set (80% of the data) and a testing set (20% of the data) using the hold-out method, while maintaining the original proportions of each class; data from the Highland Valley and the southern Gangdese belt were not included in the training set. The training data set was used to train the constructed classifiers and the testing set was utilized to evaluate the performance of the classifiers.

3.2.2. Model Hyperparameter Tuning

Model tuning was accomplished using a grid search strategy with the fivefold cross-validation method (Kohavi, 1995), which exhaustively generates candidates from a grid of parameter values. 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. For the fivefold cross validation, the entire data set was randomly divided into five equal-sized subsets. Four subsets were used to train the machine learning model, and the remaining subset was regarded as the validation subset to evaluate the classification mode. The training and evaluating processes were then repeated five times, with the output score being the mean value of all five predictions. To obtain the optimal hyperparameters for the Random Forest and Deep Neural Network models, a grid search method was used, which fits models to the data set using different combinations of the specified hyperparameters and returns scores via cross validation. 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. The detailed processes of hyperparameters selection and model construction can be found in Zou et al. (2021).

3.2.3. Model Evaluation

Once the machine learning models were constructed, they were evaluated using the testing set. The performance of both models over the testing set was evaluated based on their accuracy, precision, recall, and F1 scores. In binary classification, the accuracy score measures the proportion of the classes that are predicted correctly, whereas the precision is the number of correct predictions among all of the predictions. The recall of a class is defined as the ratio of true positive predictions/(true positive predications + false negative predications), where a true positive is defined as the number of the positive observations that are predicted correctly and a false negative is defined as the number of positive observations that are incorrectly predicted. The F1 score is the harmonic mean value between precision and recall. For model evaluation, a higher F1 score corresponds to better classification quality of the model, and vice versa. In addition, the precision–recall curve is also used for evaluating

the performance of the machine learning classifiers, where a large area under the curve (AUC; maximum = 1) implies high precision and high recall.

4. Result and Discussion

4.1. Conventional Zircon Fertility Indicators

Previous studies have demonstrated that magmas with a high-water content (≥4 wt.% H_2O at epizonal depth or >10 wt.% H_2O at high pressure; Y.-J. Lu et al., 2015; Richards et al., 2012) and a high oxidation state (FMQ +1 to +2; Richards, 2015) are key to the formation of porphyry Cu deposits. The whole-rock geochemistry of such magmas is characterized by high Sr/Y, Ce/Nd, Dy/Y, and (Ce/Nd)/Y ratios (e.g., Loucks, 2014; Y.-J. Lu et al., 2015), which reflect hornblende fractionation and suppression of plagioclase. These whole-rock geochemical signatures should be inherited by zircon that crystallize from such magmas. Accordingly, previous studies have proposed several zircon fertility indicators to discriminate fertile igneous suites from barren suites (e.g., Y.-J. Lu et al., 2015; Shen et al., 2015; Shu et al., 2019).

Based on our compiled data set, the performance of several traditional zircon fertility indicators was reassessed. Consistent with previous studies (e.g., Dilles et al., 2015; Y. Lu et al., 2016; S. Zhong et al., 2019), these indicators can relatively robustly distinguish between zircon from fertile and barren intrusions (Figure 2). They can identify fertile igneous suites well, with relatively high accuracy (from 0.65 to 0.83) and true positive rates (from 0.87 to 0.97), suggesting that these indicators may be helpful in identifying fertile igneous suites. That being said, these geochemical indicators also produce a significant number of false positive observations, with false positive rates ranging from 0.23 to 0.66 (Figure 2). This overlap likely results from the fact that many of the trace-element ratios (such as [Ce/Nd]/Y, Eu/Eu*/Y, and Ce/Nd ratios, and Eu and Ce anomaly) may not only be controlled by the water content and oxidation state of magmas but also be influenced by other magmatic processes, such as the crystallization of REE-bearing minerals before zircon (Buret et al., 2016; Loader et al., 2017, 2022). Therefore, if these discrimination diagrams with high false positive rates were applied to greenfield or brownfield exploration, the results would be inconclusive and lead to inaccurate mineralization targets. It is for this reason that machine learning models need to be applied—more of the controls on magma and mineral chemistry need to be considered when developing robust mineral exploration tools.

4.2. Machine Learning Results and Their Performances

Average performance metrics (i.e., accuracy, precision, recall, and F1 scores) for the Random Forest and Deep Neural Network models from the fivefold cross-validation process are all between 0.90 and 0.92. The normalized confusion matrix (Figure 3) shows that the Random Forest model can identify zircons from barren and fertile magmas with high classification accuracies of 0.90 and 0.93, respectively, with a false positive rate of 0.10. Zircons from barren and fertile rock are discriminated with accuracies of 0.85 and 0.95, respectively, with a false positive rate of 0.15 using the Deep Neural Network model. Compared to traditional zircon fertility indicators (Figure 2), the machine learning models developed here display a higher performance. The precision–recall curves for barren and fertile zircons are plotted in Figure 4. The AUC for the Random Forest model (Figure 4a) varies between 0.96 and 0.97 and between 0.94 and 0.95 for the Deep Neural Network model (Figure 4b). The results indicate that both constructed models can robustly distinguish between zircon from barren and fertile intrusions.

4.3. Feature Importance for Classifying Magma Fertility

Fertility classification relies entirely on the input features. Our data set contains 15 trace elements and 5 trace-element ratios of zircon grains as classification features. Therefore, it is important to understand which features contribute the most to the accurate discrimination of zircon from barren and fertile rocks. In this study, the feature importance techniques described in Section 3.1 were used to calculate a score for all of the input features for the Random Forest and Deep Neural Network models. A higher importance score means that a given feature plays a greater role in discriminating between zircon from fertile and barren rocks, and vice versa. For the Random Forest model, feature importance scores for identifying mineralization are, from high to low, Eu/Eu*, Eu/Eu*/Y, Hf, (Ce/Nd)/Y, Ce/Nd, Eu, Nd, Ce/Ce*, Dy, Ti, Sm, Lu, Ce, Yb, Tb, Gd, Tm, Er, Y, and Ho (Figure 5a).

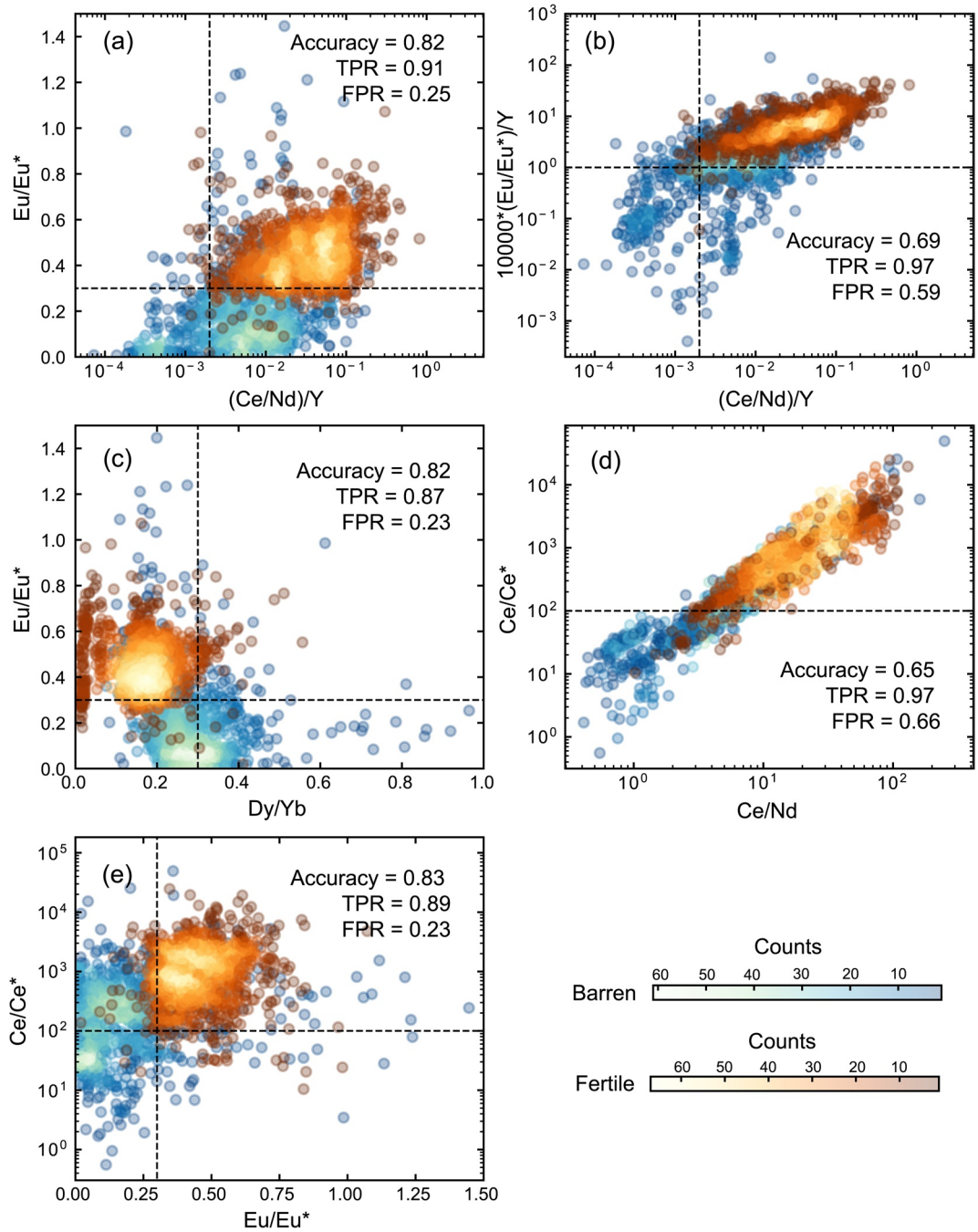

Figure 2. Scatterplots of trace-element ratios of fertile and barren zircon. The boundaries between fertile and barren zircon are from Y. Lu et al. (2016) and Pizarro et al. (2020). The accuracy, true positive rate (TPR) and false positive rate (FPR) are calculated based on these boundaries. These parameters are calculated using the same data used for training the machine learning models, thus allowing comparison between binary scatterplots and machine learning classifiers.

In contrast, feature importance scores for the Deep Neural Network model are, from high to low, (Ce/Nd)/Y, Ce/Ce*, Dy, Gd, Ce/Nd, Eu/Eu*, Hf, Ce, Eu, Nd, Y, Sm, Ho, Tb, Eu/Eu*/Y, Er, Ti, Lu, Yb, and Tm (Figure 5b). In addition, the influence of the number of features on model accuracy was assessed by sequentially adding features to the model in order of their importance score (i.e., from Eu/Eu* to Ho for the Random Forest model and from [Ce/Nd]/Y to Tm for the Deep Neural Network model). Changes in model accuracy as a function of the number of features incorporated are illustrated in Figures 5c and 5d. The results demonstrate that the accuracy of the models

(a) Random Forest

Actual \ Predicted	Barren	Fertile
Barren	0.897(TN)	0.103(FP)
Fertile	0.067(FN)	0.933(TP)

Accuracy = 0.92

(b) Deep Neural Network

Actual \ Predicted	Barren	Fertile
Barren	0.846(TN)	0.154(FP)
Fertile	0.054(FN)	0.946(TP)

Accuracy = 0.90

Figure 3. Normalized confusion matrix of the testing set to show the evaluation metrics of (a) the Random Forest and (b) the Deep Neural Network models. Abbreviations: TN, true negative; FP, false positive; FN, false negative; TP, true positive.

changes greatly up to 16 and 13 features, respectively, for the Random Forest model and Deep Neural Network models. After the models reach their greatest accuracy, adding more features to the algorithms did not increase their accuracy further, but rather slightly decreased it, indicating that the lower importance features do not have a significant impact on improving model accuracy (Figures 5c and 5d). For both models, the first 10 geochemical features have a cumulative importance of ~90%, with the trace-element ratios—Eu/Eu*, (Ce/Nd)/Y, Ce/Nd, and Ce/Ce*—having the greatest contribution to model accuracy and are of particular importance in characterizing magma fertility.

The importance of a feature in a sample may not be consistent with the entire data set. The importance of a feature to classifying a given sample can be assessed using the SHAP algorithm—the results are shown in Figure S1 in Supporting Information S1. Analysis of the scores demonstrates that high ratios of Eu/Eu*, Ce/Ce*, Ce/Nd, and Eu/Eu*/Y, and high concentrations of Eu have a positive impact on identifying fertile zircons, whereas the concentrations of Dy, Hf, and Ti are relatively low in fertile zircons. The importance of high Eu/Eu*, Ce/Nd, Ce/Ce*, and Eu/Eu*/Y ratios to identifying fertile rock is consistent with previous conclusions, which suggest that magmas with a high-water content and a high oxidation state are critical to the formation of porphyry Cu deposits (Y. Lu et al., 2016). Previous studies also demonstrated that such hydrous and fertile melts likely experienced early fractionation of amphibole, which preferentially concentrate the middle REE (e.g., Dy), depleting the residual melt in these elements (Davidson et al., 2007). This may be the main reason for the low Dy concentration in fertile zircons. Low Ti contents in fertile zircons are also consistent with the previous suggestion that they

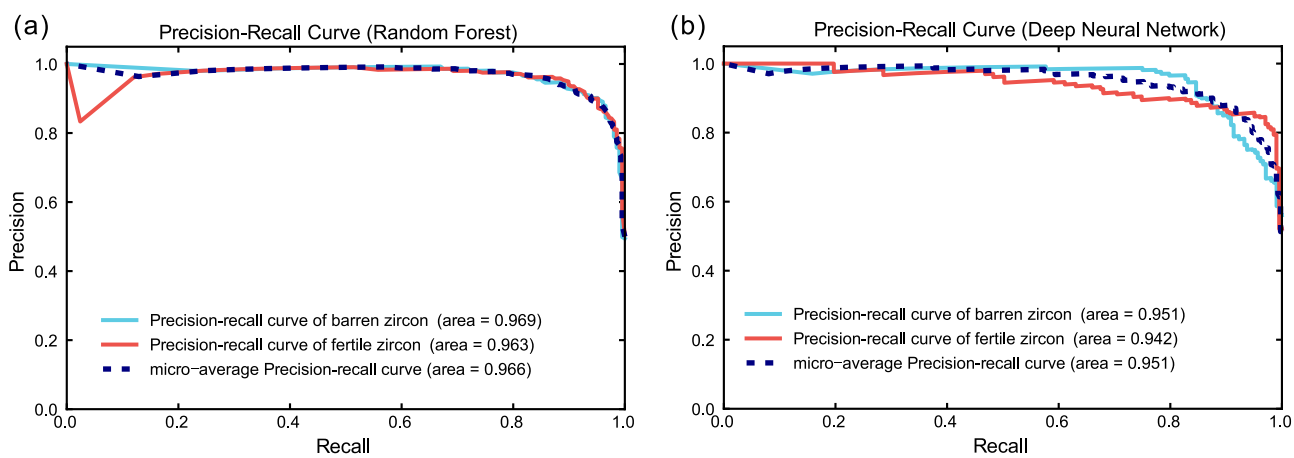

Figure 4. Precision–recall curve to show the performance of (a) the Random Forest and (b) the Deep Neural Network models.

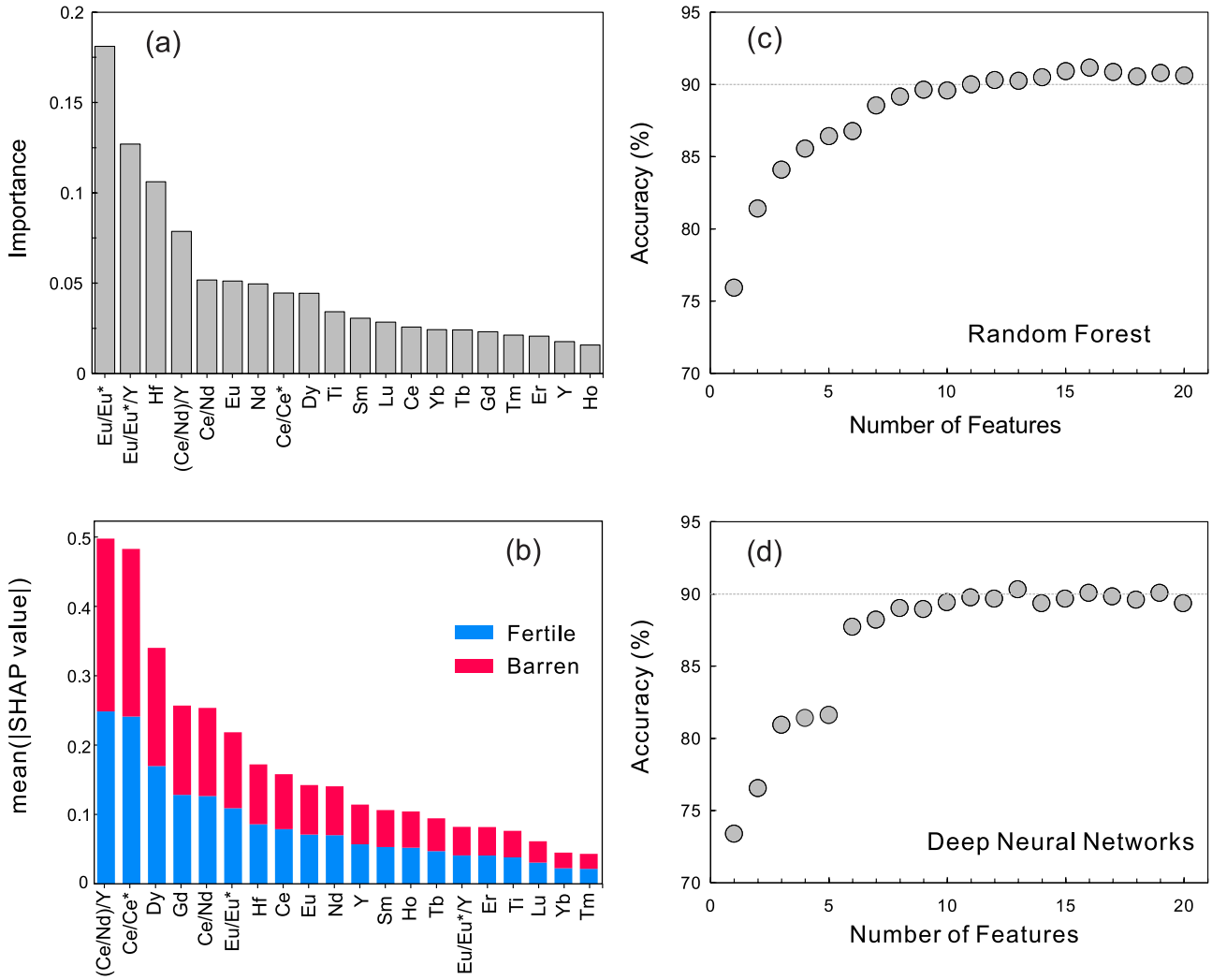

Figure 5. (a) Relative feature importance (sum equal to 1) for the individual geochemical features used in the training of the Random Forest algorithm. (b) Mean SHAP (SHapley Additive exPlanations) values for the individual geochemical features of the fertile and barren zircon utilized during the training of the deep neural network algorithm used in this study. (c) Accuracy of the Random Forest model with increasing number of features utilized in the algorithm. Starting with distance to the feature of Eu/Eu* and proceeding with features with next highest importance features. (d) Accuracy of the Deep Neural Network model with increasing number of features utilized in the algorithm. Starting with distance to the feature of (Ce/Nd)/Y and proceeding with features with next highest importance features.

crystallized from the parent magma with a relatively low temperature and a relatively long evolution process (Loucks, 2014).

The two machine learning models also revealed that low Hf contents in zircon have some influence on discriminating between fertile and barren zircons (Figure S1 in Supporting Information S1), which is inconsistent with previous conclusions. Pizarro et al. (2020) demonstrated that most zircon grains from fertile intrusions have a higher Hf concentrations (greater than 8,750 ppm) than those in barren intrusions based on a comparison between fertile and barren zircon from porphyry systems in northern Chile and southern Peru. This is not the case, however, if the global trace-element data set of zircon is considered. As illustrated in Figure S2 in Supporting Information S1, the concentration of Hf in most fertile zircons is relatively lower than that of barren zircon and exhibits significant overlap. This suggests that a relatively high Hf concentration of zircon cannot be utilized as a robust fertility indicator for porphyry Cu systems.

4.4. Model Validation

Porphyry deposits have been recognized in both subduction-related and postcollisional settings. Ore-forming magmas in both tectonic settings are thought to share some similar features, such as a high oxidation state, and high-water and volatile content (Haschke et al., 2010; Hou et al., 2015; Richards, 2009; R. Wang et al., 2020). Previous studies have demonstrated, however, that geochemical indicators of magma fertility in these tectonically distinct types of porphyry systems are different (cf., Leng et al., 2020). This highlights the inefficiency of traditional bivariate plots, which need to combine multiple fertility indicators and evaluate each indicator independently. In this study, to verify whether the machine learning models can be successfully applied to porphyry systems that occur in these two distinct geologic settings, two areas (one located in a subduction-related area in Canada and the other located in a postcollision-related area in China) were selected to assess mineralization potential.

Before validating the machine learning models, two parameters were constructed to assist in the classification of zircon grains as having crystallized from either barren or fertile magma. The *magma fertility percentage* (MFP) is defined as the proportion of zircon grains within a given sample classified by the models as having crystallized from fertile magma. The *magma fertility index* (MFI) is defined as MFI = (MFP – 0.5)/0.5 and is used to identify samples that have greater than 50% fertile zircon grains and, hence, are classified in the test as fertile rocks. In other words, a sample with a positive MFI value crystallized from a magma with a high mineralization potential, whereas a sample with a negative MFI value crystallized from a magma with limited mineralization potential. In the following examples, we first classify zircon grains using the trained machine learning models and then calculate the MFI values of different samples based on the outputs.

4.4.1. The Highland Valley Porphyry Cu District, Canada

The Highland Valley district, located within the Intermontane Belt in British Columbia, Canada, contains several large, subduction-related porphyry Cu–Mo deposits and more than 160 prospects (Byrne et al., 2017), all of which are hosted by the Late Triassic Guichon Creek Batholith (GCB). The intrusive rocks were emplaced into volcanic and sedimentary rocks of the 238–202 Ma Nicola Group, which is mainly composed of volcanic and sedimentary rocks, with island-arc affinities (D'Angelo, 2016; Mihalynuk et al., 2016). Three intrusive suites have been recognized in this district—the premineralized Gump Lake stock (ca. 218 Ma), the mineralized GCB (ca. 211–208 Ma), and the postmineralization dikes (ca. 208–206 Ma; Byrne et al., 2017; D'Angelo et al., 2017; Lee, 2021; Lee et al., 2017). Based on cross-cutting relationships, and lithogeochemical and geochronological studies (D'Angelo et al., 2017; Schiarizza, 2014), the mineralized GCB is interpreted to have crystallized from three cogenetic magma pulses represented by (a) the early Border and Highland Valley facies (including the Guichon, Guichon–Chataway transition, and Chataway subfacies; ca. 211–210.5 Ma), (b) the intermediate Bethlehem facies, and the syn-mineralization dikes and stocks (ca. 209.5–209 Ma), and (c) the late Skeena and Bethsaida facies (208.8–208.5 Ma). The syn-mineralization dikes and stocks include the Bethlehem porphyry, Jersey stock, and feldspar–quartz-rich-crowded porphyry dike. The postmineralization dikes, which are small volume dikes emplaced into the GCB, consist of the "S&P" Bethsaida porphyry and quartz–feldspar phyric, quartz-rich porphyry dike. The geological framework and evolution of the igneous rocks has been characterized by D'Angelo et al. (2017), and the chemistry of zircon grains hosted by the igneous rocks was reported by Lee (2021). Previous studies demonstrated that the mineralization event occurred at ca. 208 Ma (molybdenite Re–Os age; D'Angelo et al., 2017). The multiple magmatic events that occurred in the Highland Valley porphyry Cu district makes it an ideal location to assess the accuracy of the constructed machine learning models.

The fertility of all of the intrusive units in the Highland Valley district was predicted utilizing both the Random Forest and the Deep Neural Network models constructed in this study. As illustrated in Figure 6, fertility predictions given by both models are generally consistent with each other, demonstrating that the same input features result in similar predictions (e.g., Zhao et al., 2019). Overall, both methods predict positive MFI values for the early, intermediate, and late facies of the mineralized GCB, except for the Guichon and Guichon–Chataway transition subfacies. In contrast, negative MFI values are predicted for the premineralization Gump Lake stock and postmineralization dikes (Figure 6). In addition, from the intermediate facies of the GCB to the syn-mineralization dikes to the late facies of the GCB, the predicted MFI value gradually increases to a value of 1 (Figure 6), suggesting that (a) the mineralization potential of the magma increased gradually and (b) the main and late facies of the GCB have the highest mineralization potential, consistent with the fact that they are the main hosts of mineralization (D'Angelo et al., 2017).

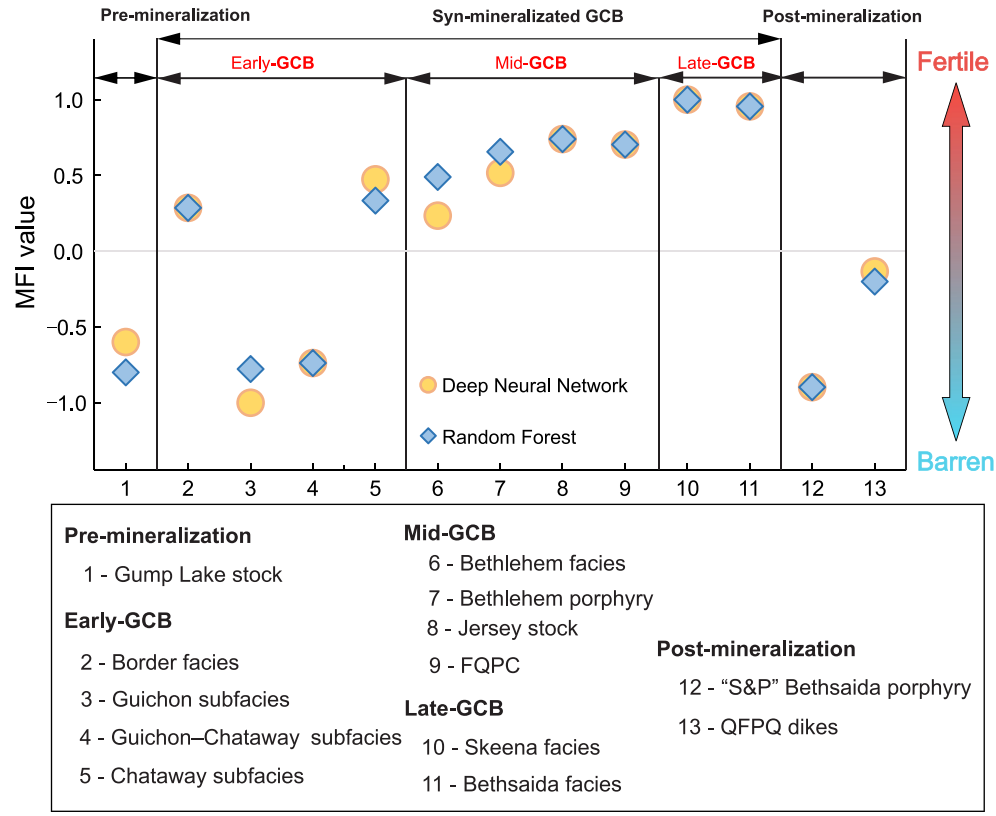

Figure 6. The magma fertility index (MFI) values of zircon in igneous rocks from different phases of the intrusive suites in the Highland Valley porphyry Cu district, Canada.

The Guichon and Guichon–Chataway transition subfacies, as the part of the early facies of the GCB, exhibit negative MFI values based on the predictions of both models (Figure 6). D'Angelo et al. (2017) demonstrated that all facies from the GCB were derived from the same magma chamber, and most likely the same mantle source, and formed from three pulses of magma emplacement. The earliest magmatic pulse formed the early facies of the GCB (Border, Guichon, Guichon–Chataway transition, and Chataway facies). Given that some samples from this pulse have elevated $^{87}Sr/^{86}Sr_i$ ratios (especially the Guichon facies; D'Angelo et al., 2017), it was suggested that the Guichon and Guichon–Chataway transition subfacies formed as a result of contamination by limestone and/or were previously contaminated by Nicola Group basalt country rocks, which lead to their lack of mineralization and negative MFI values. All of the prediction results are consistent with the conclusions of previous studies based on field observations and lithogeochemical analyses (e.g., D'Angelo et al., 2017; Lee, 2021). Therefore, we are optimistic that the machine learning models constructed here accurately and reliably differentiate between fertile and barren rocks in subduction-related porphyry systems and also can be used as a tool to identify the genetic evolution of a crystallizing magma chamber and the relation of this evolution to mineralization.

4.4.2. The Southern Gangdese Belt, Tibet, China

The Gangdese belt is a well-known porphyry Cu belt in southern Tibet, China. It experienced subduction of the Neo-Tethyan oceanic slab from the Late Triassic to the Paleocene (Yin & Harrison, 2000; D.-C. Zhu et al., 2011, 2015). During this process, voluminous calc-alkaline and high-K calc-alkaline granitic rocks and volcanic rocks were emplaced, with the former representing the Gangdese batholith (Chen et al., 2019; Chu et al., 2006; Ji et al., 2009; R. Wang et al., 2017). The closure of the Neo-Tethys Ocean, and collision between India and Eurasia are generally considered to have occurred at 55–50 Ma (Chemenda et al., 2000; Replumaz et al., 2010; D.-C. Zhu et al., 2015). Since the early stages of collision, volcanic rocks of the Paleocene Linzizong erupted and large volumes of the Paleocene–Eocene syn-collisional granitoids were emplaced (Mo et al., 2005). Subsequently, magmatic rocks occurred as Oligocene–Miocene adakite-like intrusions (33–12 Ma) and Miocene ultrapotassic–potassic igneous rocks (25–8 Ma; R. Wang et al., 2018; Yang et al., 2016). The collision-related

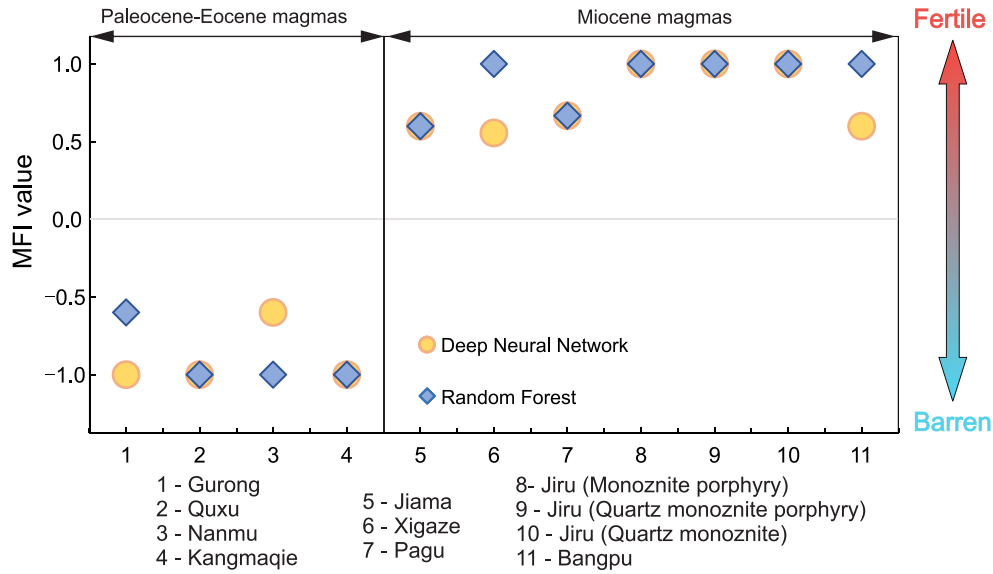

Figure 7. The magma fertility index (MFI) values of zircon in the Cenozoic igneous rocks from the southern Gangdese belt, China.

Paleocene–Eocene intrusive rocks are mostly barren, with few economic porphyry deposits, whereas the postcollision-related Miocene rocks are fertile and host several porphyry Cu–Mo ± Au deposits, such as the giant Qulong porphyry Cu–Mo deposit, the Jiama porphyry Cu–Mo–Au deposit, and the large- to small-sized Tingong, Chongjiang, Gangjiang, Bairong, Nanmu, and Lakang'e porphyry Cu–Mo deposits (Sun et al., 2020; R. Wang, Richards, Hou, Yang, & DuFrane, 2014; R. Wang, Richards, Hou, Yang, et al., 2014; Yang & Cooke, 2019). Based on whole-rock geochemistry, mineral chemistry, and zircon trace-element and isotope compositions of the Paleocene–Miocene rocks, previous studies have demonstrated that the Miocene magmas had higher water and sulfur contents, and higher oxidization states than the earlier Paleocene–Eocene magmas (Sun et al., 2020; R. Wang, Richards, Hou, Yang, & DuFrane, 2014; R. Wang, Richards, Hou, Yang, et al., 2014). Accordingly, these studies suggested that the Miocene magmas would have had a greater potential for the formation of porphyry Cu deposits.

As illustrated in Figure 7, predictions made using the Random Forest and Deep Neural Network models are generally consistent with each other, with zircons from the Paleocene–Eocene rocks having systematically negative MFI values, and zircons from the Miocene rocks having systematically positive MFI values. The model results demonstrate that most of the zircon in the Paleocene–Eocene rocks are classified as belonging to barren rock, while most of the zircon in the Miocene rocks are classified as belonging to fertile rock. This suggests that the Miocene magmas in the southern Gangdese belt had a higher mineralization potential than the Paleocene–Eocene magmas, consistent with the interpretations of previous studies. The constructed machine learning models, therefore, have the potential to be applied to robustly evaluate the fertility of magmas in both subduction and postcollisional settings.

4.5. Machine Learning-Based Predictors for Characterizing Magma Fertility

As exploration for, and consumption of, mineral resources increase, fewer outcropping or near-surface deposits will remain to be discovered and/or exploited. Mining companies require robust, accessible, and cost-effective exploration strategies to reduce the size of prospecting regions, minimize the time and monetary investment required for exploration, and improve the likelihood of identifying mineralized systems. In the case of porphyry Cu deposits, previous studies have highlighted a number of lithogeochemical and mineralogical indicators that can be used to assess the fertility of igneous rock suites, including whole-rock Al_2O_3/TiO_2, Sr/Y, La/Yb, Dy/Yb, V/Sc, and Sr/MnO (cf., Leng et al., 2020; Loucks, 2014; Wells et al., 2020), and ratios of Eu/Eu*, Eu/Eu*/Y, (Ce/Nd)/Y, Ce/Nd, and Ce/Ce* in zircon (Y. Lu et al., 2016). Although these indicators have made some progress toward this goal, significant limitations of these individual geochemical indicators are becoming increasingly

apparent. First, the composition of mineralized rocks and accessory minerals in these rocks varies widely in porphyry systems, such that a geochemical indicator that accurately identifies fertile rock in one system will not necessarily be accurate in another. In addition, to more accurately assess magma fertility, traditional methods need to combine multiple fertility indicators and evaluate each indicator independently (e.g., Chen, et al., 2019; Wells et al., 2020). This approach is less efficient, however, when compared to machine learning-based approaches that can integrate all zircon trace-element parameters simultaneously when being trained to discriminate fertile and barren suites. More importantly, the machine learning models have higher accuracies (0.90–0.93 for the Random Forest model and 0.85–0.95 for the Deep Neural Network model) and lower false positive rates (0.10 for Random Forest model and 0.15 for the Deep Neural Network model) than that of traditional zircon geochemical indicators (accuracy of 0.65–0.83 and false positive rate of 0.23–0.66). The 10%–30% increase in accuracy and 10%–50% decrease in false positive greatly improve the success ratios during prospecting and exploration and save time and resources.

Machine learning can directly capture the relationships between geochemical data and magma fertility without any a priori knowledge of the exploration region. That being said, the data-based machine learning model developed here may be biased and lead to few false positive observations. First, the current data set may not be entirely representative because the fertile data set used here comprises known porphyry Cu deposits with published zircon trace-element data. Second, elemental concentrations and ratios in zircon are assumed equally important in our constructed model. However, the concentration of some elements in zircon is almost indistinguishable among fertile and barren rock in the same region (e.g., Hf, Dilles et al., 2015), indicating that they are not as important as the other elements in distinguishing magma fertility. Regardless, both models have captured the geochemical characteristics of zircon in fertile rock, including the high ratios of Eu/Eu*, Ce/Nd, Ce/Ce*, and Eu/Eu*/Y, and the low contents of Dy, Hf, and Ti; these results are largely consistent with previous demonstrations that these signatures are important for distinguishing fertile and barren suites (Dilles et al., 2015; Loucks, 2014; Y. Lu et al., 2016; Shu et al., 2019). In addition, the machine learning predictions made for both the subduction-related Highland Valley district and the postcollision-related Gangdese belt are in good agreement with the actual observations of mineralization. Furthermore, as increasingly more geochemical data of zircon in porphyry ore systems become publicly available, the trained machine learning models will become more intelligent, and their performance will continuously improve. We are, therefore, optimistic that this machine learning approach is robust in identifying porphyry Cu mineralization in a range of tectonic environments.

4.6. Implications

The constructed Random Forest and Deep Neural Network models show comparably strong performance and can discriminate zircon from igneous rocks associated with porphyry Cu deposits from those in barren igneous systems with high accuracy and independent of geological setting. This suggests that, when combined with conventional geological, geophysical, and geochemical methods, machine learning has the potential to provide efficient and economic means to evaluate magma fertility in porphyry districts in the early stages of prospecting. More importantly, it is likely that this machine learning-based approach can be applied to other minerals, such as apatite, titanite, quartz, and magnetite, to characterize fertile and barren magmas in other magmatic–hydrothermal systems. Although the real-world applicability of our models is assessed by testing them on two well-characterized porphyry Cu–Mo deposits in different tectonic settings, it does not mean that they are only applicable to these two districts. On the contrary, more examples and data will make the model more sophisticated. Machine learning models trained with different target variables may be applied to more fields of solid Earth science, such as understanding other different magma systems.

In summary, this contribution demonstrates that machine learning methods can capture well the relationship between 20-dimensional geochemical data and magma fertility in porphyry districts and make robust predictions on the location of mineralization. The results highlight the great potential of machine learning in analyzing and processing high-dimensional data in solid Earth sciences, which will help to further understand patterns behind these data. Nevertheless, caution is still required—machine learning models should not be naively applied to assess complex problems in the geosciences prior to evaluating the complexities in machine learning methods that may skew predictions (e.g., biased data, overfitting).

5. Conclusions

This contribution demonstrates the potential of using the trace-element chemistry of zircon to construct machine learning models (Random Forest and Deep Neural Network algorithms) that can effectively and rapidly evaluate the fertility of magmas in exploration regions targeting porphyry Cu systems. Compared to traditional magma fertility indicators, this approach improves the efficiency of prospect evaluation and reduces the likelihood of false positive observations that may be caused by differences in tectonic setting. Feature importance analysis of the classification models reveals that, in addition to the previously proposed fertility indicators (i.e., high ratios of Eu/Eu*, Ce/Nd, Ce/Ce*, and Eu/Eu*/Y), low concentrations of Dy, Hf, and Ti in zircon are also important indices to distinguish between fertile and barren rock. This contribution demonstrates that the machine learning technique is a promising tool for decoding hidden information in high-dimensional geochemical data, promoting its application to development of exploration tools centered on other minerals and other mineral deposit types.

Data Availability Statement

The data sets and the code for the machine learning models developed in this study are available at http://doi.org/10.5281/zenodo.6459709; the trained models can also be found at https://github.com/shaohaozou/magma-fertility, which will be updated as more geochemical data of zircon in porphyry ore systems becomes publicly available. Further descriptions regarding the construction of the Deep Neural Network and Random Forest models are available through the Keras (https://keras.io/api/) and Scikit-Learn (https://scikit-learn.org/stable/index.html) documentation, respectively.

Acknowledgments
This work was cofunded by the National Natural Science Foundation of China (Nos. 42002089, 42102095, 41930428, and 42022021), the Provincial Science and Technology Project of Jiangxi Province (20202BABL214056), and the DHBK project from East China University of Technology (Nos. DHBK2019320 and DHBK2019317). The authors would like to thank Editor Isabelle Manighetti, an anonymous associate editor, and three anonymous reviewers for their constructive suggestions and thoughtful comments that greatly improved the quality and impact of this study.

References

Aitchison, J. (1986). The statistical analysis of geochemical compositions. *Journal of the International Association for Mathematical Geology*, *16*(6), 531–564. https://doi.org/10.1007/BF01029316

Ali, S. H., Giurco, D., Arndt, N., Nickless, E., Brown, G., Demetriades, A., et al. (2017). Mineral supply for sustainable development requires resource governance. *Nature*, *543*(7645), 367–372. https://doi.org/10.1038/nature21359

Asadi, S., Moore, F., & Zarasvandi, A. (2014). Discriminating productive and barren porphyry copper deposits in the southeastern part of the central Iranian volcano-plutonic belt, Kerman region, Iran: A review. *Earth-Science Reviews*, *138*, 25–46. https://doi.org/10.1016/j.earscirev.2014.08.001

Baldwin, J. A., & Pearce, J. A. (1982). Discrimination of productive and nonproductive porphyritic intrusions in the Chilean Andes. *Economic Geology*, *77*(3), 664–674. https://doi.org/10.2113/gsecongeo.77.3.664

Ballard, J. R., Palin, M. J., & Campbell, I. H. (2002). Relative oxidation states of magmas inferred from Ce(IV)/Ce(III) in zircon: Application to porphyry copper deposits of northern Chile. *Contributions to Mineralogy and Petrology*, *144*(3), 347–364. https://doi.org/10.1007/s00410-002-0402-5

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brugger, J., Pring, A., Reith, F., Ryan, C., Etschmann, B., Liu, W., et al. (2010). Probing ore deposits formation: New insights and challenges from synchrotron and neutron studies. *Radiation Physics and Chemistry*, *79*(2), 151–161. https://doi.org/10.1016/j.radphyschem.2009.03.071

Buret, Y., von Quadt, A., Heinrich, C., Selby, D., Wälle, M., & Peytcheva, I. (2016). From a long-lived upper-crustal magma chamber to rapid porphyry copper emplacement: Reading the geochemistry of zircon crystals at Bajo de la Alumbrera (NW Argentina). *Earth and Planetary Science Letters*, *450*, 120–131. https://doi.org/10.1016/j.epsl.2016.06.017

Byrne, K., Lesage, G., Gleeson, S. A., & Lee, R. G. (2017). Large-scale sodic–calcific alteration around porphyry copper systems: Examples from the Highland Valley copper district, Guichon batholith, south-central British Columbia. *Geoscience BC Summary of Activities*, *2017*(1), 213–222.

Campbell, I. H., Ballard, J. R., Palin, J. M., Allen, C., & Faunes, A. (2006). U–Pb zircon geochronology of granitic rocks from the Chuquicamata–El Abra porphyry copper belt of northern Chile: Excimer laser ablation ICP-MS analysis. *Economic Geology*, *101*(7), 1327–1344. https://doi.org/10.2113/gsecongeo.101.7.1327

Chayes, F. (1960). On correlation between variables of constant sum. *Journal of Geophysical Research*, *65*(12), 4185–4193. https://doi.org/10.1029/JZ065i012p04185

Chelle-Michou, C., Chiaradia, M., Béguelin, P., & Ulianov, A. (2015). Petrological evolution of the magmatic suite associated with the Coroccohuayco Cu (–Au–Fe) porphyry–skarn deposit, Peru. *Journal of Petrology*, *56*(9), 1829–1862. https://doi.org/10.1093/petrology/egv056

Chelle-Michou, C., Chiaradia, M., Ovtcharova, M., Ulianov, A., & Wotzlaw, J.-F. (2014). Zircon petrochronology reveals the temporal link between porphyry systems and the magmatic evolution of their hidden plutonic roots (the Eocene Coroccohuayco deposit, Peru). *Lithos*, *198*, 129–140. https://doi.org/10.1016/j.lithos.2014.03.017

Chemenda, A. I., Burg, J.-P., & Mattauer, M. (2000). Evolutionary model of the Himalaya–Tibet system: Geopoem: Based on new modelling, geological and geophysical data. *Earth and Planetary Science Letters*, *174*(3–4), 397–409. https://doi.org/10.1016/S0012-821X(99)00277-0

Chen, X., Richards, J. P., Liang, H., Zou, Y., Zhang, J., Huang, W., et al. (2019). Contrasting arc magma fertilities in the Gangdese belt, southern Tibet: Evidence from geochemical variations of Jurassic volcanic rocks. *Lithos*, *324–325*, 789–802. https://doi.org/10.1016/j.lithos.2018.12.008

Cherniak, D. J., & Watson, E. B. (2003). Diffusion in zircon. *Reviews in Mineralogy and Geochemistry*, *53*(1), 113–143. https://doi.org/10.2113/0530113

Chu, M.-F., Chung, S.-L., Song, B., Liu, D., O'Reilly, S. Y., Pearson, N. J., et al. (2006). Zircon U–Pb and Hf isotope constraints on the Mesozoic tectonics and crustal evolution of southern Tibet. *Geology*, *34*(9), 745–748. https://doi.org/10.1130/G22725.1

Claiborne, L. L., Miller, C. F., Flanagan, D. M., Clynne, M. A., & Wooden, J. L. (2010). Zircon reveals protracted magma storage and recycling beneath Mount St. Helens. *Geology*, *38*(11), 1011–1014. https://doi.org/10.1130/G31285.1

Cooke, D. R., Hollings, P., & Walshe, J. L. (2005). Giant porphyry deposits: Characteristics, distribution, and tectonic controls. *Economic Geology*, *100*(5), 801–818. https://doi.org/10.2113/gsecongeo.100.5.801

Correa, K. J., Rabbia, O. M., Hernández, L. B., Selby, D., & Astengo, M. (2016). The timing of magmatism and ore formation in the El Abra porphyry copper deposit, northern Chile: Implications for long-lived multiple-event magmatic-hydrothermal porphyry systems. *Economic Geology*, *111*(1), 1–28. https://doi.org/10.2113/econgeo.111.1.1

D'Angelo, M. (2016). *Geochemistry, petrography and mineral chemistry of the Guichon Creek and Nicola batholiths, south-central British Columbia* (unpublished M.Sc. thesis). Lakehead University.

D'Angelo, M., Miguel, A., Hollings, P., Byrne, K., Piercey, S., & Creaser, R. A. (2017). Petrogenesis and magmatic evolution of the Guichon Creek batholith: Highland Valley porphyry Cu ± (Mo) district, south-central British Columbia. *Economic Geology*, *112*(8), 1857–1888. https://doi.org/10.5382/econgeo.2017.4532

Davidson, J., Turner, S., Handley, H., Macpherson, C., & Dosseto, A. (2007). Amphibole "sponge" in arc crust? *Geology*, *35*(9), 787–790. https://doi.org/10.1130/G23637A.1

Dilles, J. H., Kent, A. J. R., Wooden, J. L., Tosdal, R. M., Koleszar, A., Lee, R. G., & Farmer, L. P. (2015). Zircon compositional evidence for sulfur-degassing from ore-forming arc magmas. *Economic Geology*, *110*(1), 241–251. https://doi.org/10.2113/econgeo.110.1.241

Fiorentini, M. L., & Garwin, S. L. (2010). Evidence of a mantle contribution in the genesis of magmatic rocks from the Neogene Batu Hijau district in the Sunda Arc, South Western Sumbawa, Indonesia. *Contributions to Mineralogy and Petrology*, *159*(6), 819–837. https://doi.org/10.1007/s00410-009-0457-7

Haschke, M., Ahmadian, J., Murata, M., & McDonald, I. (2010). Copper mineralization prevented by arc-root delamination during Alpine–Himalayan collision in central Iran. *Economic Geology*, *105*(4), 855–865. https://doi.org/10.2113/gsecongeo.105.4.855

Hezarkhani, A. (2006). Petrology of the intrusive rocks within the Sungun porphyry copper deposit, Azerbaijan, Iran. *Journal of Asian Earth Sciences*, *27*(3), 326–340. https://doi.org/10.1016/j.jseaes.2005.04.005

Holliday, J. R., & Cooke, D. R. (2007). Advances in geological models and exploration methods for copper ± gold porphyry deposits. *Proceedings of Exploration*, *7*, 791–809.

Hoskin, P. W. O., & Schaltegger, U. (2003). The composition of zircon and igneous and metamorphic petrogenesis. *Reviews in Mineralogy and Geochemistry*, *53*(1), 27–62. https://doi.org/10.2113/0530027

Hou, Z., Yang, Z., Lu, Y., Kemp, A., Zheng, Y., Li, Q., et al. (2015). A genetic linkage between subduction- and collision-related porphyry Cu deposits in continental collision zones. *Geology*, *43*(3), 247–250. https://doi.org/10.1130/G36362.1

Ji, W.-Q., Wu, F.-Y., Chung, S.-L., Li, J.-X., & Liu, C.-Z. (2009). Zircon U–Pb geochronology and Hf isotopic constraints on petrogenesis of the Gangdese batholith, southern Tibet. *Chemical Geology*, *262*(3–4), 229–245. https://doi.org/10.1016/j.chemgeo.2009.01.020

Jiang, Y.-H., Jiang, S.-Y., Ling, H.-F., & Dai, B.-Z. (2006). Low-degree melting of a metasomatized lithospheric mantle for the origin of Cenozoic Yulong monzogranite-porphyry, east Tibet: Geochemical and Sr–Nd–Pb–Hf isotopic constraints. *Earth and Planetary Science Letters*, *241*(3–4), 617–633. https://doi.org/10.1016/j.epsl.2005.11.023

Kingma, D. P., & Ba, J. (2014). ADAM: A method for stochastic optimization. *ArXiv Preprint ArXiv:1412.6980*.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection (pp. 1137–1145). Paper presented at the International Joint Conference on Artificial Intelligence.

Kuwatani, T., Nagata, K., Okada, M., Watanabe, T., Ogawa, Y., Komai, T., & Tsuchiya, N. (2015). Machine-learning techniques for geochemical discrimination of 2011 Tohoku tsunami deposits. *Scientific Reports*, *4*(1), 7077. https://doi.org/10.1038/srep07077

Lang, J. R., & Titley, S. R. (1998). Isotopic and geochemical characteristics of Laramide magmatic systems in Arizona and implications for the genesis of porphyry copper deposits. *Economic Geology*, *93*(2), 138–170. https://doi.org/10.2113/gsecongeo.93.2.138

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lee, R. G. (2021). Using zircon trace element composition to assess porphyry copper potential of the Guichon Creek batholith and Highland Valley copper deposit, south-central British Columbia. *Mineralium Deposita*, *24*, 215–238. https://doi.org/10.1007/s00126-020-00961-1

Lee, R. G., Dilles, J. H., Tosdal, R. M., Wooden, J. L., & Mazdab, F. K. (2017). Magmatic evolution of granodiorite intrusions at the El Salvador porphyry copper deposit, Chile, based on trace element composition and U/Pb age of zircons. *Economic Geology*, *30*, 245. https://doi.org/10.2113/econgeo.112.2.245

Leng, C.-B., Chen, X., Zhang, J., Ma, X., Tian, F., Guo, J., & Zhang, L. (2020). Lithogeochemical and mineral chemical footprints of porphyry Cu ± Mo ± Au deposits: A review. *Acta Geologica Sinica*, *94*(11), 3189–3212. https://doi.org/10.19762/j.cnki.dizhixuebao.2020200

Li, J., Qin, K., Li, G., Cao, M., Xiao, B., Chen, L., et al. (2012). Petrogenesis and thermal history of the Yulong porphyry copper deposit, eastern Tibet: Insights from U–Pb and U–Th/He dating, and zircon Hf isotope and trace element analysis. *Mineralogy and Petrology*, *105*(3), 201–221. https://doi.org/10.1007/s00710-012-0211-0

Liu, Y., Hu, Z., Gao, S., Günther, D., Xu, J., Gao, C., & Chen, H. (2008). In situ analysis of major and trace elements of anhydrous minerals by LA-ICP-MS without applying an internal standard. *Chemical Geology*, *257*(1–2), 34–43. https://doi.org/10.1016/j.chemgeo.2008.08.004

Loader, M. A., Nathwani, C. L., Wilkinson, J. J., & Armstrong, R. N. (2022). Controls on the magnitude of Ce anomalies in zircon. *Geochimica et Cosmochimica Acta*, *42*, 242–257. https://doi.org/10.1016/j.gca.2022.03.024

Loader, M. A., Wilkinson, J. J., & Armstrong, R. N. (2017). The effect of titanite crystallisation on Eu and Ce anomalies in zircon and its implications for the assessment of porphyry Cu deposit fertility. *Earth and Planetary Science Letters*, *472*, 107–119. https://doi.org/10.1016/j.epsl.2017.05.010

Lösing, M., & Ebbing, J. (2021). Predicting geothermal heat flow in Antarctica with a machine learning approach. *Journal of Geophysical Research: Solid Earth*, *126*, e2020JB021499. https://doi.org/10.1029/2020JB021499

Loucks, R. R. (2014). Distinctive composition of copper-ore-forming arcmagmas. *Australian Journal of Earth Sciences*, *61*(1), 5–16. https://doi.org/10.1080/08120099.2013.865676

Lu, Y., Loucks, R., Fiorentini, M., McCuaig, T., Evans, N. J., Yang, Z.-M., et al. (2016). Zircon compositions as a pathfinder for porphyry Cu ± Mo ± Au deposits. *Society of Economic Geologists. Special Publications Series*, *19*, 329–347.

Lu, Y.-J., Loucks, R. R., Fiorentini, M. L., Yang, Z.-M., & Hou, Z.-Q. (2015). Fluid flux melting generated postcollisional high Sr/Y copper ore–forming water-rich magmas in Tibet. *Geology*, *43*(7), 583–586. https://doi.org/10.1130/G36734.1

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Proceedings of the 31st international conference on neural information processing systems* (pp. 4768–4777).

Lusty, P. A. J., & Gunn, A. G. (2015). Challenges to global mineral resource security and options for future supply. *Geological Society, London, Special Publications*, *393*(1), 265–276. https://doi.org/10.1144/SP393.13

Mihalynuk, M. G., Diakow, L. J., Friedman, R. M., & Logan, J. M. (2016). Chronology of southern Nicola arc stratigraphy and deformation. *Geological Fieldwork*, 2016–1.

Mo, X.-X., Dongn, G.-C., Zhao, Z.-D., Zhou, S., Wang, L.-L., Qiu, R.-Z., & Zhang, F.-Q. (2005). Spatial and temporal distribution and characteristics of granitoids in the Gangdese, Tibet and implication for crustal growth and evolution. *Geological Journal of China Universities*, *11*(3), 281–290.

Munoz, M., Charrier, R., Fanning, C. M., Maksaev, V., & Deckart, K. (2012). Zircon trace element and O-Hf isotope analyses of mineralized intrusions from El Teniente ore deposit, Chilean Andes: Constraints on the source and magmatic evolution of porphyry Cu–Mo related magmas. *Journal of Petrology*, *53*(6), 1091–1122. https://doi.org/10.1093/petrology/egs010

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. *Mineralium Deposita*. https://doi.org/10.1007/s00126-021-01086-9

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Petrelli, M., Caricchi, L., & Perugini, D. (2020). Machine learning thermo-barometry: Application to clinopyroxene-bearing magmas. *Journal of Geophysical Research: Solid Earth*, *125*, e2020JB020130. https://doi.org/10.1029/2020JB020130

Petrelli, M., & Perugini, D. (2016). Solving petrological problems through machine learning: The study case of tectonic discrimination using geochemical and isotopic data. *Contributions to Mineralogy and Petrology*, *171*(10), 81. https://doi.org/10.1007/s00410-016-1292-2

Pizarro, H., Campos, E., Bouzari, F., Rousse, S., Bissig, T., Gregoire, M., & Riquelme, R. (2020). Porphyry indicator zircons (PIZs): Application to exploration of porphyry copper deposits. *Ore Geology Reviews*, *19*, 103771. https://doi.org/10.1016/j.oregeorev.2020.103771

Porter, T. M. (2016). The geology, structure and mineralisation of the Oyu Tolgoi porphyry copper–gold–molybdenum deposits, Mongolia: A review. *Geoscience Frontiers*, *7*(3), 375–407. https://doi.org/10.1016/j.gsf.2015.08.003

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

Reid, M. R., Vazquez, J. A., & Schmitt, A. K. (2011). Zircon-scale insights into the history of a supervolcano, Bishop Tuff, Long Valley, California, with implications for the Ti-in-zircon geothermometer. *Contributions to Mineralogy and Petrology*, *161*(2), 293–311. https://doi.org/10.1007/s00410-010-0532-0

Replumaz, A., Negredo, A. M., Guillot, S., & Villaseñor, A. (2010). Multiple episodes of continental subduction during India/Asia convergence: Insight from seismic tomography and tectonic reconstruction. *Tectonophysics*, *483*(1–2), 125–134. https://doi.org/10.1016/j.tecto.2009.10.007

Rezvanbehbahani, S., Stearns, L. A., Kadivar, A., Walker, J. D., & van der Veen, C. J. (2017). Predicting the geothermal heat flux in Greenland: A machine learning approach. *Geophysical Research Letters*, *44*, 12271–12279. https://doi.org/10.1002/2017GL075661

Richards, J. P. (2009). Postsubduction porphyry Cu–Au and epithermal Au deposits: Products of remelting of subduction-modified lithosphere. *Geology*, *37*(3), 247–250. https://doi.org/10.1130/G25451A.1

Richards, J. P. (2011). High Sr/Y arc magmas and porphyry Cu ± Mo ± Au deposits: Just add water. *Economic Geology*, *106*(7), 1075–1081. https://doi.org/10.2113/econgeo.106.7.1075

Richards, J. P. (2015). The oxidation state, and sulfur and Cu contents of arc magmas: Implications for metallogeny. *Lithos*, *233*, 27–45. https://doi.org/10.1016/j.lithos.2014.12.011

Richards, J. P. (2016). Clues to hidden copper deposits. *Nature Geoscience*, *9*(3), 195–196. https://doi.org/10.1038/ngeo2656

Richards, J. P., & Kerrich, R. (2007). Special paper: Adakite-like rocks: Their diverse origins and questionable role in metallogenesis. *Economic Geology*, *102*(4), 537–576. https://doi.org/10.2113/gsecongeo.102.4.537

Richards, J. P., Spell, T., Rameh, E., Razique, A., & Fletcher, T. (2012). High Sr/Y magmas reflect arc maturity, high magmatic water content, and porphyry Cu ± Mo ± Au potential: Examples from the Tethyan arcs of central and eastern Iran and western Pakistan. *Economic Geology*, *107*(2), 295–332. https://doi.org/10.2113/econgeo.107.2.295

Schiarizza, P. (2014). Geological setting of the granite mountain batholith, south-central British Columbia. *Geological Fieldwork*, 2015–1.

Shen, P., Hattori, K., Pan, H., Jackson, S., & Seitmuratova, E. (2015). Oxidation condition and metal fertility of granitic magmas: Zircon trace-element data from porphyry Cu deposits in the Central Asian orogenic belt. *Economic Geology*, *110*(7), 1861–1878. https://doi.org/10.2113/econgeo.110.7.1861

Shore, J., & Johnson, R. (1980). Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Transactions on Information Theory*, *26*(1), 26–37. https://doi.org/10.1109/TIT.1980.1056144

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. *Mineralium Deposita*, *54*(5), 645–656. https://doi.org/10.1007/s00126-019-00867-7

Sillitoe, R. H. (2010). Porphyry copper systems. *Economic Geology*, *105*(1), 3–41. https://doi.org/10.2113/gsecongeo.105.1.3

Smythe, D. J., & Brenan, J. M. (2016). Magmatic oxygen fugacity estimated using zircon-melt partitioning of cerium. *Earth and Planetary Science Letters*, *453*, 260–266. https://doi.org/10.1016/j.epsl.2016.08.013

Sun, X., Lu, Y., Li, Q., & Li, R. (2020). A downgoing Indian lithosphere control on along-strike variability of porphyry mineralization in the Gangdese belt of southern Tibet. *Economic Geology*, *116*(1), 29–46. https://doi.org/10.5382/econgeo.4768

Wainwright, A. J., Tosdal, R. M., Wooden, J. L., Mazdab, F. K., & Friedman, R. M. (2011). U–Pb (zircon) and geochemical constraints on the age, origin, and evolution of Paleozoic arc magmas in the Oyu Tolgoi porphyry Cu–Au district, southern Mongolia. *Gondwana Research*, *19*(3), 764–787. https://doi.org/10.1016/j.gr.2010.11.012

Wang, G.-G., Ni, P., Yao, J., Wang, X.-L., Zhao, K.-D., Zhu, R.-Z., et al. (2015). The link between subduction-modified lithosphere and the giant Dexing porphyry copper deposit, South China: Constraints from high-Mg adakitic rocks. *Ore Geology Reviews*, *67*, 109–126. https://doi.org/10.1016/j.oregeorev.2014.12.004

Wang, R., Richards, J. P., Hou, Z., Yang, Z., & DuFrane, S. A. (2014). Increased magmatic water content—The key to oligo-Miocene porphyry Cu–Mo Au formation in the eastern Gangdese belt, Tibet. *Economic Geology*, *109*(5), 1315–1339. https://doi.org/10.2113/econgeo.109.5.1315

Wang, R., Richards, J. P., Hou, Z.-Q., Yang, Z.-M., Gou, Z.-B., & DuFrane, S. A. (2014). Increasing magmatic oxidation state from Paleocene to Miocene in the eastern Gangdese belt, Tibet: Implication for collision-related porphyry Cu–Mo Au mineralization. *Economic Geology*, *109*(7), 1943–1965. https://doi.org/10.2113/econgeo.109.7.1943

Wang, R., Tafti, R., Hou, Z., Shen, Z., Guo, N., Evans, N. J., et al. (2017). Across-arc geochemical variation in the Jurassic magmatic zone, southern Tibet: Implication for continental arc-related porphyry Cu Au mineralization. *Chemical Geology*, *451*, 116–134. https://doi.org/10.1016/j.chemgeo.2017.01.010

Wang, R., Weinberg, R. F., Collins, W. J., Richards, J. P., & Zhu, D. (2018). Origin of postcollisional magmas and formation of porphyry Cu deposits in southern Tibet. *Earth-Science Reviews*, *181*, 122–143. https://doi.org/10.1016/j.earscirev.2018.02.019

Wang, R., Zhu, D., Wang, Q., Hou, Z., Yang, Z., Zhao, Z., & Mo, X. (2020). Porphyry mineralization in the Tethyan orogen. *Science China Earth Sciences*, *63*, 2042–2067. https://doi.org/10.1007/s11430-019-9609-0

Wang, S., Zhou, T., Yuan, F., Fan, Y., Cooke, D. R., Zhang, L., et al. (2016). Geochemical characteristics of the Shujiadian Cu deposit related intrusion in Tongling: Petrogenesis and implications for the formation of porphyry Cu systems in the Middle–Lower Yangtze River Valley metallogenic belt, eastern China. *Lithos*, *252*, 185–199. https://doi.org/10.1016/j.lithos.2016.02.013

Wells, T. J., Meffre, S., Cooke, D. R., Steadman, J. A., & Hoye, J. L. (2020). Porphyry fertility in the Northparkes district: Indicators from whole-rock geochemistry. *Australian Journal of Earth Sciences*, *67*(5), 717–738. https://doi.org/10.1080/08120099.2020.1715477

Wilkinson, J. J. (2013). Triggers for the formation of porphyry ore deposits in magmatic arcs. *Nature Geoscience*, *6*(11), 917–925. https://doi.org/10.1038/ngeo1940

Williams, M., Schoneveld, L., Mao, Y., Klump, J., Gosses, J., Dalton, H., et al. (2020). Pyrolite: Python for geochemistry. *Journal of Open Source Software*, *5*, 2314. https://doi.org/10.21105/joss.02314

Wotzlaw, J.-F., Bindeman, I. N., Stern, R. A., D'Abzac, F.-X., & Schaltegger, U. (2015). Rapid heterogeneous assembly of multiple magma reservoirs prior to Yellowstone supereruptions. *Scientific Reports*, *5*(1), 1–10. https://doi.org/10.1038/srep14026

Xiao, B., Chen, H., Hollings, P., Han, J., Wang, Y., Yang, J., & Cai, K. (2017). Magmatic evolution of the Tuwu–Yandong porphyry Cu belt, NW China: Constraints from geochronology, geochemistry and Sr–Nd–Hf isotopes. *Gondwana Research*, *43*, 74–91. https://doi.org/10.1016/j.gr.2015.09.003

Yang, Z.-M., & Cooke, D. R. (2019). Porphyry copper deposits in China. In Z. Chang & R. J. Goldfarb (Eds.), *Mineral deposits of China* (Vol. 22, pp. 133–187). Society of Economic Geologists. Retrieved from https://www.segweb.org/store/SearchResults.aspx?Category=SP

Yang, Z.-M., Goldfarb, R., & Chang, Z.-S. (2016). Generation of postcollisional porphyry copper deposits in southern Tibet triggered by subduction of the Indian continental plate. In J. P. Richards (Ed.), *Tectonics and metallogeny of the Tethyan orogenic belt* (Vol. 19, pp. 279–300). Society of Economic Geologists. Retrieved from https://www.segweb.org/store/detail.aspx?id=EDOCSP19

Yin, A., & Harrison, T. M. (2000). Geologic evolution of the Himalayan–Tibetan orogen. *Annual Review of Earth and Planetary Sciences*, *28*(1), 211–280. https://doi.org/10.1146/annurev.earth.28.1.211

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. *Geophysical Research Letters*, *46*, 5234–5242. https://doi.org/10.1029/2019GL082322

Zhong, R., Deng, Y., & Yu, C. (2021). Multi-layer perceptron-based tectonic discrimination of basaltic rocks and an application on the Paleoproterozoic Xiong'er volcanic province in the North China Craton. *Computers & Geosciences*, *149*, 104717. https://doi.org/10.1016/j.cageo.2021.104717

Zhong, S., Seltmann, R., Qu, H., & Song, Y. (2019). Characterization of the zircon Ce anomaly for estimation of oxidation state of magmas: A revised Ce/Ce* method. *Mineralogy and Petrology*, *113*(6), 755–763. https://doi.org/10.1007/s00710-019-00682-y

Zhu, D.-C., Wang, Q., Zhao, Z.-D., Chung, S.-L., Cawood, P. A., Niu, Y., et al. (2015). Magmatic record of India–Asia collision. *Scientific Reports*, *5*(1), 1–9.

Zhu, D.-C., Zhao, Z.-D., Niu, Y., Mo, X.-X., Chung, S.-L., Hou, Z.-Q., et al. (2011). The Lhasa Terrane: Record of a microcontinent and its histories of drift and growth. *Earth and Planetary Science Letters*, *301*(1–2), 241–255. https://doi.org/10.1016/j.epsl.2010.11.005

Zhu, J.-J., Richards, J. P., Rees, C., Creaser, R., DuFrane, S. A., Locock, A., et al. (2018). Elevated magmatic sulfur and chlorine contents in ore-forming magmas at the Red Chris porphyry Cu–Au deposit, northern British Columbia, Canada. *Economic Geology*, *113*(5), 1047–1075. https://doi.org/10.5382/econgeo.2018.4581

Zou, S., Chen, X., Xu, D., Brzozowski, M. J., Lai, F., Bian, Y., et al. (2021). A machine learning approach to tracking crustal thickness variations in the eastern North China Craton. *Geoscience Frontiers*, *12*(5), 101195. https://doi.org/10.1016/j.gsf.2021.101195