

## Article

# Prediction of Au-Polymetallic Deposits Based on Spatial Multi-Layer Information Fusion by Random Forest Model in the Central Kunlun Area of Xinjiang, China

Yuepeng Zhang <sup>1</sup>, Xiaofeng Ye <sup>2</sup>, Shuyun Xie <sup>1,\*</sup>, Jianbiao Dong <sup>1</sup>, Oraphan Yaisamut <sup>1,3</sup> , Xuwei Zhou <sup>1</sup> and Xiaoying Zhou <sup>2</sup>

<sup>1</sup> State Key Laboratory of Geological Processes and Mineral Resources (GPMR), Faculty of Earth Sciences, China University of Geosciences, Wuhan 430074, China

<sup>2</sup> Geophysical and Geochemical Prospecting Group, Geological and Mineral Exploration and Development Bureau, Changji 831100, China

<sup>3</sup> Geotechnology Division, Department of Mineral Resources, Ministry of Natural Resources and Environment, 75/10 Rama VI Road, Ratchathewi, Bangkok 10400, Thailand

\* Correspondence: tinaxie@cug.edu.cn

**Abstract:** In recent years, there has been a growing emphasis on combining intelligent prospecting algorithms, such as random forest, with extensive geological and mineral data for the purpose of quantitatively predicting exploration geochemistry. This approach holds significant importance for enhancing the accuracy of target delineation. The central Kunlun area in Xinjiang possesses highly favorable ore-forming geological conditions, offering excellent prospects for mineral exploration. However, the depletion of shallow deposits coupled with a decade-long gap in geological exploration have presented increasing challenges in the quest to discover substantial metal resources. Consequently, there is now a severe shortage of reserve assets in the region, prompting an urgent need for the implementation of new theories, methods, and technologies in mineral resource investigation and evaluation efforts. The researchers used geological and regional geochemical data to construct a random forest metallogenic discriminant model for predicting the mineralization of gold polymetallic minerals in the central Kunlun area of Xinjiang and delineating the metallogenic target area. Two different sampling methods were compared to quantitatively predict gold polymetallic mineral resources. The results indicate that the selected training samples offer higher prediction accuracy and reliability by fully capturing the complex information of the original data. The random forest model using select training samples has valuable applications in metallogenic prospect prediction and potential division due to its ability to consider the actual exploration cost and identify small areas with high potential and a high proportion of ore. This study significantly improves prediction accuracy, reduces exploration risk, and expands the use of machine learning algorithms in mathematical geology in the central Kunlun area of Xinjiang.



**Citation:** Zhang, Y.; Ye, X.; Xie, S.; Dong, J.; Yaisamut, O.; Zhou, X.; Zhou, X. Prediction of Au-Polymetallic Deposits Based on Spatial Multi-Layer Information Fusion by Random Forest Model in the Central Kunlun Area of Xinjiang, China. *Minerals* **2023**, *13*, 1302. <https://doi.org/10.3390/min13101302>

Academic Editor: Paul Alexandre

Received: 17 July 2023

Revised: 25 September 2023

Accepted: 26 September 2023

Published: 8 October 2023

**Keywords:** spatial multi-information fusion; random forest; metallogenic prediction; central Kunlun; Xinjiang



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Mineral resources have specific properties, such as exploration risk, output concealment, non-renewability, uncertainty of cognition, and other factors [1]. Metallogenic prediction is mainly based on the study of metallogenic systems and deposit models. It summarizes the metallogenic regularity, comprehensively utilizes geological, geophysical, geochemical, and remote sensing techniques, and predicts the location, quantity, and quality of potential mineral resources through geological similarity analogy, statistical analysis, and other methods [2]. How to extract useful information from various sources of geological

information for comprehensive processing and analysis to achieve the purpose of mineral prediction has always been a problem discussed in the field of geochemical exploration.

The early metallogenic prediction model is mainly a comprehensive summary of the theory and method of metallogenic prediction. Since the 1970s, many geologists have carried out a lot of research and practice on the aspects of mining area evaluation, total resource evaluation, computer programs, and evaluation methods, and they have gradually formed the theory and method system of quantitative evaluation of mineral resources [3–9]. In the past 30 years, with the development of computers and GIS combined with traditional methods, spatial information fusion models have been widely used in mineral quantitative prediction. These spatial information fusion models can be divided into three categories: data-driven models [10–15], knowledge-driven models [16–21], and hybrid-driven models [22,23]. Keyan Xiao et al. [24–26] developed a mineral resources evaluation system with many functions on a MAPGIS platform, including various geological data processing and information extraction functions, as well as comprehensive prediction and analysis methods of mineral resources, and realized various algorithms such as the feature analysis method, evidence weight method, BP neural network method, and cluster analysis. It makes full use of various geological and mineral data accumulated in the past to effectively identify the existence of deposits to achieve the maximum prospecting effect with less investment. In addition, these evaluation and prediction methods are also widely used in the field of agricultural geochemistry [27].

In the field of geological science, especially in the direction of metallogenic prediction, domestic and foreign scholars have adopted a variety of machine-learning algorithms [28–30]. Among them, the decision tree algorithm (DT), random forest algorithm (RF), support vector machine (SVM), and artificial neural network (ANN) are the most widely used in geosciences [31–34]. Among them, DT-based algorithms need to estimate fewer parameters and are easy to apply. Therefore, they have a high degree of automation but are easily overshadowed by the tendency of data overfitting [35]. For this reason, it is gradually replaced by more advanced and simpler machine learning algorithms; the support vector machine of the kernel function method and the random forest of the ensemble tree method have become very effective methods in metallogenic prediction. For example, Shuyun Xie [36] applied the BP neural network and fuzzy evidence weight model to the metallogenic prediction of lead/zinc deposits in Guangxi, China. This study points out that the classification, delineation, and prediction accuracy of the BP neural network algorithm in the target area are higher than those of the fuzzy evidence weight model, which confirms the effectiveness of the BP neural network in the field of metallogenic prediction. Rodriguez et al. [37] applied an artificial neural network, regression tree, random forest, and support vector machine algorithms in machine learning algorithms to predict mineralization in the Rodalquilar (Spain) mining area, and the prediction results were compared and evaluated comprehensively.

The prediction of gold and polymetallic mineralization plays a crucial role in mineral exploration and resource assessment. Machine learning methods have gained significant attention in this field due to their ability to analyze complex geological data and predict mineral potential areas. Among various machine learning algorithms, the random forest algorithm has shown great efficacy in mineral resource prediction. This paper aims to comprehensively elucidate the necessity and advantages of selecting the random forest algorithm for predicting gold and polymetallic mineralization in the central Kunlun area of Xinjiang, China. The central Kunlun area in Xinjiang, China, has attracted attention due to its significant potential for gold and polymetallic mineralization. Accurate prediction of mineralization in this region is crucial for effective resource exploration and management. Traditional geostatistical methods have limitations in dealing with complex geological data. Therefore, the application of advanced machine learning algorithms, such as random forest, is imperative for improved mineralization prediction [38–42].

Over the past few years, the application of machine learning algorithms, specifically the random forest algorithm, has gained significant attention in the field of mineral prediction. Mineral prediction plays a crucial role in the exploration and assessment of mineral

resources, aiding in informed decision-making for mining projects. The random forest algorithm, belonging to the ensemble learning family, has shown great potential due to its ability to handle complex datasets and provide accurate predictions [36].

Several recent research papers have explored the application of the random forest algorithm in mineral prediction using various data sources, including remote sensing imagery, spectral indices, and terrain variables [43–45]. These studies have demonstrated the effectiveness of random forest in accurately identifying and mapping lithological compositions, hydrothermal alterations, and geological features related to mineral deposits. One notable aspect of the random forest algorithm is its capability to handle high-dimensional datasets and effectively select relevant features for mineral prediction. This allows researchers to extract valuable information from complex datasets, improving the accuracy of mineral predictions. Additionally, the adaptability of random forest to different types of mineral exploration data, such as Sentinel-2 data or ASTER data, further enhances its applicability in diverse geological settings. The recent advancements in random forest-based mineral prediction have also focused on addressing challenges encountered in traditional methods, such as overcoming the limitations of manual interpretation and reducing subjectivity in the analysis. By leveraging the power of machine learning and data-driven approaches, these studies have demonstrated the potential for improved mineral resource assessments and exploration strategies [43,44].

However, despite the promising results, there are still areas that require further investigation and refinement [36]. These include optimizing input feature selection, exploring domain-specific feature engineering techniques, and enhancing the interpretability of the random forest model. Additionally, the combination of random forest with other algorithms or integration with multi-source data could potentially enhance prediction accuracy, expanding the application of this algorithm in mineral prediction. The application of the random forest algorithm in mineral prediction has shown promising results in recent years. Its ability to handle complex datasets, accurately predict mineral compositions, and overcome limitations of traditional approaches has demonstrated its significant potential in advancing mineral exploration and resource assessment [43–46]. Continued research and development in this field will contribute to further improvements in prediction accuracy and the integration of advanced technologies for sustainable and efficient mineral resource management.

Random forest is an ensemble learning method that combines multiple decision trees to make predictions. It is known for its capability to handle high-dimensional data, consider feature interactions, and provide reliable predictions based on variable importance measures. These characteristics make it well suited for modeling complex geological systems and predicting mineral potential areas. To apply the random forest algorithm, a comprehensive dataset consisting of geological, geochemical, and geophysical attributes should be gathered. This dataset should include variables related to mineralization processes, lithology, alteration, structural features, as well as environmental factors. Proper data preprocessing steps, such as attribute selection, data normalization, and handling missing values, need to be performed to ensure the quality and integrity of the dataset [42–44].

One of the significant advantages of the random forest algorithm is its ability to assess the importance of input features. By evaluating the variable importance measures provided by the algorithm, geologists and mineral resource experts can gain insights into the key factors influencing gold and polymetallic mineralization in the central Kunlun area [42–44]. This information is invaluable for prioritizing exploration efforts and understanding the underlying geological processes.

In conclusion, the selection of the random forest algorithm for predicting gold and polymetallic mineralization in the central Kunlun area of Xinjiang, China, is necessary and advantageous. The algorithm's ability to handle complex geological data, assess feature importance, provide ensemble learning, and offer interpretability and visualization enhances the accuracy and applicability of mineralization prediction models [42,43]. By leveraging the random forest algorithm, mineral exploration and resource assessment

efforts in the central Kunlun area can be significantly improved, leading to more efficient and sustainable resource management practices.

After comprehensively comparing the sensitivity and accuracy of the above four algorithms in the delineation of the model parameters and the data scale of the selected scenic area, the researchers believe that the random forest algorithm is superior to the other three algorithms in terms of prediction accuracy and parameter sensitivity. Therefore, this paper selects the random forest algorithm to carry out metallogenic prediction and target delineation of mineral resources in the central Kunlun area of Xinjiang.

## 2. Regional Geology

### 2.1. Basic Geological Background

The study area is located in southern Qiemo County, Bayingolin Mongolian Autonomous Prefecture, in the Xinjiang Uygur Autonomous Region, adjacent to the Tibet Autonomous Region, within the junction zone of the Kunlun block and Bayankhara Plate (Figure 1). This region covers an area of 15,756 km<sup>2</sup>. The geotectonic units in the study area belong to two Grade-I tectonic units. The northern part of the study area is located in the Qin-Qi-Kun orogenic system, while the southern part is located in the Sanjiang orogenic system of Tibet [38]. There are five grade III tectonic units within this region. From north to south, these units are the Apar-Mengya ophiolitic mélange belt of the Arguin arc basin system, distributed in a small area in the northern part of the study area; the North Kunlun magmatic arc of the East Kunlun arc basin system; the subduction accretionary complex belt on the southern slope of the Eastern Kunlun; the Muztag–Xidashan–Buqingshan ophiolite mélange belt; and the Hoh Xil–Songpan foreland basin in the Bayankala massif [38]. In addition, two large fold and fault structural belts are located in the study area: the fold-fault zone of the Lower Carboniferous in Feiyunshan and the Triassic fault-fold belt in Yingshishan, Pingling [39]. This complex geological background allowed for the development of element enrichment and mineralization in the study area following the proto-Tethys back-arc basin extinction, the formation and extinction of the Paleo-Tethys Ocean, and the geological evolution associated with the uplift of the Qinghai–Tibet Plateau [40,41].

### 2.2. Metallogenic Belt

The study area spans three metallogenic belts from north to south, namely the Karamiran (compound gully arc belt) Au–Cu–Ag ore belt, Huangyangling (fold belt) Sb–Hg–Au–Cu ore belt, and Yunwuling (fold belt) Cu–Au ore belt [41].

#### 2.2.1. Karamiran (Compound Gully Arc Belt) Au–Cu–Ag Ore Belt

The Karamiran belt is approximately east–west-trending. The middle part of this belt is mainly Silurian in origin, while Carboniferous minerals are distributed in the northern and southern margins. The Silurian and Devonian systems within this belt were built from deep–semideep-sea terrestrial source clastic rocks sandwiched with basalt, carbonate, and siliceous rocks. This belt mainly contains gold and silver ore. The main types of mineralization associated with this belt include ophiolite-type chromium, asbestos ore, tough shear-zone crushing, and altering rock-type gold mines, continental sedimentary coal mines, copper-bearing sandstone, and rock salt mines [38]. However, the Cr, Cu, and Ag deposits are distributed outside the study area, and only Au deposits and a large number of rock salt deposits are found within the study area [41].

#### 2.2.2. Huangyangling (Fold Belt) Sb–Hg–Au–Cu Ore Belt

The Huangyangling belt is located in the eastern section of the Muzi mineralization belt and is structurally a continental marginal active zone. In this belt, mineralization mainly involves antimony and mercury, while copper and gold mineralization has also been recorded. After a preliminary exploration, a Wollongong–Huangyangling–Changshangou antimony mercury ore belt was demarcated, and the Huangyangling antimony deposit, Wollongong antimony deposit, and Changshangou mercury deposit were discovered along

with several antimony and mercury ore sites. This belt has great prospects and is one of the major antimony ore prospect areas in Xinjiang. Although Sb and Hg deposits are present in this mineralization belt, they are distributed outside the study area, and only Au occurrence and placer gold occurrence are located within the study area [38–41].

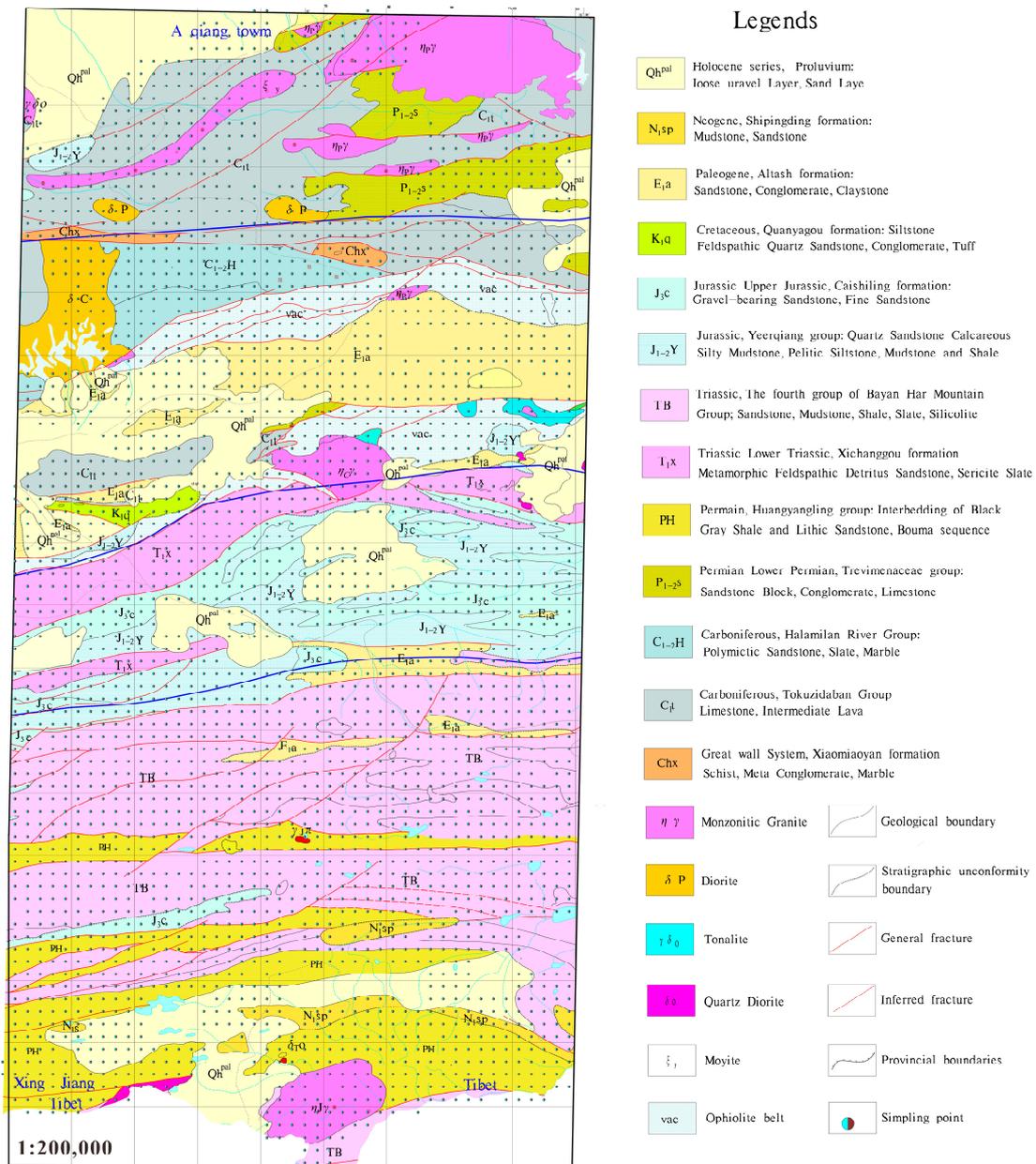


Figure 1. Regional geological map of the study area.

2.2.3. Yunwuling (Fold Belt) Cu–Au Ore Belt

The Yunwuling belt is located in the eastern section of the Muzi mineralization belt and is structurally a continental marginal active zone. The area around the Yunwuling belt exhibits porphyry-type copper mineralization associated with Tertiary granite and gold mineralization associated with the construction of Triassic volcanic rocks. The former mineralization region is centered on the sub-shallow potassium-long granite mass of Yunwuling, with some small oblique long granite porphyry bodies distributed on the periphery, while the latter is associated with copper mineralization. Although Cu and Au mineralization regions have been found in this mineralization belt, these regions are

distributed outside the study area, and only some Au occurrence and placer Au occurrence are found within the study area [38–41].

### 2.3. Characteristics of Mineral Resources

Within the field of study, previous methods of geochemical evaluation were employed in order to detect anomalies. The presence of anomalies was confirmed to be objective and reliable through the finding of ore spots and mineralization hotspots. However, the correlation between elements and the relationship between deposits, a large number of element anomalies failure to discover the deposits of the corresponding elements, and other issues have not been studied in depth.

The study shows that the central part of the study area is significantly enriched with gold, lead, and zinc, which are also the most important ore-forming elements in the area. The gold element is significantly enriched in the southern part of the study area. These characteristics provide detailed basic data for further exploring the relationship between element anomalies and minerals, and the relationship of the correlation between elements and minerals. At present, the metal mineral occurrences found in the study area are mainly gold, lead, and zinc. According to the genetic types, they can be divided into the following three types.

#### 2.3.1. Sedimentary Type

The placer gold occurs in the gravel layer (frozen soil layer) with strong gypsum salinization, and the thickness of this layer is generally 20–30 cm. The gold grade of the mineralized body is 0.5–0.8 g/m<sup>3</sup>, generally 3 × 3 × 1 mm<sup>3</sup> melon seed gold and a small amount of granular gold with different thicknesses.

There are many 1–2 mm wide quartz veins in the Hanzhugou placer gold deposit.

#### 2.3.2. Volcanic Type

The lead ore body is lenticular and cystic in the east of Yesanggang, which is produced along the cracks of silicified limestone, with a thickness of about 20–50 m. The main minerals are Galena and Sphalerite. Symbiotic minerals include pyrite, chalcopyrite, and secondary malachite.

#### 2.3.3. Skarn Type

Hydrothermal alteration and gold mineralization at the Guanshuigou: The mineralized bodies are predominantly layered, exhibiting lens-shaped distribution. There are four visible gold-bearing mineralization zones on the surface, each ranging from 200 to 500 meters in length and several meters to tens of meters in width.

## 3. Data Sources and Main Research Methods

### 3.1. Data Sources and Introduction

The research area is located in a high, cold mountainous region in the Eastern Kunlun Mountains, with elevations ranging from 4500 to 6400 m. A total of 3076 stream sediment samples with grain sizes of –10+80 mesh were collected at a density of 1.28 samples/4 km<sup>2</sup> according to the Regional Geochemical Exploration Work Progress standards (DZ/T 0167-2006). We identified 10 elements and 2 oxides. The quality analyses of the samples and test data representing 10 elements and 2 oxides, namely Au, Ba, Bi, Co, Cu, Fe<sub>2</sub>O<sub>3</sub>, MgO, Pb, Sn, Ti, V, and Zn, were conducted following the Geological and Mineral Laboratory Test Quality Management Work Progress standards (DZ/T 0130-2006 established in China (Ministry of Natural Resources, People's Republic of China 2006). The analysis method, detection limit, reporting rate, qualified internal rate, number of sample repetitions, and qualified abnormal rate of each element met the requirements outlined by the above specifications [41].

### 3.2. Random Forest Algorithm (RF)

Random forest is an ensemble learning algorithm based on the decision tree. It is a very popular and efficient algorithm at present. It can not only realize Bayesian classification mapping to minimize classification error, but it can also realize the construction strategy model evaluator of the regression model [42].

The given training dataset  $L = \{(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)\}$  is composed of observed random vectors  $(X, Y)$  with  $n$  instances. Where the vector  $X = \{X^1, X^2, \dots, X^p\}$ ,  $X \in \mathbb{R}^p$  is a predictor or an explanatory variable with  $p$  attributes,  $Y = \dagger$  is the class label or digital response. The principle of random forest is to randomly select  $k$  independent and identically distributed sample sets  $\{L_1, L_2, \dots, L_k\}$  (Each sample set contains  $n$  instances) from the learning sample  $L$  using the Bagging method, and randomly selecting  $m$  ( $m \leq p$ ) attributes from the  $p$  sample attributes in each sample set to construct a decision tree, which is composed of  $k$  decision trees [43]. Using the Bagging method has the following advantages: When random attributes are needed, the Bagging method can improve performance. Moreover, Bagging can be used to continuously estimate the generalization error of the combined tree set, as well as to estimate the strength and correlation, which are performed by the Out-of-bag data [44].

In this paper, the decision tree is constructed using the Gini index to select the partition attributes. The purity of the dataset  $L_k$  can be defined by the Gini value:

$$\begin{aligned} \text{Gini}(L_k) &= \sum_{c=1}^{|\dagger|} \sum_{c' \neq c} p_c p_{c'} \\ &= 1 - \sum_{c=1}^{|\dagger|} p_c^2 \end{aligned} \quad (1)$$

where  $p_c$  is the proportion of  $c$ -class samples in the dataset  $L_k$ ,  $\text{Gini}(L_k)$  reflects the probability that two samples are randomly selected from the dataset  $L_k$ , and their category labels are inconsistent; that is, the smaller the  $\text{Gini}(L_k)$ , the higher the purity of the dataset  $L_k$ . The Gini index of attribute  $a$  is defined as follows:

$$\text{Gini\_index}(L_k, a) = \sum_{v=1}^V \frac{|L_k^v|}{|L_k|} \text{Gini}(L_k^v) \quad (2)$$

where  $V$  is the value of discrete attribute  $a$ . In the candidate attribute set  $m$ , the attribute with the smallest Gini index is selected as the optimal partition attribute; that is,  $a_* = \arg \min_{a \in m} \text{Gini\_index}(L_k, a)$ .

#### 3.2.1. Undersampling Method

The undersampling method is to eliminate the harm of skewed distribution by discarding the intrinsic samples of most classes. The simplest but most effective method is random undersampling. That is, randomly selecting samples matching the number of minority classes from most classes for training, but it involves randomly eliminating the examples of most classes, resulting in the loss of important information. Liu et al. [45] proposed Easy Ensemble's undersampling method in 2009; that is, multiple subsets are independently extracted from most cases, and a classifier is established for each subset. Then, Adaboost is used to combine all the generated classifiers to make the final decision, and the experimental results show that the method has strong generalization ability. Based on this idea, this paper proposes ensemble random undersampling based on the random sampling method; that is, a training dataset is established by random sampling from the majority class samples many times without putting back, and the machine learning algorithm is used as the base learner for training. Finally, the prediction results are integrated. The advantage is that when the number of cycles is enough, all samples have the opportunity

to participate in the training, and the integrated model will contain more information than a single subset.

The pseudo-code of the binary classification problem is as follows Algorithm 1:

---

**Algorithm 1: Ensemble undersampling**

---

**Input:** The training sample  $D\{(x, y)\}$ ,  $D^+$  is the majority class sample,  $D^-$  is the minority class sample;  
 Base learner  $\Psi$ ;  
 Training rounds  $T$ .

**Output: Integrated model N.**

**Process:**

1: **for**  $t = 1, 2, \dots, T$  **do**

2: The subset  $D_t^+$  is randomly selected from  $D^+$ , and the size of the subset  $D_t^+$  is consistent with  $D^-$ ;

3: We use the base learner to train a single model  $N_t$  on the dataset  $D_t^+ \cup D^-$ :  

$$N_t(x) = \Psi(x)$$

4: **end for**

5: Integration of results

$$N(x) = \operatorname{argmax}_{y \in \{y^+, y^-\}} \sum_{t=1}^T (N_t(x) = y)$$


---

### 3.2.2. Performance Evaluation Parameters

The results of the metallogenic prediction model based on the random forest algorithm can be represented by the confusion matrix (Table 1) [46]. Among them, TP and FP are the number of true/false positive samples, and TN and FN are the number of true/false negative samples, respectively. In the mineral prediction, the prospective area is positive, and the non-prospective area is negative. The error rate becomes an unsuitable evaluation metric in the presence of grade imbalances or cost disparities. Therefore, this paper uses G-mean [47] and the area under the ROC curve AUC [48] as performance evaluation indicators, defined in Equations (3)–(6).

**Table 1.** Confusion matrix.

	Predicted Positive (PP)	Predicted Negative (PN)
Actual Positive (AP)	TP (True Positive)	FN (False Negative)
Actual Negative (AN)	FP (False Positive)	TN (True Negative)

AUC is a reliable performance indicator for measuring imbalance and cost-sensitive problems. The ROC curve is employed in binary classification scenarios to visually represent the performance of a model. It accomplishes this by graphing the true positive ratio (TPR) along the  $y$ -axis and the false positive ratio (FPR) along the  $x$ -axis. The ROC curve describes the relative trade-off between gains (true positive) and losses (false positive). AUC is the area under the line, which combines the performance of the classification method of all possible values of FPR.

$$\text{Accuracy}(Acc) = \frac{TP + TN}{TP + FP + TN + FN} \tag{3}$$

$$\text{True Positive Rate}(Acc_+) = \frac{TP}{TP + FN} = \text{Recall} \tag{4}$$

$$\text{True Negative Rate}(Acc_-) = \frac{TN}{TN + FP} \tag{5}$$

$$\text{G-mean} = \sqrt{Acc_+ \times Acc_-} \tag{6}$$

### 3.2.3. Hold-Out Test

A model is very important to its predictive ability in the new sample, so it is often not enough to evaluate its fitting ability only on the original dataset. It is also necessary to evaluate the generalization ability of the model, that is, to evaluate the predictive ability of the model to the new sample (test set). In this paper, the data are divided into a training set and a test set using the method of leaving out. A total of 80% of the data is used to train the model, and the obtained model is predicated on the test set of 20% data. The accuracy of the model is evaluated, and the corresponding ROC curve and AUC value are obtained.

## 4. Metallogenic Prediction and Target Area Delineation of Au Polymetallic Deposits

### 4.1. Geochemical Characteristics of Different Geological Units

When using stream sediments to express the geochemical characteristics of the strata, to understand the distribution and distribution characteristics of the elements in this area, the average values of the whole region and each geological unit are counted to reflect the content variation characteristics of the elements in this area. The coefficient of variation is used to reflect the relative dispersion degree between different elements, and the regional concentration coefficient (the average value of the elements in the geological unit/the average value of the elements in the whole area) is used to reflect the enrichment and dilution degree of the elements in the geological unit. This parameter is also called a contrast; greater than 1.2 is considered to be enriched, and less than 0.8 is considered to be depleted. The average value, standard deviation, and coefficient of variation of stream sediment samples in each geological unit are listed in the following table, and the Great Wall system, Permian system, Triassic system, Paleogene system, and Neogene system which may be related to mineralization, are discussed.

#### 4.1.1. Great Wall System (Ch)

It can be seen from Table 2 that compared with the whole area, the relatively enriched elements (including oxides) are Au (1.9) and Sn (1.3), which are a set of elements closely associated with Au and some elements related to granite. The relatively depleted elements are elements (including oxides), Ba (0.7), Co (0.7), Mn (0.7), Fe<sub>2</sub>O<sub>3</sub> (0.7), and the iron group elements are depleted, indicating that the basic ultrabasic rocks in the strata are not developed.

**Table 2.** Geochemical characteristic parameter table of stream sediment elements in geological units.

Parameter	Geological Unit	N	E	K	J	T	P	C	Ch	Whole Area
Au	Mean Value	1.0	1.1	0.6	0.7	0.8	1.1	1.6	2.1	1.1
	Standard Deviation	0.8	1.2	0.4	0.7	0.6	2.2	1.6	2.8	1.5
	Coefficient of Variation	0.8	1.1	0.7	0.9	0.8	2.0	1.0	1.2	1.4
	Concentration Coefficient	0.9	1.0	0.5	0.7	0.7	1.0	1.4	1.9	1.0
Ba	Mean Value	639.6	756.5	677.9	1598.3	793.9	580.1	526.3	513.9	766.7
	Standard Deviation	502.1	531.1	130.1	6135.3	1776.1	326.5	219.6	95.3	2307.8
	Coefficient of Variation	0.8	0.7	0.2	3.8	2.2	0.5	0.4	0.1	3.0
	Concentration Coefficient	0.8	1.0	0.9	2.0	1.0	0.7	0.6	0.6	1.0
Bi	Mean Value	0.3	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
	Standard Deviation	0.4	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2
	Coefficient of Variation	1.1	0.5	0.2	0.5	0.3	0.7	0.5	0.4	1.0
	Concentration Coefficient	1.5	1.0	0.5	1.0	1.0	1.0	1.0	1.0	1.0
Co	Mean Value	11.6	9.7	7.6	8.9	12.1	10.7	10.1	7.8	10.7
	Standard Deviation	3.0	4.5	2.2	3.9	3.1	3.6	3.6	3.3	4.3

Co

Table 2. Cont.

Parameter	Geological Unit	N	E	K	J	T	P	C	Ch	Whole Area
Cu	Mean Value	21.5	21.5	18.7	25.2	26.6	20.7	27.6	28.2	24.4
	Standard Deviation	4.9	9.8	3.8	9.9	8.9	7.4	16.6	22.7	11.1
	Coefficient of Variation	0.2	0.5	0.2	0.3	0.3	0.3	0.6	0.7	0.4
	Concentration Coefficient	0.9	0.9	0.8	1.0	1.1	0.8	1.1	1.1	1.0
Fe	Mean Value	4.7	3.7	3.1	4.3	5.2	4.4	4.0	3.1	4.4
	Standard Deviation	1.2	1.7	0.7	1.8	1.4	1.4	1.3	1.0	1.7
	Coefficient of Variation	0.3	0.5	0.2	0.4	0.2	0.3	0.3	0.3	0.4
	Concentration Coefficient	1.1	0.8	0.7	0.9	1.1	1.0	0.9	0.7	1.0
Mg	Mean Value	1.3	1.6	1.2	1.4	1.7	1.6	2.0	1.6	1.7
	Standard Deviation	0.4	0.8	0.4	0.6	0.3	0.4	0.7	1.0	0.8
	Coefficient of Variation	0.3	0.5	0.3	0.4	0.2	0.3	0.3	0.6	0.5
	Concentration Coefficient	0.8	0.9	0.7	0.8	1.0	0.9	1.1	0.9	1.0
Sn	Mean Value	2.7	1.8	1.4	1.7	2.1	2.1	2.1	2.9	2.3
	Standard Deviation	1.3	0.8	0.2	0.5	0.8	1.0	0.7	1.0	2.0
	Coefficient of Variation	0.5	0.4	0.1	0.3	0.4	0.4	0.3	0.3	0.9
	Concentration Coefficient	1.2	0.8	0.6	0.7	0.9	0.9	0.9	1.2	1.0
Ti	Mean Value	3160.2	2395.3	1861.3	1880.9	2805.9	2515.4	2509.1	2244.3	2574.3
	Standard Deviation	1840.7	979.1	676.5	875.1	939.3	967.3	795.5	631.9	1162.0
	Coefficient of Variation	0.5	0.4	0.3	0.4	0.3	0.4	0.3	0.3	0.4
	Concentration Coefficient	1.2	0.9	0.7	0.7	1.0	1.0	1.0	0.9	1.0
V	Mean Value	65.4	60.8	53.4	49.5	64.0	59.3	71.1	55.0	63.1
	Standard Deviation	19.8	25.4	22.1	18.2	17.5	22.7	26.2	32.9	25.8
	Coefficient of Variation	0.3	0.4	0.4	0.3	0.2	0.4	0.3	0.6	0.4
	Concentration Coefficient	1.0	0.9	0.8	0.7	1.0	0.9	1.1	0.9	1.0
Pb	Mean Value	20.5	15.2	14.8	17.6	21.8	19.2	15.3	15.6	18.4
	Standard Deviation	5.6	21.3	3.4	13.5	21.9	7.6	5.4	3.2	15.1
	Coefficient of Variation	0.3	1.4	0.2	0.7	1.0	0.4	0.3	0.2	0.8
	Concentration Coefficient	1.1	0.8	0.8	0.9	1.1	1.0	0.8	0.8	1.0
Zn	Mean Value	68.9	50.1	46.1	64.6	78.3	65.5	64.8	60.0	66.0
	Standard Deviation	18.2	22.7	11.9	79.3	65.0	25.6	21.5	24.1	48.6
	Coefficient of Variation	0.3	0.4	0.2	1.2	0.8	0.4	0.3	0.4	0.7
	Concentration Coefficient	1.0	0.7	0.7	0.9	1.1	1.0	0.9	0.9	1.0

Note: Content unit: Fe is Fe<sub>2</sub>O<sub>3</sub>, Mg is MgO, oxide is %, other elements is ×10<sup>-6</sup>; Ch is Great Wall System; C is Carboniferous; P is Permian; T is Triassic; J is Jurassic; K is Cretaceous; E is Paleogene; N is Neogene.

From the coefficient of variation, Au (1.3) and Cu (0.8) are extremely uneven and strongly differentiated elements, which are likely to be locally mineralized. Ba (0.2) and Pb (0.2) are evenly distributed in the strata, and there is no metallogenic basis.

#### 4.1.2. Permian System (P)

There are no relatively depleted elements (including oxides). It is almost consistent with the level of element content in the whole region. From the coefficient of variation, Au (2.0) is a strongly differentiated element, which may be locally mineralized, and other elements are evenly distributed.

#### 4.1.3. Triassic System (T)

The only relatively depleted element is Au (0.7), and the other elements are evenly distributed. It shows that the element content in the stratum is close to the average level of the whole area.

From the coefficient of variation, Au (0.8), Ba (2.2), Pb (1.0), and Zn (0.8) are strongly differentiated elements, which have the possibility of mineralization. MgO (0.2) is an undivided element.

#### 4.1.4. Paleogene System (E)

There is no relatively enriched element (including oxides), and the relatively depleted element is Zn (0.8).

From the coefficient of variation, Au (1.08) and Pb (1.40) are strongly differentiated elements, which may be partially mineralized.

#### 4.1.5. Neogene System (N)

The relatively enriched elements (including oxides) are Ti (1.2). The relatively depleted element is MgO (0.8).

From the coefficient of variation, Au (0.8) and Ba (0.8) are strongly differentiated elements, which have great metallogenic possibility. Cu (0.2) belongs to the undivided elements.

### 4.2. Data Preprocessing and Dataset Establishment

#### 4.2.1. Evidence Layer Information Extraction

The data used in this paper include 1:200,000 regional geological map (strata, faults, and mineral distribution) and 1:200,000 stream sediment geochemical data. The sampling layout was designed to balance uniformity and rationality, maximizing control over the measurement area to cover the entire study area effectively. The obtained mixed sample was placed in the center of the grid and assigned a value to represent the catchment area within that grid. The data analysis, processing, and extraction are as follows.

##### (1) Stratum

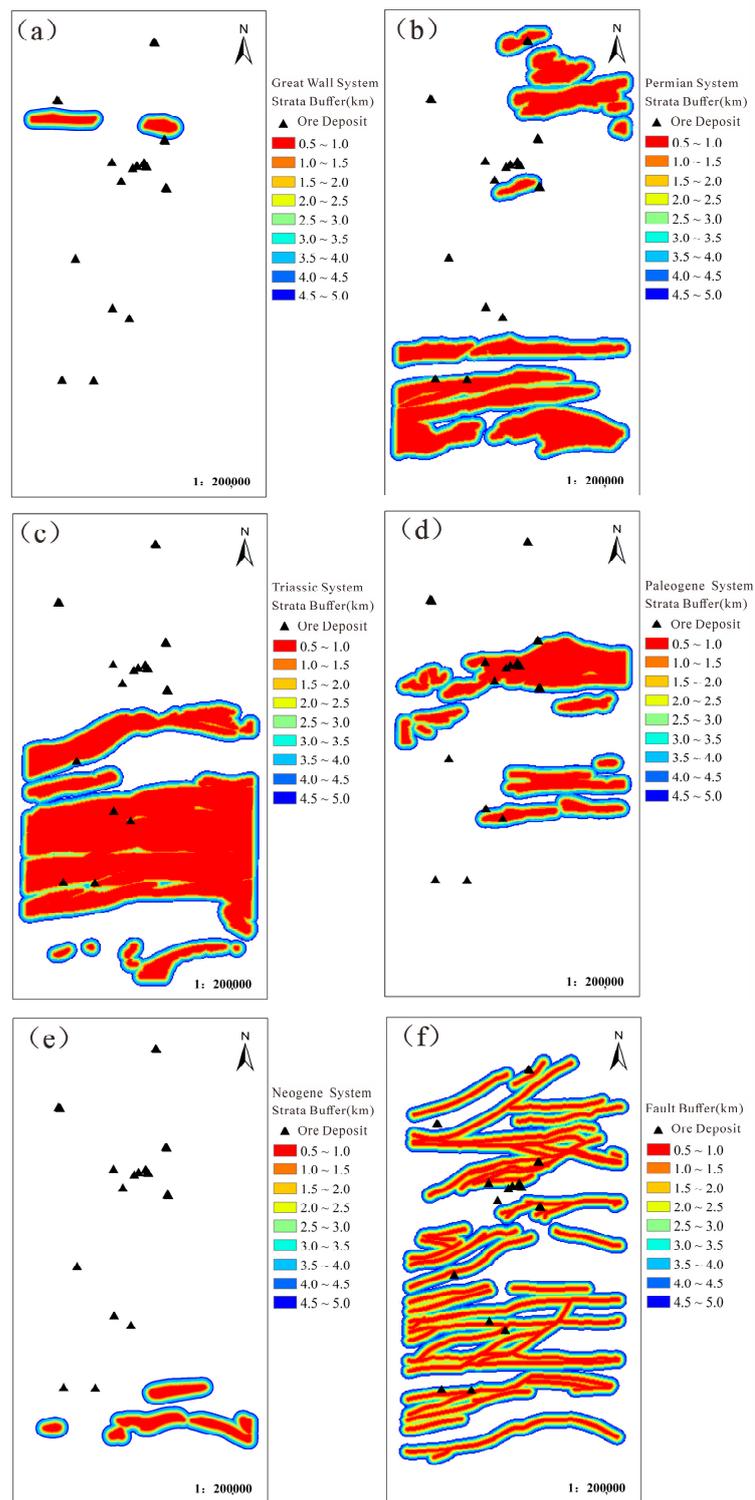
According to the geochemical characteristics of the geological units in 4.1, it is concluded that the Great Wall system, Permian system, Triassic system, Paleogene system, and Neogene system are the strata closely related to mineralization in the study area. The known ore spots also correspond well with these strata, so these five strata are selected, respectively, and a multi-ring buffer zone with a distance of 500 m and 10 rings is established (Figure 2).

##### (2) Regional fracture structure

The deep faults of different natures in different periods in the study area are very developed. The multi-stage and multi-stage magmatic–hydrothermal activities and the distribution of related ophiolitic mélangé belts provide favorable geological conditions for the enrichment/depletion of elements. As an important basis for regional prospecting, this paper selects the main ore-controlling fault structure as the evidence layer and establishes a multi-ring buffer zone with a spacing of 500 m and 10 rings (Figure 2).

##### (3) Regional geochemistry

The results of factor analysis showed that four factors with eigenvalues  $\geq 1$  were extracted, and the cumulative variance contribution rate was 75.58% (Table 3). The factor loading matrix was rotated using the Kaiser normalized maximum variance method, and each factor score for each sample was calculated. The eigenvalue represents the variance of the factor, which highlights the importance of the factor.



**Figure 2.** A multi-ring buffer zone; (a) the Great Wall system; (b) Permian; (c) Triassic; (d) Paleogene; (e) Neogene; (f) faults. Note: These ore spots are gold ore spots and gold mineral occurrence.

**Table 3.** Principal factor variance explanation rates are derived based on the factor analysis.

Component	Initial Eigenvalues			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.65	38.75	38.75	4.31	35.91	35.91
2	1.85	15.44	54.19	1.95	16.23	52.14
3	1.56	13.01	67.19	1.80	15.00	67.14
4	1.01	8.39	75.58	1.01	8.44	75.58
5	0.92	7.67	83.25			
6	0.64	5.35	88.60			
7	0.43	3.61	92.21			
8	0.33	2.76	94.96			
9	0.27	2.21	97.18			
10	0.19	1.58	98.75			
11	0.10	0.81	99.56			
12	0.05	0.44	100.00			

Note: Factors with an eigenvalue  $\geq 1$  are selected, and cumulative indicates the frequency at which each factor explains the total variance.

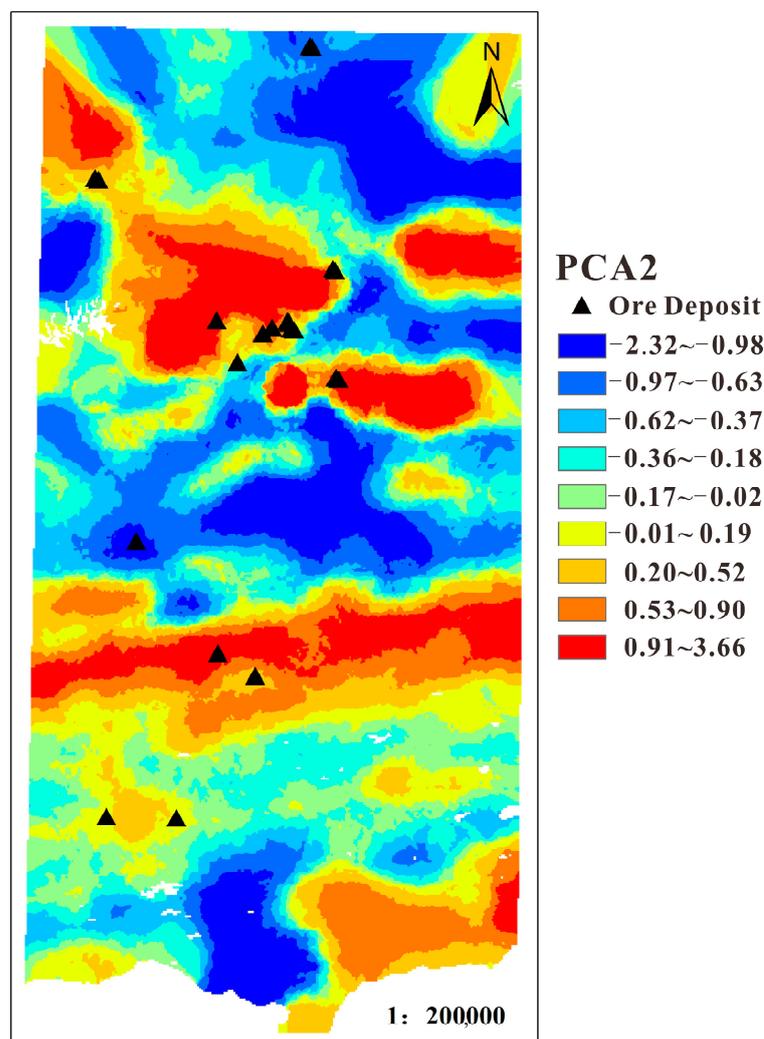
For factor 1, the variance explanation rate of the Co-V-Ti-Fe<sub>2</sub>O<sub>3</sub>-MgO-Cu group was 38.75%; this element group is linked to the suture movement of the Carboniferous-Permian plate in the Central Kunlun area and is associated with the outcropping of ultramafic and mafic rocks, mafic intrusive rocks, mafic volcanic rocks, and ophiolite belts; the observed anomalies correspond well to the Carboniferous-Permian ophiolite belt [49–51].

For factor 2, the variance explanation rate of the Pb-Zn group was 15.44%; this result may have been related to medium-moderate and medium-high temperature hydrothermal mineralization [52], reflecting the mineralization of these elements in the Central Kunlun area. Yesanggang East lead ore and Hongyu lead and zinc ore deposits have been found in the study area.

For factor 3, the variance explanation rate of the Bi-Sn-Au group was 13.01%, reflecting the contribution of the gold element mineralization group in the study area [53]. Liuzonggou and Hanzhugou placer gold deposits have been found in the area. Sn and Bi are mainly related to high-temperature volcanic hydrothermal activity, and as Bi is associated with a combination of the characteristic indicator elements of the tailing halo of gold deposits, it is often closely associated with Au.

For factor 4, the variance explanation rate of Ba was 8.39%. In addition, barite has been found in the sedimentary rocks in the eastern part of the study area as a result of Ba mineralization [54].

To express the results of factor analysis more intuitively, factors 2, 3, and 4 are combined into one factor, and these two main factors are manually extracted. The twelve target elements are divided into two main factors. Factor one, Co, V, Ti, Fe<sub>2</sub>O<sub>3</sub>, MgO, and Cu are a group of background elements closely related to regional geological background. Pb, Zn, Bi, Sn, Au, and Ba are potential ore-forming element combinations with metallogenic prospects. Among them, the Pb-Zn group may have been related to medium-moderate and medium-high temperature hydrothermal mineralization. Bi-Sn-Au is related to gold mineralization in the area; Ba is closely related to the common barite mineralization in the area. Therefore, the extraction factor two is reclassified by the classification number 10, and the Kriging interpolation method is used as one of the evidence layers of this training prediction (Figure 3).



**Figure 3.** PCA2 factor score plot after reclassification. Note: These ore spots are gold ore spots and gold mineral occurrence.

Combined with the existing geological and regional geochemical data and based on the above analysis and processing, seven evidence layers were finally selected for the prediction and evaluation of mineral resource prospects of gold polymetallic deposits in the central Kunlun area of Xinjiang: (1) Great Wall strata; (2) Permian strata; (3) Triassic strata; (4) Paleogene strata; (5) Neogene strata; (6) regional fault structure; (7) factor score of regional geochemical data PCA2.

#### 4.2.2. Establishment of Datasets

There are a series of practical difficulties in the application of machine learning in mineral prediction, such as the imbalance of label data (mineralization exists or not) [31]. At present, more than 20 large and small deposits have been found in the central Kunlun area of Xinjiang, and a few specific locations of non-deposit areas containing different information have been marked, and there is a problem of obvious imbalance of label data. Therefore, the researchers selected 27 gold polymetallic ore deposits and 10 non-prospective ore locations from the region, extracted the input evidence layer values, and created a training dataset. Each training dataset consists of a set of input feature vectors and a binary classification value, where 1 is a metallogenic prospect and 0 is a non-metallogenic prospect. The prediction dataset is stream sediment grid data occurrence. Due to the obvious data imbalance problem in the training dataset, this paper uses two undersampling methods for comparative study. One is to use the integrated random undersampling method mentioned

above for sampling. Second, by fully considering the geological factors, the typical representative known occurrence is selected. The two are verified and evaluated by the leave-out method and the model performance evaluation parameters. Finally, the metallogenic prediction results of these two different undersampling methods are evaluated.

### 4.3. Random Forest Algorithm Implementation

#### 4.3.1. Parameter Optimization

The parameterization of a random forest has a great influence on its robustness and generalization ability, which affects the accuracy of mineral prediction. The size and depth of the random forest tree are crucial to the performance of machine learning. Too large or too small may lead to unsatisfactory results.

Unlike most machine learning methods, the random forest generation prediction model only needs to set two important parameters: the number of decision trees (trees) and the number of evidence features that each node uses to make the decision tree grow (m). Breiman [42] proved that with the increase in trees, the generalization error is always convergent, and there is no overfitting problem in overtraining. On the other hand, reducing the number of m will lead to a decrease in the correlation between trees, thereby improving the accuracy of the model. To optimize these parameters, a large number of experiments were carried out using different numbers of trees and feature numbers. The value range of trees was set to 2 to 6, and the interval was 1. The number of feature variables m was 2, 3, and 4, respectively. The results show (Figure 4, Table 4) that the average RF accuracy is 0.85, the standard deviation is 0.16, and it has high accuracy and stability. It can be seen that RF is less sensitive to parameter changes. The stable performance of RF is mainly attributed to the combination of multiple classifiers trained under certain conditions. On the one hand, the evidence features used for tree induction are randomly selected, which reduces the correlation between individual models, reduces the generalization error, and provides a very stable prediction. In addition, the feature selection method adds a re-sampling (Bagging) of the training data for each tree, which helps to increase the diversity of the models that make up the whole and prevents the decision tree from overfitting the data.

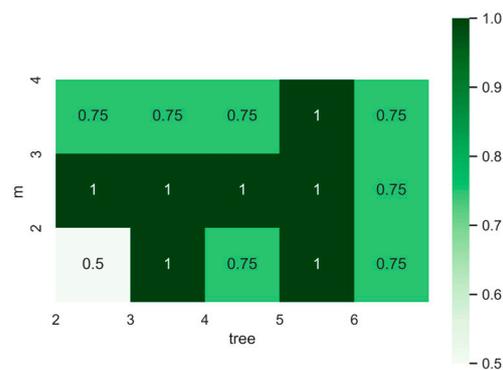


Figure 4. Accuracy of random forest model under hyperparameters.

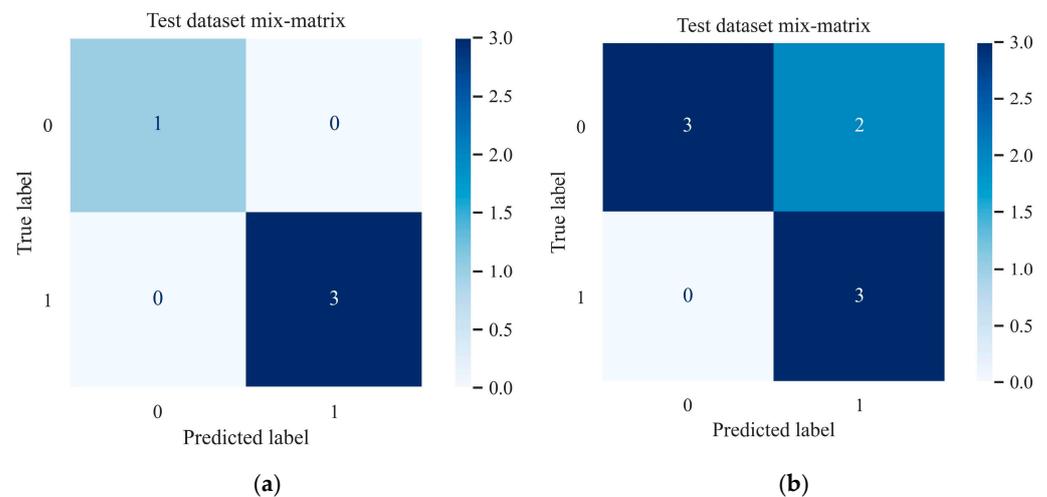
Table 4. Hyperparameter accuracy statistics under random forest model.

	RF
Min	0.50
Max	1.00
Mean	0.85
Std	0.16

#### 4.3.2. Prediction Performance Assessment

Random undersampling is used to sample the unbalanced data (27 ore spots and 10 non-ore spots) of the training dataset to ensure that the data ratio of 0 and 1 in the new

dataset is approximately 1:1. Then, the newly generated data is trained and predicted by the hold-out method (80% training and 20% test), and the final prediction results are integrated. For the divided training dataset and prediction dataset, 15 parallel experiments were carried out by Tree taking 2–6, m taking 2, 3, and 4, and the best results were taken. Figure 5 lists the confusion matrix of the test set prediction of the random forest model under different sampling methods. It can be seen that the number of positive and negative samples of the correct classification is far greater than the number of samples of the wrong classification, and the model application is feasible. Based on this confusion matrix, the calculated evaluation parameters are shown in Table 5 (see Equations (3)–(6) for the calculation method). The random forest model shows different performances with different sampling methods.



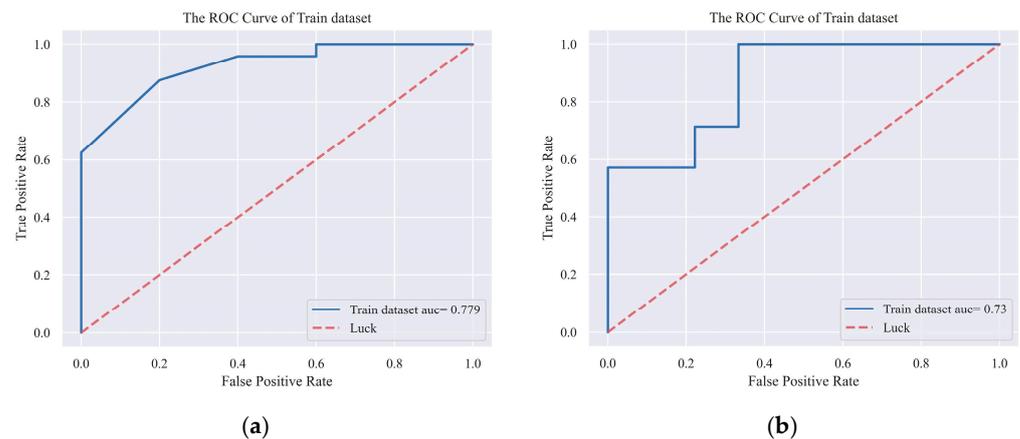
**Figure 5.** The confusion matrix of the random forest model under different sampling methods. (a) Select training samples. (b) Integrated random undersampling.

**Table 5.** Prediction performance of random forest model under different sampling methods.

	Select Training Samples	Integrated Random Undersampling
	RF	RF
Acc	1.00	0.75
Acc+	1.00	1.00
Acc-	1.00	0.60
G-mean	1.00	0.87

As far as different sampling methods are concerned, the evaluation parameters of selecting representative training samples for training are higher than the results of integrated random undersampling. This is because the sampling method of selecting representative samples is to eliminate some known occurrence, to make the training model fit better, and to make the prediction results of the test set more accurate.

The ROC curve can evaluate the prediction performance of the high-probability region, and the predicted classification results are judged by different discriminant thresholds. The area under the curve (AUC) value is used to evaluate the overall performance of different prediction models [55]. The RF curve under the selected training samples has a relatively high AUC value (Figure 6a), while the ROC curve under the integrated random undersampling (Figure 6b) is slightly inferior, but the overall gap is not large. This may be related to RF as an integrated model [42]. As an integrated algorithm based on decision trees as the base learner, RF ensures the independence of individual models by random sampling of training variables and samples, thereby improving the generalization ability. In the face of data changes in the random sampling process, its fitting effect is also more stable.

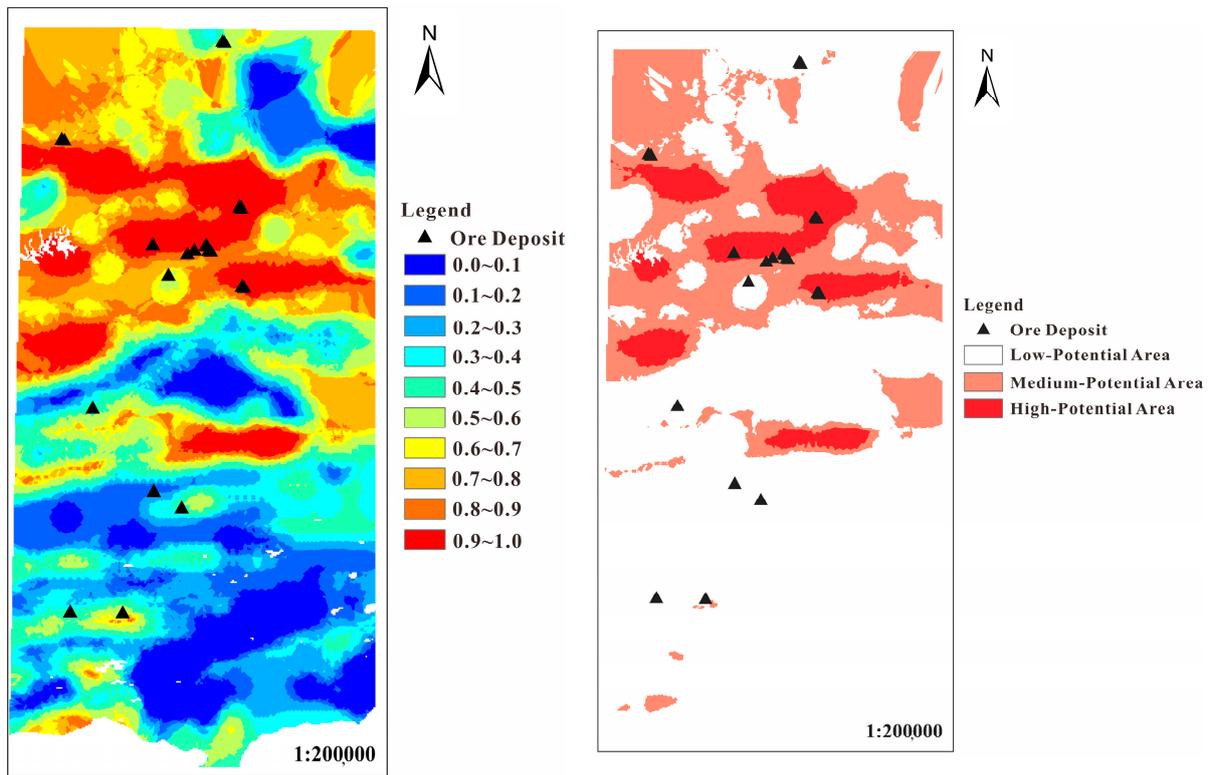


**Figure 6.** ROC curve and AUC value of random forest model under different sampling methods. (a) Select training samples. (b) Integrated random undersampling.

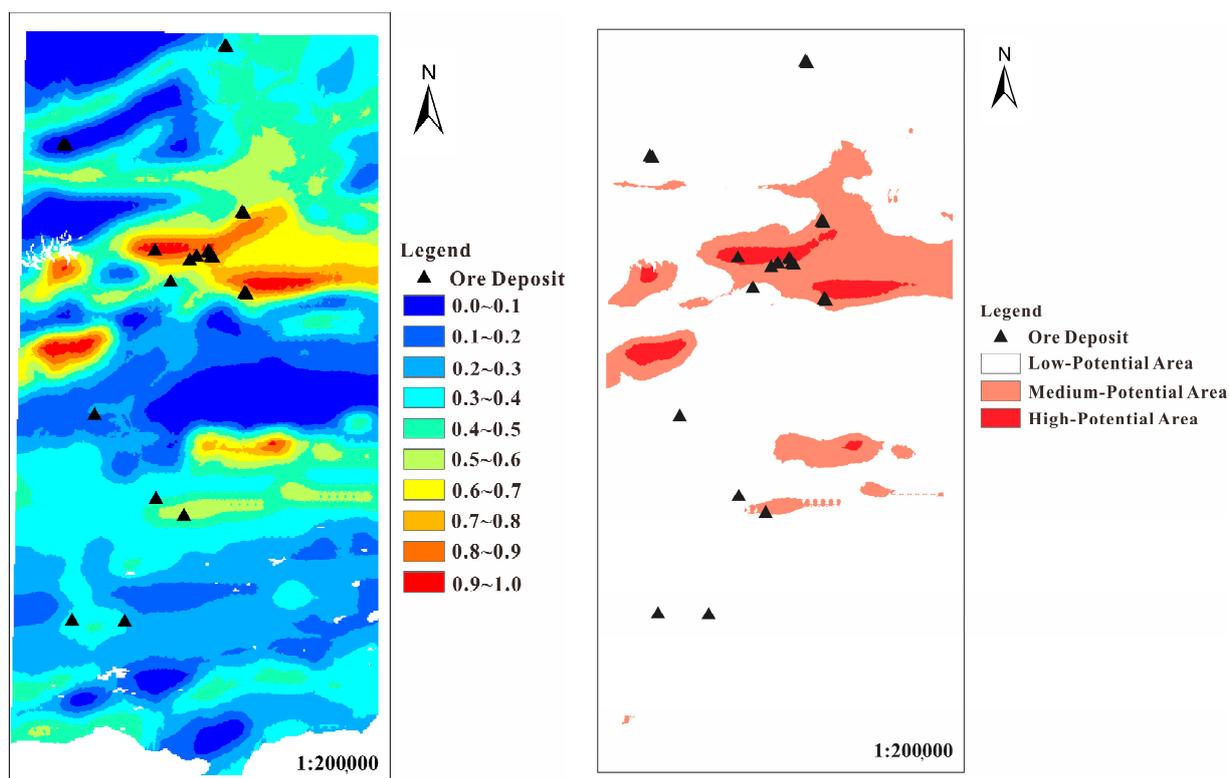
In short, from the above results, it can be concluded that the prediction performance parameters in the training process show that the prediction results of selecting representative training samples are better than the results of ensemble random undersampling.

#### 4.3.3. Metallogenic Prediction Results

Figures 7 and 8 show the prospect map of random forest metallogenic prediction under the training of integrated random undersampling and representative sample selection, respectively. The predicted value of each point is expressed by a 0–1 floating point value, which indicates the metallogenic probability value at this point. The regions with probabilities greater than 0.5 and 0.8 are extracted as delineated mineralization potential zones and high-potential zones, while regions below 0.5 are considered low-potential zones.



**Figure 7.** Random forest metallogenetic probability and metallogenetic potential division trained by under-integrated random undersampling. Note: These ore spots are gold ore spots and gold mineralization occurrence.



**Figure 8.** Random forest metallogenetic probability and metallogenetic potential division trained by selecting representative samples. Note: These ore spots are gold ore spots and gold mineralization occurrence.

Table 6 shows the statistical information for each metallogenetic potential area under different sampling methods. It can be observed that the metallogenetic prediction results predicted by selecting training samples have the following advantages compared with the integrated random undersampling method: (1) The number of ore spots misjudged in the low-potential area is small, and the proportion of ore deposits is low. (2) The number of ore spots correctly predicted in the metallogenetic potential area is large, and the ore-bearing ratio is high. (3) The number of ore-bearing occurrences and the proportion of ore-bearing occurrences in the high-potential area are higher; the above results show that the machine learning training model under the method of selecting training samples has learned the complex information of the original data to a greater extent and has been successfully applied to the metallogenetic prediction results and prospect division.

**Table 6.** Statistical information of different metallogenic potential areas.

		Integrated Random Undersampling	Select Training Samples
		RF	RF
Low-potential areas	Known number of occurrence	7	5
	Area (km <sup>2</sup> )	11,240	13,768
	Ore-bearing ratio (%)	0.062	0.036
Mineralization potential area	Known number of occurrence	9	10
	Area (km <sup>2</sup> )	3447	1702
	Ore-bearing ratio (%)	0.261	0.588
High-potential areas	Known number of occurrence	11	12
	Area (km <sup>2</sup> )	1068	286
	Ore-bearing ratio (%)	1.030	4.196

From the above metallogenic prospect prediction and metallogenic potential division results, it can be concluded that the selection of the training sample method is more reliable than the integrated random undersampling method in the metallogenic potential area due to the large number of accurately predicted ore spots and the high proportion of ore content per unit area.

## 5. Conclusions

Based on the previous research results, the researchers applied the random forest algorithm under different sampling methods for the first time to predict and delineate the 10 elements and 2 oxides of 3076 stream sediments collected by the medium–small-scale regional geochemical exploration in the central Kunlun area of Xinjiang. It provides a basis for the future medium–small-scale metallogenic elements from qualitative identification to quantitative prediction and discusses the prediction effect of a random forest algorithm on gold polymetallic minerals in the central Kunlun area of Xinjiang under different sampling methods. The main conclusions are as follows:

- (1) In this study, the strata, fault structure, and geochemical information are extracted, and the known ore spots and geochemical data are used to form a training set and a prediction set to construct a random forest model. The difference between the integrated random undersampling method and the selection of the training sample method is compared, which expands the prospecting idea of machine learning algorithms in mathematical geology in the central Kunlun area of Xinjiang.
- (2) For different sampling methods, the performance evaluation parameters of the training process show that the prediction accuracy of the selected training samples is higher, indicating that the fitting effect and generalization ability are stronger. Ensemble random undersampling can weaken the difference between the base learners and achieve consistency in the overall results. The results of metallogenic prospect prediction and potential area division better explain this.
- (3) The sampling method of selecting training samples is more reliable because it can fully learn the complex information of the original data, the number of ore spots accurately predicted by the prediction results is greater, and the proportion of ore is higher. For the random forest model of different sampling methods, the random forest algorithm under the selected training samples has more reference value and further exploration significance in the actual exploration problems considering the cost because of its small area of high-potential prediction area and high proportion of ore per unit area.

**Author Contributions:** Conceptualization, Y.Z. and S.X.; methodology, Y.Z. and S.X.; software, Y.Z., J.D., and X.Z. (Xuwei Zhou); validation, Y.Z. and S.X.; formal analysis, Y.Z.; data curation, Y.Z. and X.Y.; writing—original draft preparation, Y.Z., O.Y.; writing—review and editing, Y.Z. and S.X.; visualization, Y.Z., X.Z. (Xuwei Zhou), and X.Y.; supervision, Y.Z., S.X., and X.Z. (Xiaoying Zhou); project administration, Y.Z. (Xiaoying Zhou); funding acquisition, S.X. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was financially supported by funds from the National Natural Science Key Fund Project (71132008), the Northwestern Basic Geological Survey and Data Update Project of China Geological Survey (1212010911051), the Natural Science Key Project of Inner Mongolia Education Department (NJZZ11067), and the Inner Mongolia Natural Science Foundation (2015BS0702).

**Data Availability Statement:** Not applicable.

**Acknowledgments:** We are very thankful for all the editors and reviewers who have helped us improve the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Kong, W.H.; Xiao, K.Y.; Chen, J.P.; Sun, L.; Li, N. A combined prediction method for reducing prediction uncertainty in the quantitative mineral resources prediction. *Earth Sci. Front.* **2021**, *28*, 128–138, (In Chinese with English abstract). [\[CrossRef\]](#)
2. Liu, G.; Wang, Y.Z.; Xue, T.; Wu, C.Y.; Xue, B.; Tang, T.T.; Liu, S.M. Mineral Resource Spatial Association Analysis and Prediction: A Case Study in Western China. *Geoscience* **2019**, *33*, 751–758, (In Chinese with English abstract). [\[CrossRef\]](#)
3. Agterberg, F.; Kelly, A. Geomathematical methods for use in prospecting. *Can. Min. J.* **1971**, *5*, 61–72.
4. Griffiths, J.; Menzie, D.; Labovitz, M. Exploration for and evaluation of natural resources. In Proceedings of the AAPG Research Symposium, Probability Methods in Oil Exploration, Stanford, CA, USA, 20–22 August 1975; pp. 20–22.
5. Singer, D.A. RESIN, a FORTRAN IV program for determining the area of influence of samples or drill holes in resource target search. *Comput. Geosci.* **1976**, *2*, 249–260. [\[CrossRef\]](#)
6. Singer, D.A. Basic concepts in three-part quantitative assessments of undiscovered mineral resources. *Nonrenewable Resour.* **1993**, *2*, 69–81. [\[CrossRef\]](#)
7. Singer, D.A. Progress in integrated quantitative mineral resource assessments. *Ore Geol. Rev.* **2010**, *3*, 242–250. [\[CrossRef\]](#)
8. Wu, D.C.; Lu, W.Q.; Wang, G.P. 3D geological modeling and metallogenic prediction of Yimaquan M14 magnetic anomaly area in Geermu City of Qinghai. *Miner. Resour. Geol.* **2023**, *37*, 55–61, 71, (In Chinese with English abstract). [\[CrossRef\]](#)
9. Song, W.; Zheng, L.; Liu, J.; Cao, S.; Xie, Z. Genesis, metallogenic model, and prospecting prediction of the Nibao gold deposit in the Guizhou Province, China. *Acta Geochim.* **2023**, *42*, 136–152. [\[CrossRef\]](#)
10. Carranza, E.J.M.; Hale, M. Logistic Regression for Geologically Constrained Mapping of Gold Potential, Baguio District, Philippines. *Explor. Min. Geol.* **2001**, *3*, 165–175. [\[CrossRef\]](#)
11. Li, W.; Neubauer, F.; Liu, Y.J.; Genser, J.; Ren, S.M.; Han, G.Q.; Liang, C.Y. Paleozoic Evolution of the Qimantage Magmatic Arcs, Eastern Kunlun Mountains: Constraints from Zircon Dating of Granitoids and Modern River Sands. *J. Asian Earth Sci.* **2013**, *77*, 183–202. [\[CrossRef\]](#)
12. Seraj, R.R.R. A hybrid GIS-assisted framework to integrate Dempster-Shafer theory of evidence and fuzzy sets in risk analysis: An application in hydrocarbon exploration. *Geocarto Int.* **2021**, *36*, 5a8. [\[CrossRef\]](#)
13. Behera, S.; Panigrahi, M.K. Mineral prospectivity modelling using singularity mapping and multifractal analysis of stream sediment geochemical data from the auriferous Hutti-Maski schist belt, S. India. *Ore Geol. Rev.* **2021**, *131*, 104029. [\[CrossRef\]](#)
14. Koike, K.; Matsuda, S.; Suzuki, T.; Ohmi, M. Neural Network-Based Estimation of Principal Metal Contents in the Hokuroku District, Northern Japan, for Exploring Kuroko-Type Deposits. *Nat. Resour. Res.* **2002**, *2*, 135–156. [\[CrossRef\]](#)
15. Porwal, A.; Carranza, E.J.M.; Hale, M. Artificial Neural Networks for Mineral-Potential Mapping: A Case Study from Aravalli Province, Western India. *Nat. Resour. Res.* **2003**, *3*, 155–171. [\[CrossRef\]](#)
16. Choi, S.; Moon, W.M.; Choi, S.-G. Fuzzy logic fusion of W-Mo exploration data from Seobyeg-ri, Korea. *Geosci. J.* **2000**, *2*, 43–52. [\[CrossRef\]](#)
17. Luo, X.; Dimitrakopoulos, R. Data-driven fuzzy analysis in quantitative mineral resource assessment. *Comput. Geosci.* **2003**, *1*, 3–13. [\[CrossRef\]](#)
18. Liu, Y.; Cheng, Q.M.; Xia, Q.L.; Wang, X.Q. Mineral potential mapping for tungsten polymetallic deposits in the Nanling metallogenic belt, South China. *J. Earth Sci.* **2014**, *4*, 689–700. [\[CrossRef\]](#)
19. Wang, W.L.; Xie, S.Y.; Carranza, E.J.M. Introduction to the thematic collection: Applications of innovations in geochemical data analysis. *Geochem. Explor. Environ. Anal.* **2022**, *23*, 1–2. [\[CrossRef\]](#)
20. Carranza, E.J.M. Weights of Evidence Modeling of Mineral Potential: A Case Study Using Small Number of Prospects, Abra, Philippines. *Nat. Resour. Res.* **2004**, *3*, 173–187. [\[CrossRef\]](#)

21. Yang, F.; Xie, S.Y.; Hao, Z.; Carranza, E.J.M.; Song, Y.; Liu, Q.; Xu, R.; Nie, L.; Han, W.; Wang, C. Geochemical Quantitative Assessment of Mineral Resource Potential in the Da Hinggan Mountains in Inner Mongolia, China. *Minerals* **2022**, *12*, 434. [[CrossRef](#)]
22. Brown, W.; Groves, D.; Gedeon, T. Use of Fuzzy Membership Input Layers to Combine Subjective Geological Knowledge and Empirical Data in a Neural Network Method for Mineral-Potential Mapping. *Nat. Resour. Res.* **2003**, *3*, 183–200. [[CrossRef](#)]
23. Kim, Y.H.; Choe, K.U.; Ri, R.K. Application of fuzzy logic and geometric average: A Cu sulfide deposits potential mapping case study from Kapsan Basin, DPR Korea. *Ore Geol. Rev.* **2019**, *107*, 239–247. [[CrossRef](#)]
24. Xiao, K.Y.; Zhang, X.H.; Song, G.Y.; Chen, Z.H.; Liu, D.L.; Wang, S.L. Development of GIS-Based Mineral Resources Assessment System. *Earth Sci.* **1999**, *5*, 525–528, (In Chinese with English abstract). [[CrossRef](#)]
25. Cui, C.Q.; Wang, B.; Zhao, Y.X.; Wang, Q.; Sun, Z.M. China's regional sustainability assessment on mineral resources: Results from an improved analytic hierarchy process-based normal cloud model. *J. Clean. Prod.* **2019**, *210*, 105–120. [[CrossRef](#)]
26. Karapurkar, D.D. RS and GIS based studies on Sediment yield from a tropical watershed: A case study of the Gangolli Catchment, Karnataka. In Proceedings of the Sedimentation, Tectonics, Mineral Resources and Sustainable Development, Hyderabad, India, 7–8 November 2019.
27. Xie, S.Y.; Wan, X.; Dong, J.B.; Wan, N.; Jiang, X.N.; Carranza, E.J.M.; Wang, X.Q.; Chang, L.H.; Tian, Y. Quantitative Prediction of Potential Areas Likely to Yield Se-rich and Cd-low Rice using Fuzzy Weights-of-Evidence Method. *Sci. Total Environ.* **2023**, *889*, 164015. [[CrossRef](#)] [[PubMed](#)]
28. Greff, K.; Srivastava, R.K.; Koutnik, J.; Steunebrink, B.R.; Schmidhuber, J. LSTM: A Search Space Odyssey. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *10*, 2222–2232. [[CrossRef](#)] [[PubMed](#)]
29. Agterberg, F. *Geomathematics: Theoretical Foundations, Applications and Future Developments*; Springer: Cham, Switzerland, 2014; p. 552.
30. Zhou, Y.Z.; Chen, S.; Zhang, Q.; Xiao, F.; Wang, S.G. Advances and Prospects of Big Data and Mathematical Geoscience. *Acta Petrol. Sin.* **2018**, *2*, 255–263.
31. Rodriguez-Galiano, V.; Sanchez-Castillo, M.; Chica-Olmo, M.; Chica-Rivas, M. Machine Learning Predictive Models for Mineral Prospectivity: An Evaluation of Neural Networks, Random Forest, Regression Trees and Support Vector Machines. *Ore Geol. Rev.* **2015**, *71*, 804–818. [[CrossRef](#)]
32. Sun, T.; Chen, F.; Zhong, L.X.; Liu, W.M.; Wang, Y. GIS-based mineral prospectivity mapping using machine learning methods: A case study from Tongling ore district, eastern China. *Ore Geol. Rev.* **2019**, *109*, 26–49. [[CrossRef](#)]
33. Nahool, T.A.; Anwar, M.; Yahya, G.A. Utilization of the random forest method for studying some heavy mesons spectra via machine learning technique. *Int. J. Mod. Phys. A Part. Fields Grav. Cosmol.* **2022**, *37*, 2250219. [[CrossRef](#)]
34. Beucher, A.; Siemssen, R.; Fröjdö, S.; Österholm, P.; Martinkauppi, A.; Edén, P. Artificial Neural Network for Mapping and Characterization of Acid Sulfate Soils: Application to Sirppujoki River Catchment, Southwestern Finland. *Geoderma* **2015**, *247–248*, 38–50. [[CrossRef](#)]
35. Herrera, M.; Torgo, L.; Izquierdo, J.; Pérez-García, R. Predictive models for forecasting hourly urban water demand. *J. Hydrol.* **2010**, *1–2*, 141–150. [[CrossRef](#)]
36. Xie, S.Y.; Huang, N.; Deng, J.; Wu, S.L.; Zhan, M.G.; Carranza, E.J.M.; Zhang, Y.P.; Meng, F.X. Quantitative Prediction of Prospectivity for Pb–Zn Deposits in Guangxi (China) by Back-propagation Neural Network and Fuzzy Weights-of-Evidence Modeling. *Geochem. Explor. Environ. Anal.* **2022**, *22*, 1–10. [[CrossRef](#)]
37. Rodriguez-Galiano, V.F.; Ghimire, B.; Rogan, J.; Chica-Olmo, M.; Rigol-Sanchez, J.P. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS J. Photogramm. Remote Sens.* **2012**, *67*, 93–104. [[CrossRef](#)]
38. Dong, L.H.; Zhang, L.C.; Li, W.D. Division and characteristics of geotectonic units in Xinjiang. In Proceedings of the 6th Tianshan Geological and Mineral Resources Symposium, Urumqi, China, 1 January 2008; pp. 25–32. (In Chinese).
39. Pan, Y.S. Formation and Uplifting of the Qinghai-Tibet Plateau. *Earth Sci. Front.* **1999**, *3*, 153–163, (In Chinese with English abstract). [[CrossRef](#)]
40. Liu, C.Y.; Liu, T. Discovery and Significance of Porphyritic Copper Mineralization in YunwuNing of XinJiang. *Xinjiang Geol.* **1998**, *2*, 185–187, (In Chinese with English abstract).
41. Zhang, Y.P.; Ye, X.F.; Xie, S.Y.; Zhou, X.Y.; Awadelseid, S.F.; Yaisamut, O.; Meng, F.X. Implication of multifractal analysis for quantitative evaluation of mineral resources in the Central Kunlun area, Xinjiang, China. *Geochem. Explor. Environ. Anal.* **2022**, *22*, geochem2021-083. [[CrossRef](#)]
42. Breiman, L. Random forests. *Mach. Learn.* **2001**, *1*, 5–32. [[CrossRef](#)]
43. Martins, T.F.; Seoane, J.C.S.; Tavares, F.M. Cu-Au exploration target generation in the eastern Carajás Mineral Province using random forest and multi-class index overlay mapping. *J. S. Am. Earth Sci.* **2022**, *116*, 103790. [[CrossRef](#)]
44. Oshiro, T.M.; Perez, P.S.; Baranauskas, J.A. How Many Trees in a Random Forest? In *Machine Learning and Data Mining in Pattern Recognition*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 154–168.
45. Liu, T.Y. EasyEnsemble and Feature Selection for Imbalance Data Sets. In Proceedings of the 2009 International Joint Conference on Bioinformatics, Systems Biology and Intelligent Computing, Shanghai, China, 3–5 August 2009; pp. 517–520.
46. Xiong, Y.; Zuo, R. Effects of misclassification costs on mapping mineral prospectivity. *Ore Geol. Rev.* **2017**, *82*, 1–9. [[CrossRef](#)]
47. Feng, K.; Hong, H.; Tang, K.; Wang, J. Decision making with machine learning and roc curves. *arXiv* **2019**, arXiv:1905.02810v1. [[CrossRef](#)]

48. Kim, E.; Kim, W.; Lee, Y. Combination of multiple classifiers for the customer's purchase behavior prediction. *Decis. Support Syst.* **2003**, *2*, 167–175. [[CrossRef](#)]
49. Cui, X.L.; Liu, T.T.; Wang, W.H.; Jing, M.; Bai, Y. Characteristics of geochemistry and prospecting direction of stream sediments in Buqingshan area, East Kunlun Mountains. *Geophys. Geochem. Explor.* **2011**, *35*, 573–578, (In Chinese with English abstract).
50. Guo, Y.; Gong, F.Z.; Ning, J.S.; Liu, Y.C.; Liu, Z. Comparative study of the content area fractal method and the traditional statistical method for determining the anomaly lower limit: A case study of Au element of stream sediment survey in Awengcuo area of Tibet. *Miner. Resour. Geol.* **2018**, *4*, 736–741, (In Chinese with English abstract).
51. Thanh, T.N.; Tuyen, D.V. Identification of Multivariate Geochemical Anomalies Using Spatial Autocorrelation Analysis and Robust Statistics. *Ore Geol. Rev.* **2019**, *111*, 102985. [[CrossRef](#)]
52. Feng, C.Y.; Wang, S.; Li, G.C.; Ma, S.C.; Li, D.S. Middle to Late Triassic granitoids in the Qimantage area, Qinghai Province, China: Chronology, geochemistry and metallogenic significances. *Acta Petrol. Sin.* **2012**, *28*, 665–678, (In Chinese with English abstract).
53. Guo, Z.F.; Deng, J.F.; Xu, Z.Q.; Mo, X.X.; Luo, Z.H. Late Palaeozoic-Mesozoic Intracontinental, Orogenic Process and Inter Medate-Acidic Igneous Rocks from the Eastern KunLun Mountains of NorthWestern China. *Geoscience* **1998**, *3*, 51–59, (In Chinese with English abstract).
54. Zheng, M.T.; Zhang, L.C.; Zhu, M.T.; Li, Z.Q.; He, L.D.; Shi, Y.J.; Dong, L.H.; Feng, J. Geological characteristics, formation age and genesis of the Kalaizi Ba-Fe deposit in West Kunlun. *Earth Sci. Front.* **2016**, *5*, 252–265, (In Chinese with English abstract). [[CrossRef](#)]
55. Omar, L.; Ivrisimtzis, I. Using theoretical ROC curves for analysing machine learning binary classifiers. *Pattern Recognit. Lett.* **2019**, *128*, 447–451. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.