

# Mineral prospectivity mapping of tungsten polymetallic deposits using machine learning algorithms and comparison of their performance in the Gannan region, China

Yue Liu<sup>1</sup> and Lou Yonghang<sup>1</sup>

<sup>1</sup>China University of Geosciences

November 22, 2022

## Abstract

The current study aimed at assessing the capabilities of five machine learning models in term of mapping tungsten polymetallic prospectivity in the Gannan region of China. The five models include logistic regression (LR), support vector machine (SVM), random forest (RF), convolutional neural network (CNN), and light gradient boosting machine (LGBM) models. Geochemical, lithostratigraphic, and structural datasets were used to generate 16 evidential maps, which were integrated into the machine learning models. Tungsten polymetallic deposits were randomly separated into two parts: 80% for training and 20% for validating. Performances of the models were evaluated through receiver operating characteristic (ROC) and K-fold cross validation, with an emphasis on the variable influence within different machine learning methods. The results show that the models are especially sensitive to the chemical elements: Be, Bi, Pb and Cd, implying that these are closely related to tungsten polymetallic mineralization. Compared to other models, the LGBM and CNN models performed best, while the LR model was the most stable. The results also indicated that the CNN model can predict maximum known deposits within a minimum area, based on the prediction-area plot analysis of the five models, while the RF model can capture the most well-known deposits within the smallest study area. Finally, eighteen prospective areas were delineated according to the predicting results of the machine learning models, which will provide important guidance for further tungsten polymetallic exploration and associated studies.

1 **Mineral prospectivity mapping of tungsten polymetallic deposits using machine learning algorithms**  
2 **and comparison of their performance in the Gannan region, China**

3 **Yonghang Lou, Yue Liu\***

4 Faculty of Earth Resources; State Key Laboratory of Geological Process and Mineral Resources, China University  
5 of Geosciences, Wuhan 430074, China

6 Corresponding author: Yue Liu ([liyue7870@cug.edu.cn](mailto:liyue7870@cug.edu.cn))

7 **Key Points:**

- 8 • Five machine learning models were used and compared for tungsten prospectivity analysis in the Gannan  
9 region, China
- 10 • Performances of the machine learning models were evaluated through ROC, P-A plot and K-fold cross  
11 validation
- 12 • CNN and LGBM models are more efficient in mapping favorable areas for tungsten exploration.

13

14

## 15 **Abstract**

16 The current study aimed at assessing the capabilities of five machine learning models in term of mapping tungsten  
17 polymetallic prospectivity in the Gannan region of China. The five models include logistic regression (LR), support  
18 vector machine (SVM), random forest (RF), convolutional neural network (CNN), and light gradient boosting machine  
19 (LGBM) models. Geochemical, lithostratigraphic, and structural datasets were used to generate 16 evidential maps,  
20 which were integrated into the machine learning models. Tungsten polymetallic deposits were randomly separated  
21 into two parts: 80% for training and 20% for validating. Performances of the models were evaluated through receiver  
22 operating characteristic (ROC) and K-fold cross validation, with an emphasis on the variable influence within different  
23 machine learning methods. The results show that the models are especially sensitive to the chemical elements: Be, Bi,  
24 Pb and Cd, implying that these are closely related to tungsten polymetallic mineralization. Compared to other models,  
25 the LGBM and CNN models performed best, while the LR model was the most stable. The results also indicated that  
26 the CNN model can predict maximum known deposits within a minimum area, based on the prediction-area plot  
27 analysis of the five models, while the RF model can capture the most well-known deposits within the smallest study  
28 area. Finally, eighteen prospective areas were delineated according to the predicting results of the machine learning  
29 models, which will provide important guidance for further tungsten polymetallic exploration and associated studies.

## 30 **1 Introduction**

31 Mineral prospectivity mapping (MPM) is concerned with quantifying and mapping the likelihood that mineral  
32 deposits are present at a certain location, which require the application of diverse methods and techniques to integrate  
33 multi-sources spatial geoscience datasets (Carranza 2009, 2017). The application of machine learning methods for  
34 MPM improves efficiency and accuracy of mineral exploration and has become a major trend (Li et al., 2020; Wang  
35 et al., 2020a; Liu et al. 2022). Machine learning is an important branch of artificial intelligence, and has a strong ability  
36 for mineral prediction. With the introduction of big data technology and artificial intelligence into geosciences, various  
37 machine learning and deep learning methods have been developed for quantitative assessment and modeling of mineral  
38 resources, such as logistic regression (LR) (Xiong and Zuo, 2018), support vector machine (SVM) (Maepa et al., 2021;  
39 Ghezelbash et al., 2021), decision tree (Sun et al., 2022), random forest (RF) (Carranza & Laborte 2015a, b; Sun et  
40 al., 2019), convolutional neural network (CNN) (Li et al., 2022), maximum entropy (Liu et al., 2018), Boltzmann  
41 machine (Chen, 2015), and extreme learning machine regression (Chen & Wu, 2017). These methods provide

42 powerful in modeling of mineral prospectivity by handling complex and nonlinear relationships among multisource  
43 geospatial datasets. Many successful case studies have made the use of machine learning the most typical data mining  
44 paradigm in mineral prospectivity analysis.

45 In the present study, five advanced machine learning methods including LR, SVM, RF, CNN and the optical  
46 gradient propulsion machine (LGBM), were introduced for tungsten polymetallic prospectivity mapping in the  
47 Gannan region. The LR is a traditional machine learning method that can deal with linear relationships between  
48 evidential variable, while the relationships between geological layers are nonlinear. When the LR model is combined  
49 with other machine learning algorithms, its performance can be improved as demonstrated by previous studies (Chen  
50 & Zhao, 2021; Kost et al., 2021; Sinaice et al., 2021). The SVM method based on Vapnik-Chervonenkis (VC)  
51 dimension theory and the principle of structural risk minimization, aims to find the global optimal solution, balance  
52 the complexity and fitting accuracy of the model, and avoid over-fitting problems (Zhou et al., 2017; Meng et al.,  
53 2019; Ghezelbash et al., 2019; Wang et al., 2020a). The RF is a popular machine leaning method that is frequently  
54 used frequently for mineral prospectivity analysis mainly due to the following advantages: it is relatively robust to  
55 outliers; it combines continuous and categorical data as variable inputs; and it can overcome black-box limitations of  
56 artificial neural networks (Rodriguez et al., 2015; Youssef et al., 2016; Wang et al., 2020b; Xiang et al., 2020; Parsa  
57 & Maghsoudi, 2021). The CNN is an efficient deep learning method, which can process complex data through its high  
58 capacity of feature learning and convolutional filtering (Lecun et al., 1998; Li et al., 2021). The LGBM is a new  
59 machine learning method that has a fast running speed, low memory consumption, and is widely used in biology,  
60 medicine, chemistry, social science (Sun et al., 2020b; Zafari et al., 2020; Ou et al., 2022; Vaulet et al., 2022), and  
61 lithofacies classification. However, it is rarely used to integrate multisource geospatial data for MPM.

62 The Gannan region is situated at the southern part of the Jiangxi Province and the eastern part of the Nanling  
63 structural belt, where endows with advantageous metallogenic geological conditions and have discovered a large  
64 number of tungsten polymetallic resources (Yu et al., 2010; Chen et al., 2018; Legros et al., 2020; Zhang et al., 2021).  
65 Mesozoic granites have been studied extensively in the Gannan region owing to their close relationships with tungsten  
66 polymetallic mineralization (Yu et al., 2010; Shu et al., 2011; Liu et al., 2013, 2015; Wang et al., 2017; Zhang et al.,  
67 2017a; Guo et al., 2020; Wang et al., 2020c; Cao et al., 2021; Wu et al., 2021; Li, T., 2022). The study aimed to  
68 compare the performance and robustness of the five machine learning algorithms, and further, to determine the most  
69 appropriate model in terms of mapping favorable areas for tungsten polymetallic exploration in the Gannan region.

70 The performances of the individual machine learning models were evaluated by multiple performance metrics  
71 including receiver operating characteristic (ROC) curve, K-fold cross validation, and prediction–area (P-A) plot  
72 analysis. Additionally, the influence of the variables was compared to determine the optimum ore-controlling variables.  
73 Three types of datasets were considered in this study (lithostratigraphic, fault, and geochemistry data), from which 16  
74 variables were derived to construct machine learning models. In addition, these predicting maps obtained from the  
75 five machine learning models were used to identify and delineate favorable areas in the study area, so that decision  
76 makers can grasp the basic information of regional mineral resource potential and make a better plan for future  
77 exploration and development.

## 78 **2 Study Area**

### 79 **2.1 Geological Setting**

80 Gannan region is located in the eastern part of the Nanling structural belt has undergone complex tectonic-  
81 magmatic activities, and is dominated by large folds and faults that occurred mainly in the early Paleozoic, Triassic,  
82 and late Jurassic eras (Figure 1). In the early Paleozoic and Triassic eras, the region was affected by the Tethys domain  
83 and Hercynian Indosinian orogeny, resulting in the development of NW-SE and S-N trending faults. In the Late  
84 Triassic, the continental collision between the North and South China blocks led to the formation of a new subduction  
85 system, and the formation of larger-scale W-Sn deposits in the Gannan metallogenic belt (Mao et al., 2013a, Yuan et  
86 al., 2019). In the late Jurassic, the region was affected by the Yanshan movement, and the fold changed to NE trending.  
87 Concurrently, the region was subducted and compressed by the Pacific plate and intracontinental orogeny, and finally  
88 formed a basin-mountain system where the base alternates with granitic volcanic intrusive rocks.

89 The strata of the Gannan region are composed of Sinian to Quaternary strata. The deepest sedimentary strata  
90 consist of feldspar quartz sandstone, siltstone, silty slate, phyllite, and shale ranging from Sinian to Silurian,  
91 characterized by a super thick sub-deep sea flysch clastic rock structure. Metamorphic rock turbidite sediments exist  
92 from the Ordovician period. The strata cover a large area of Devonian~Triassic carbonate rocks and sandstones, and  
93 develop coastal facies, shallow sea facies, and marine continental facies. The stratum is thinner than bedrock as a  
94 caprock. The Sinian to Ordovician and Devonian strata are considered source beds of Yanshanian granite tungsten  
95 deposits owing to high tungsten content (Chen et al., 2018). Above these strata, Jurassic~Cretaceous clastic and  
96 volcanic rocks and red beds were deposited in rift basins (Xu, X. S., 2005; Cao et al., 2018; Jiang et al., 2018). However,

97 because of Caledonian uplift strata, the Silurian strata in Gannan were denuded and has therefore been removed over  
98 time.

99 During the Caledonian, Hercynian, Indosinian, and Yanshanian periods, three magmatic intrusion cycles  
100 occurred in Gannan. The Yanshanian magmatic intrusion cycle was the most frequent, forming granite widely in  
101 Gannan. This mineralization is related to Mesozoic granitoid magmatism (e.g., Chen et al., 2013; Mao et al., 2013b;  
102 Cai et al., 2017; Xie et al., 2018). Scholars have found that the Mesozoic granite is closely related to mineral resources  
103 such as tungsten, tin, molybdenum, bismuth, copper, lead, zinc, silver, rare earth, rare metals, and uranium (Wang et  
104 al., 2020b). In the Gannan region, granites are mainly acid granites, with few intermediate acid and basic rocks (mainly  
105 monzogranite, and porphyritic, biotite, and migmatite granites). The main minerals are quartz, potassium feldspar and  
106 plagioclase (Chen et al., 2018). The granites in the Nanling area show multi-stage characteristics. In the Caledonian  
107 period, the study area predominantly exposed the Ordovician and Silurian monzogranite, biotite and migmatite granite.  
108 The study area exposes Indosinian Middle Triassic monzogranite, in a small distribution area. In the Yanshanian  
109 period, the area exposed Cretaceous and Jurassic biotite granite and monzogranite, with the largest distribution area.

## 110 **2.2 Tungsten polymetallic mineralization**

111 The Nanling belt is one of the most important tungsten polymetallic metallogenic belts in the world, and forms  
112 many ore deposits, such as Dachang, Shizhuyuan, Qitianling, Dajishan, and Xihuashan deposits (Hu & Zhou, 2012;  
113 Liu et al., 2013, 2014). Tungsten polymetallic mineralization in the region was mainly controlled by fold, fault, and  
114 multi-stage magmatic activities. Ore-controlling faults are generally NE and EW trending. Tungsten mineralized  
115 bodies are mainly skarn type (skarn froze complex type) and quartz vein type (Mao et al., 2011; Hu & Zhou, 2012;  
116 Chen et al., 2013). Spatially, tungsten polymetallic deposits are mainly distributed in the contacts of the late Jurassic  
117 granite intrusions and its surrounding rocks (such as siliciclastic rock and carbonate rock) (Liu et al. 2013, 2014, 2015).  
118 Quartz vein-type tungsten polymetallic deposits which occur in the Gannan region are closely related to Yanshanian  
119 granites (Chen et al., 2013; Xiong et al., 2020). Metal minerals are mainly predominantly wolframite, cassiterite,  
120 molybdenite, bismuth pyroxene, scheelite, pyrite, cinnabar, and chalcopyrite (Cao et al., 2021). Ore minerals are  
121 predominantly quartz, fluorite, feldspar, wollastonite, topaz, sericite, muscovite, and calcite. Associated beneficial  
122 components usually contain tin, molybdenum, bismuth, plumbum, and silver. The skarn type tungsten polymetallic  
123 ore body is strictly controlled by the contact zone of the rock mass, and usually occurs intermittently along the direction  
124 of the contact zone. Large-scale wide bodies are often developed in the concave of the rock mass. The main ore

125 minerals are scheelite, galena, and sphalerite, and the secondary metal minerals are molybdenite, bismuth pyroxene,  
126 pyrite, magnetite, and chalcopyrite.

127 The Yanshanian and Indosinian periods were the main stage of Tungsten polymetallic mineralization in the  
128 Nanling belt (Liu et al, 2014). The Jurassic era is the most important period of formation for the tungsten polymetallic  
129 deposits (Guo et al., 2020). The Gannan region has widespread Jurassic batholiths including granodiorite, biotite  
130 granite, two-mica granite, A-type granite, minor gabbro, and syenite (Zhang et al., 2017b). The Yanshanian ore-  
131 forming event has typical characteristics of a short duration in Gannan. Therefore, in this area, mineralization  
132 processes show unexpected, unique, and complex characteristics.

### 133 **3 Materials and Data Preprocessing**

134 The datasets used in this study came from the Chinese Geological Survey (CGS), including stream sediment  
135 geochemical data, rock mass, faults, and tungsten polymetallic deposits. An overview of the MPM procedures is  
136 shown in Figure 2.

#### 137 **3.1 Geochemical criteria**

138 In terms of geochemical elements, there are many geochemical elements related to tungsten polymetallic  
139 deposits. In this study, the factor load matrix of geochemical elements was obtained by factor analysis (FA) of the  
140 geochemical data and can obtain the biplot of component 1 (F1) vs. component 2 (F2) (Figure 3).

141 Figure 3 shows that elements with large component 2 values have a high correlation with mineralization. A total  
142 of 11 chemical elements (Be, Bi, F, Li, Pb, Sn, Zn, W, Y, Cd and U) with values greater than 0.5 (considered to be  
143 more important in the same composition) were extracted to construct the dataset. Most of the extracted geochemical  
144 elements were similar to previous research results (Liu et al., 2013; Wang et al., 2020c), demonstrating that the  
145 geochemical element data were reliable and realistic. The 11 selected elements were mapped by means of an inverse  
146 distance weighted (IDW) interpolation method, on a 1x1 km cell resolution (Figure 4).

#### 147 **3.2 Granite mass**

148 Mineralization related to magmatic activity in the Nanling region is of a long-term nature and inheritance (Wang  
149 et al., 2020b). Tungsten polymetallic deposits are closely related to the distribution of magmatic rocks in the study  
150 area. The widely distributed granite in the study area is mainly from the Yanshanian period. Genetically, most of the  
151 Nanling region W-Sn-polymetallic deposits have a relationship with Mesozoic granitoids and can be divided into 10

152 types (Wang et al., 2020b). Using a physically constrained variational automatic encoder (VAE), Xiong et al. (2021)  
153 identified geochemical patterns associated with tungsten polymetallic mineralization in the Nanling Mountain Range.  
154 The results also indicated that the formation of deposits is closely related to granite intrusions. Using singularity  
155 analysis, Chen et al. (2015) identified a weak gravity anomaly caused by buried granites in the Nanling region, which  
156 is closely related to the spatial distribution of tungsten polymetallic deposits. These granitic bodies form a system in  
157 the Earth's crust that supplies sufficient material and energy for ore deposit formation. The closer the contact zone of  
158 rock mass, the higher the possibility of mineral formation. According to the distributions of granite rock mass during  
159 the Caledonian, Yanshanian and Indosinian geological movements, granitic rocks were separately buffered outside  
160 into eight buffer bands with 1 km interval.

### 161 **3.3 Faults**

162 Owing to strong tectonic folding in the Yanshanian period, the NE trending faults and fault systems have  
163 developed in the study area. The regional distribution, occurrence, shape, and scale of tungsten polymetallic deposits  
164 were strictly controlled by the fault structures of this period and to the tectonic activities in the region. The mineral  
165 resources are enriched in the areas where the structures developed along the direction of the faults. Chen et al. (2015)  
166 found that linear structures, including bends and intersections, are closely related to the spatial distribution of deposits,  
167 because most deposits occur along these linear structures. Structure is the main influencing factor on the mineralization  
168 of tungsten polymetallic deposits. Moreover, at the fault junctions, structural stress is complex and diverse, and likely  
169 to form minerals. High density zones of fault intersections and lineaments have favorable physical characteristics for  
170 hosting ore-forming materials. Thus, fault lineaments were buffered into eight buffer bands with 1 km interval (Figure  
171 5a) and fault intersection density was produced using ArcGIS software (Figure 5b).

## 172 **4 Methods**

### 173 **4.1 Support vector machine**

174 To meet the criterion of structural risk minimization, the SVM method maps data from low to high dimensions,  
175 and then locates the line, curve, or hyperplane that can optimally distinguish distinct categories in the data.

176 There are two types of SVMs: linear support and nonlinear support vector machines. Linear SVMs look for the  
177 optimal line that can divide data features in the feature space (Figure 6a). The premise of this method is that the data  
178 is linearly separable, as follows:

179 
$$f(x) = (\vec{w}, x) + b \tag{1}$$

180 where  $\vec{w}$  is the normal vector of the classification line,  $b$  is the constant term of the classification line, and  
 181  $h=2//\vec{w} //$  is the classification interval. The points on the upper and lower sides on Figure 6a represents positive and  
 182 negative samples. The black bold line is the optimal classification line found in theory, and the points passed by the  
 183 two dotted lines on the upper and lower sides are support vectors. By solving the relationship between  $y$  and  $(\vec{w}, x) +$   
 184  $b$ , and maximizing the classification interval  $h$ , the accuracy of the data feature division by the optimal classification  
 185 line can be high. However, most data used in classification problems do not conform to linear separability in practical  
 186 scenarios. Hence, there is a nonlinear SVM. Through nonlinear transformation, the input variable  $x$  is mapped into the  
 187 high-dimensional space to satisfy the linear segmentation condition and calculate the optimal classification surface  
 188 (Figure 6b). The kernel function can implement this data mapping transformation, through the widely used Polynomial  
 189 Kernel Function (Poly), Radial Basis Function (RBF), Sigmoid Kernel Function (Sigmoid), and radial basis function.

190 **4.2 Logistic regression**

191 The method of logistic regression uses continuous and discrete data, and qualitative dependent variables. The  
 192 principle of this method is that the result of linear regression  $(-\infty, \infty)$  is mapped to  $(0,1)$  by a logical function. The  
 193 linear regression function is  $y$ , where  $\theta$  represents the different contributions of each variable  $x$ , to the dependent  
 194 variable  $y$ . However, the linear regression function can only be used with the quantitative dependent variable. When  
 195 the dependent variable is qualitative, it is no longer applicable, and the logical regression method overcomes this,  
 196 using the following formula:

197 
$$y = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_3 + \dots + \theta_n X_n = \theta^T x \tag{2}$$

198 The Sigmoid function is its logical function, which has an ‘‘S curve,’’ and has a value range of 0–1. When the  
 199 independent variable  $X$  approaches positive infinity, the dependent variable approaches 1, and  $X$  approaches negative  
 200 infinity, the dependent variable approaches 0. In binary classification, the sigmoid function is used to output the event  
 201 probability,  $P$  is the positive class probability of an event, and  $1-P$  is its negative class probability. This is written as  
 202  $g(x)$ , through logit transformation:

203 
$$g(x) = \frac{1}{1+e^{-x}} \tag{3}$$

204 Therefore, the functional expression of logistic regression (formula 2) into formula 3 to obtain formula 4,  
 205 constitutes the logistic regression function:

206

207

$$\begin{cases} y = \theta^T x \\ g(x) = \frac{1}{1+e^{-x}} \end{cases} \quad (4)$$

### 208 **4.3 Random Forest**

209 A bagging algorithm, which was proposed by Beriman (1996), is a method of sample calculations with return.  
210 Through  $m$  random samples of a dataset,  $m$  new training sets are formed,  $m$  weak learners are established, and finally  
211 set into a strong learner (Figure 7). This ensures that each data sample in the initial data exists in the new dataset and  
212 model. Concurrently, the unselected data samples contain  $m$  out of bag data (OOB), which is used to evaluate the  
213 constructed model.

214 The predecessor of RF is decision tree (DT) model. It selects the data features according to certain conditions to  
215 achieve the goal, and it is easy to produce the problem of over-fitting. The RF model can be considered as a set of  
216 decision trees which are established randomly, the results of each decision tree are finally synthesized into a single  
217 result by means of averaging, voting and majority results (Figure 8). The relationships between the trees are  
218 independent with each other. It randomly takes samples from the dataset, each having its own characteristics. The  
219 samples are diverse and, therefore, the results are different. It only uses some data features from the randomly selected  
220 samples. The method adopts the highly accurate bagging integration algorithm that does not cause over-fitting, and  
221 strongly resists the interference of outliers. When the feature dimension is large, because of the unique nature between  
222 samples, the RF method can process data efficiently and maintain high accuracy.

### 223 **4.4 Convolutional neural network**

224 Lecun (1998) proposed the convolution neural network (CNN), LeNet-5, and achieved superior results. A  
225 convolutional neural network has the basic characteristics of local connection; the nodes of the convolution layer are  
226 only connected with some nodes of the previous layer, and are only used to learn local features (Figure 9a). This  
227 increases learning rate speed and avoids overfitting problems. Additionally, the extracted data feature is independent  
228 of the data position. It also has aspects of weight sharing, which means that a parameter in the convolution kernel is  
229 shared in the same convolution layer (Figure 9b). Thus, the parameters on the convolution layer can be reduced.

230 This CNN includes input, convolution, pooling, full connection, and output layers (Figure 10). It performs a  
231 convolution operation on the convolution layer, which is the core operation of a CNN. Using a fixed size convolution  
232 kernel (such as  $3 \times 3$ ) to enhance input image data features (horizontally and vertically), the feature map is obtained,

233 and  $n$  data maps can be obtained by scanning and recalculating (Figure 9c). Next the activation function is introduced  
234 for map operation, which solves nonlinear problems and increases nonlinear expression abilities (also known as the  
235 incentive layer). The appropriate activation function can avoid gradient disappearance or explosion problems. An  
236 increase in convolution layers increases network speed and accuracy. Common nonlinear activation functions include  
237 the Sigmoid, Tanh, and Reluactivation functions. Additionally, pooling operations can reduce model parameters and  
238 complexity. There are two common pooling operations (Figure 9d). The first is maximum pooling, which is the  
239 maximum value in the local range of the characteristic map as the pooling result. The second is average pooling, which  
240 outputs the average characteristics of the local region. Average pooling can collect global features, while maximum  
241 pooling can extract the most significant local features, suppress background interference, and improve the invariance  
242 of the network to the translation and geometric deformation of the input feature map. When fully connected, every  
243 layer is connected to all nodes in the previous layer, and learnt data features from the convolution and pooled layers  
244 are integrated, and then mapped to the sample marker location.

#### 245 **4.5 Light Gradient Boosting Machine**

246 The LGBM is an open source and efficient Gradient Boosting Decision Tree (GBDT) algorithm released by  
247 Microsoft in 2017. It is an optimization of the Extreme Gradient Boosting (XGBoost) algorithm. The LGBM optimizes  
248 the optimal split point search strategy, leaf growth strategy, gradient sampling method, feature attribute binding  
249 method, support class feature computation, and parallel learning of the decision tree. This reduces the training time  
250 and memory consumption of the model, while ensuring prediction accuracy. This algorithm is iterative and after  
251 several iterations, each weak classifier is weighted to become a stronger classifier (Figure 11).

252 The LGBM is an improvement on the XGBoost and GBDT as follows: (1) The LGBM uses histogram methods  
253 to process data, converting continuous values into discrete values, and increase processing speed. It divides the  
254 eigenvalues into “bins,” and converts them into  $n$  integers, to obtain a histogram with a width of  $n$ , and uses it as index.  
255 Then, histograms accumulate the required statistical data, and identify the optimal segmentation point. (2) Generally,  
256 the growth of the decision tree is based on a level-wise strategy, which divides all the leaf nodes in the same layer,  
257 while traversing the data (Figure 12a). However, when dealing with a large number of complex datasets, this algorithm  
258 cannot consider leaf nodes with very little information. The LGBM adopts the leaf growth method with depth limits.  
259 This selects the leaf node with the most information after splitting each time, and divides the leaf node further  
260 (Figure 12b). This improves accuracy by reducing the splitting of the nodes with little information, thereby reducing

261 model error. (3) The LGBM model uses a Gradient-based One-Side Sampling (Goss) algorithm in the sampling  
262 process. The considerable information of large gradient samples plays an important role in histogram construction.  
263 The Goss algorithm randomly samples only the small gradient samples proportionally, only if the distribution structure  
264 of the data is invariable. (4) Based on the Exclusive Feature Bundling (EFB) algorithm, the LGBM model can reduce  
265 the attribute dimensions and feature number used to construct the histogram. The complexity of the calculation is also  
266 simplified.

## 267 **5 Results and discussion**

### 268 **5.1 Model prediction and mineral mapping**

269 In this study, the obtained datasets were used in the model for training, and the models were debugged by means  
270 of hyperparameter optimization and grid search in the Python software. A total of 93 tungsten polymetallic deposit  
271 and 71, 179 non-deposit locations were used in the study. The datasets were randomly divided into 80% (for training)  
272 and 20% (for tests). Five trained machine models were used to predict the actual data in the study area. The probability  
273 of the models predicted identification of tungsten polymetallic deposits was imported into ArcGIS software for  
274 prospectivity mapping (Figure 13). The potential targets show a clear spatial relationship with the distribution of  
275 known tungsten polymetallic deposits. However, there are still some regions where this do not occur. The five models  
276 are fairly similar to the predicting areas of tungsten polymetallic deposits in the study area. The central and western  
277 parts of the study area were predicted as highly favorable areas for tungsten polymetallic mineralization almost  
278 simultaneously by the five models. However, in the north, there were clear differences among the five models.

### 279 **5.2 Model Assessment**

280 Receiver Operating Characteristic (ROC) is a curve with false positive rate (FPR) as abscissa and true positive  
281 rate (TPR) as ordinate. It is frequently used to evaluate the performance of models in mineral prospectivity mapping  
282 (Liu et al. 2014a; Chen & Wu, 2016). Figure 14 indicates that overall, the AUC values of the five models performed  
283 well, with all AUC values  $>0.85$  indicated by the training data and validating data. Specifically, the CNN and LGBM  
284 had superior classification and prediction performance than the other three models. This is indicated by the ROC  
285 curves (AUC values  $>0.9$ ), while the LR model had the lowest performance (as indicated by the AUC value of 0.85).

286 The K-fold cross validation is the most common method used in machine learning to evaluate and optimize model  
287 accuracy. In this study, the training data were inputted into the five models for K-fold cross validation. The prediction

288 performance of the model was quantified and compared with the original AUC values (Figure 15). The K-fold AUC  
289 and original AUC values ranged from 0.84 to 0.93, which supports the effectiveness and rationality of the constructed  
290 data. The results showed that the K-fold AUC of the LGBM model was the highest (0.89), indicating the most reliable  
291 results; however, the SVM model had the lowest reliability (0.84). This method showed that the K-fold AUC values  
292 of all models were reduced compared with the original AUC values. The AUC value of the CNN model decreased the  
293 most, by 0.0811, showing poor stability and implying that it is the most sensitive model among the five models. The  
294 LR model was the most robust, with an AUC difference of only 0.0053.

### 295 **5.3 Variables contribution analysis**

296 The importance of the variables to each model differs. The model assigns different weightings to different  
297 variables, which also represents the degree of learning and variable sensitivity. The "Permutation Importance" method  
298 can change the data in one column in the test dataset and other columns remain unchanged. It then quantifies the  
299 change of the model prediction ability after the data is disrupted, keeping the parameters unchanged after the model  
300 training. This method was applied to the five models, and the degree of influence of 16 variables on the prediction  
301 results was obtained, which showed the sensitivity of the models to each variable (Figure 16).

302 All models were highly sensitive to the variable W, which is consistent with other studies. In the LR and SVM  
303 models, W, Sn, Cd, Bi, Li, Be and Pb had similar high influences on the predicted results, while Zn was relatively  
304 low. Similarly, the two models have similar recognition characteristics and sensitivity of the five models in granite  
305 rock mass data and fault data. Conversely, Caledonian granite, Yanshanian granite, and fault distribution had little or  
306 no influence in the RF model. The learning results of the five models for these variables are consistent with the  
307 previous research conclusion that the anomalies of W, Sn, Cd, Bi, Be, Pb, and other elements are closely related to  
308 the tungsten polymetallic deposits in southern Jiangxi (Liu et al., 2014a, 2019; Xiong et al., 2021). Additionally, these  
309 elements, specifically W and Sn, have similar geochemical properties and closely relate to the residual magma,  
310 resulting from partial melting of the continental crust (Liu et al, 2014a). Table 1 shows the calculated cumulative  
311 contribution rate of each type of variable for each model.

312 The geological chemical elements is the most dominant factor in all models, with a contribution rate >78% and,  
313 therefore, this factor has the largest impact on the prediction performance of each model. Among the five models, the  
314 RF model is the most sensitive to element data (cumulative contribution rate >97%), indicating that this data is  
315 imperative in this model. However, the model may have overlooked other information that has a large impact on the

316 results, in the process of learning the other two types of data (granite mass and faults). For all models, the influence  
317 of granite rock mass data and fault data was different. The combined proportion of these two in the RF model was the  
318 lowest (combined total 2.86%). The granite rock mass data and fault data had the highest influence in the SVM model,  
319 at 12.83% and 8.61%, respectively (combined total 21.44%). The distribution of these two types of data was similar  
320 in the CNN and LGBM models, and both types of data account for approximately 5%. In the LR model, granite data  
321 (8.07%) is slightly higher than fault data (4.98%). The contribution rates to the models from these two types of data  
322 depends on the models overall learning and data feature extraction. The RF model may reduce the proportion of these  
323 two variables excessively, which makes the result less reliable. The proportion distribution of the SVM model was  
324 logical.

325 However, Li et al. (2020b) used the “Synthetic Minority Over Sampling Technology” (SMOTE) method to deal  
326 with the imbalance of data classes in the Nanling Mountain Range. Subsequently, the RF model was used to predict  
327 the mineralization and achieved good results. However, Sun et al. (2020a) generated three datasets to test the impact  
328 of negative sample data using the RF model. It captured 66.95% of the known ore within 9% of the study area,  
329 indicating that an imbalance in class can have an extremely large impact. This shows that the RF model may have  
330 great disadvantages in dealing with class unbalanced data sets.

#### 331 **5.4 Prediction–area plot**

332 The P–A plot analysis was accepted to determine suitable threshold that can evaluate prediction rate of the  
333 tungsten polymetallic deposits with smaller area including more deposits. Many researchers have evaluated the model  
334 performance and quality through prediction–area (P-A) plots (e.g., Nezhad et al., 2017; Zhang et al. 2017a;  
335 Roshanravan et al., 2018; Liu et al., 2019). Yousefi et al. (2012, 2013) plotted and compared distribution against  
336 predicted percentage of known ore deposit. These plots are used to identify mineral deposits based on area size. The  
337 smaller the area, the higher the probability of accuracy (Yousefi & Carranza, 2015). The results show the intersection  
338 point of the two curves from the five models (Figure 17). The performances of the SVM, RF, CNN and LGBM models  
339 are similar; however, the LR model intersection point is closer to the center of the curve.

340 Using these plots, the proportion of study area corresponding to each intersection point, and the percentage of  
341 known ore occurrences were calculated (Table 2). The CNN model achieved the best results, with 81% of the known  
342 ore deposit occurrences identified, in approximately 19% of the study area, and 0.57% of known deposits in high

343 potential areas. The SVM model produced the worst results, identifying 73% of the known ore deposits, in  
344 approximately 27% of the study area, and only 0.36% of known deposits in high potential areas.

### 345 **5.5 Prospectivity analysis**

346 Based on the prediction results of five models, 18 high potential areas for tungsten polymetallic exploration were  
347 delineated as shown in the Figure 18. The results show that the areas 12 and 13 are the most important in the study  
348 area as they have the widest area and the clearest directivity among the high potential areas. Additionally, they have  
349 a high coincidence with the distribution of known deposits. This shows that the metallogenic potential of the two areas  
350 has strong reliability. The five models have high potential prediction effects, and most areas capture known deposits  
351 in the areas 3, 5 ~7, 9, 11, and 16 ~18. This shows that the potential effect of these areas has a certain reliability. For  
352 the areas 1, 2, 8, 10 and 14, the LR, SVM and RF models have similar predicting results. It was concluded that these  
353 areas are linearly distributed along the fault, showing a high spatial correlation with the faults. For area 15, in addition  
354 to CNN model, other four models predicted high potential metallogenic areas, which corresponds with known deposit  
355 areas. The LR, RF, and LGBM models predict that the area 4 is a high potential area for the formation of tungsten  
356 deposits.

### 357 **6 Conclusions**

- 358 (1) According to ore-controlling factors of tungsten polymetallic deposits in the Gannan region, sixteen evidential  
359 variables were produced from three types of datasets including geochemical, fault, and lithostratigraphic data,  
360 which were integrated into the LR, SVM, RF, CNN and LGBM models. Our studies indicate that the five models  
361 have strong sensitivity to geochemical elements Be, Bi, Pb and Cd in addition to W and Sn, which can be  
362 identified by the contribution rates of the variables, implying that these elements are closely related to tungsten  
363 polymetallic mineralization and have a symbiotic association system.
- 364 (2) The CNN and LGBM models are more efficient in mapping favorable areas for tungsten exploration in the study  
365 area based on comparative studies. The RF and CNN models performed well according to the P-A plot. However,  
366 the performance of the P-A plot for the LGBM model was poor, with a clear over-fitting phenomenon. The RF  
367 model assigns an overly high weighting to geochemical data in the learning variables process, while the fault  
368 data may be ignored.

369 (3) Based on the prediction results of the five models, the study area was divided into 18 favorable areas for tungsten  
370 exploration, of which 10 areas have higher reliability and show strong spatial relationships with known deposits.  
371 however, the remaining areas need to be investigated and validated further, because there were no clear, unified  
372 conclusions from the models.

### 373 **Acknowledgments**

374 This study was financially supported by the National Natural Science Foundation of China (42172325) and the  
375 Fundamental Research Funds for the Central Universities, China University of Geosciences (Wuhan) (CUG2106202).

### 376 **Author Contributions:**

377 Conceptualization: Yue Liu, Yonghang Lou

378 Methodology: Yonghang Lou, Yue Liu

379 Writing – original draft: Yonghang Lou

380 Writing – review & editing: Yue Liu

381 Supervision: Yue Liu

### 382 **Conflict of Interest**

383 The authors declare no conflicts of interest relevant to this study.

### 384 **Data Availability Statement**

385 The training and testing data sets for tungsten polymetallic prospectivity analysis based on machine learning methods  
386 are provided in the online data set [https://figshare.com/articles/dataset/Data\\_and\\_Python\\_code\\_rar/21120022](https://figshare.com/articles/dataset/Data_and_Python_code_rar/21120022).

### 387 **References**

388 Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123-140.

389 Cai, Y., Feng, Z., Shao, T., Hu, R., Zhou, Y., & Xu, J. (2017). New precise zircon U-Pb and muscovite <sup>40</sup>Ar- <sup>39</sup>Ar  
390 geochronology of the Late Cretaceous W-Sn mineralization in the Shanhu orefield, South China. *Ore Geology*  
391 *Reviews*, 84, 338-346.

392 Carranza, E. J. M. (2009). Controls on mineral deposit occurrence inferred from analysis of their spatial pattern and  
393 spatial association with geological features. *Ore Geology Reviews*, 35, 383–400.

394 Carranza, E. J. M. (2017). Natural resources research publications on geochemical anomaly and mineral potential  
395 mapping, and introduction to the special issue of papers in these fields. *Natural Resources Research*, 26(4), 379-  
396 410.

397 Carranza, E. J. M., & Laborte, A. G. (2015a). Random forest predictive modeling of mineral prospectivity with small  
398 number of prospects and data with missing values in Abra (Philippines). *Computers & Geosciences*, 74, 60-70.

399 Carranza, E. J. M., & Laborte, A. G. (2015b). Data-driven predictive mapping of gold prospectivity, Baguio district,  
400 Philippines: Application of Random Forests algorithm. *Ore Geology Reviews*, 71, 777-787.

401 Cao, J., Wu, Q., Yang, X., Kong, H., Li, H., Xi, X., Huang, Q., Liu, B. (2018). Geochronology and Genesis of the  
402 Xitian W-Sn Polymetallic Deposit in Eastern Hunan Province, South China: Evidence from Zircon U-Pb and  
403 Muscovite Ar-Ar Dating, Petrochemistry, and Wolframite Sr-Nd-Pb Isotopes. *Minerals*, 8(3).

404 Cao, J. Y., Li, H., Algeo, T. J., Yang, L. Z., & Tamehe, L. S. (2021). Two-stage magmatism and tungsten mineralization  
405 in the Nanling Range, South China: Evidence from the Jurassic Helukou deposit. *American Mineralogist*, 106(9),  
406 1488-1502.

407 Chen, G. X., Huang, N., Wu, G. P., Luo, L., Wang, D. T., & Cheng, Q. M. (2022). Mineral prospectivity mapping  
408 based on wavelet neural network and Monte Carlo simulations in the Nanling W-Sn metallogenic province. *Ore*  
409 *Geology Reviews*, 143.

410 Chen, G. X., Liu, T. Y., Sun, J. S., Cheng, Q. M., Sahoo, B., Zhang, Z. J., & Zhang, H. L. (2015). Gravity method for  
411 investigating the geological structures associated with W-Sn polymetallic deposits in the Nanling Range, China.  
412 *Journal of Applied Geophysics*, 120, 14-25.

413 Chen, J., Wang, R. C., Zhu, J. C., Lu, J. J., & Ma, D. S. (2013). Multiple-aged granitoids and related tungsten-tin  
414 mineralization in the Nanling Range, South China. *Science China-Earth Sciences*, 56(12), 2045-2055.

415 Chen, L.-L., Ni, P., Li, W.-S., Ding, J.-Y., Pan, J.-Y., Wang, G.-G., & Yang, Y.-L. (2018). The link between fluid  
416 evolution and vertical zonation at the Maoping tungsten deposit, Southern Jiangxi, China: Fluid inclusion and stable  
417 isotope evidence. *Journal of Geochemical Exploration*, 192, 18-32.

418 Chen, Y. L. (2015). Mineral potential mapping with a restricted Boltzmann machine. *Ore Geology Reviews*, 71, 749–  
419 760.

420 Chen, Y. L., & Wu, W. (2016). A prospecting cost-benefit strategy for mineral potential mapping based on ROC curve  
421 analysis. *Ore Geology Reviews*, 74, 26-38.

422 Chen, Y. L., & Wu, W. (2017). Mapping mineral prospectivity using an extreme learning machine regression. *Ore*  
423 *Geology Reviews*, 80, 200-213.

424 Chen, Y. L., & Zhao, Q. Y. (2021). Mineral exploration targeting by combination of recursive indicator elimination  
425 with the l(2-)regularization logistic regression based on geochemical data. *Ore Geology Reviews*, 135.

426 Fan, J. L., Ma, X., Wu, L. F., Zhang, F. C., Yu, X., & Zeng, W. Z. (2019). Light Gradient Boosting Machine: An  
427 efficient soft computing model for estimating daily reference evapotranspiration with local and external  
428 meteorological data. *Agricultural Water Management*, 225.

429 Ghezelbash, R., Maghsoudi, A., & Carranza, E. J. M. (2019). Performance evaluation of RBF-and SVM-based  
430 machine learning algorithms for predictive mineral prospectivity modeling: integration of SA multifractal model  
431 and mineralization controls. *Earth Science Informatics*, 12(3), 277-293.

432 Ghezelbash, R., Maghsoudi, A., Bigdeli, A., & Carranza, E. J. M. (2021). Regional-scale mineral prospectivity  
433 mapping: Support vector machines and an improved data-driven multi-criteria decision-making technique. *Natural*  
434 *Resources Research*, 30(3), 1977–2005.

435 Guo, X. F., Wu, K. X., Tian, Y., & Zhang, X. B. (2020). U-Pb geochronology and Raman spectroscopy of zircons from  
436 the granites in the Xihuashan and Tieshanlong Deposit: Implications for W-Sn mineralization in Southern Jiangxi  
437 Province, South China. *Resource Geology*, 70(2), 141-156.

438 Hu, R. Z., & Zhou, M. F. (2012). Multiple Mesozoic mineralization events in South China-an introduction to the  
439 thematic issue. *Mineralium Deposita*, 47(6), 579-588.

440 Jiang, W. C., Li, H., Evans, N. J., Wu, J. H., & Cao, J. Y. (2018). Metal Sources of World-Class Polymetallic W-Sn  
441 Skarns in the Nanling Range, South China: Granites versus Sedimentary Rocks? *Minerals*, 8(7).

442 Kost, S., Rheinbach, O., & Schaeben, H. (2021). Using logistic regression model selection towards interpretable  
443 machine learning in mineral prospectivity modeling. *Geochemistry*, 81(4).

444 Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition.  
445 *Proceedings of the Ieee*, 86(11), 2278-2324.

446 Legros, H., Harlaux, M., Mercadier, J., Romer, R. L., Poujol, M., Camacho, A., Marignac, C., Cuney, M., Wang, R.  
447 C., Charles, N., & Lespinasse, M. Y. (2020). The world-class Nanling metallogenic belt (Jiangxi, China): W and

448 Sn deposition at 160 Ma followed by 30 m.y. of hydrothermal metal redistribution. *Ore Geology Reviews*, 117.

449 Li, S., Chen, J. P., Liu, C., & Wang, Y. (2021). Mineral Prospectivity Prediction via Convolutional Neural Networks  
450 Based on Geological Big Data. *Journal of Earth Science*, 32(2), 327-347.

451 Li, T., Zuo, R. G., Zhao, X. F., & Zhao, K. D. (2022). Mapping prospectivity for regolith-hosted REE deposits via  
452 convolutional neural network with generative adversarial network augmented data. *Ore Geology Reviews*, 142.

453 Li, T. F., Xia, Q. L., Zhao, M. Y., Gui, Z., & Leng, S. (2020). Prospectivity Mapping for Tungsten Polymetallic Mineral  
454 Resources, Nanling Metallogenic Belt, South China: Use of Random Forest Algorithm from a Perspective of Data  
455 Imbalance. *Natural Resources Research*, 29(1), 203-227.

456 Liu, L. M., Cao, W., Liu, H. S., Ord, A., Qin, Y. Z., Zhou, F. H., & Bi, C. X. (2022a). Applying benefits and avoiding  
457 pitfalls of 3D computational modeling-based machine learning prediction for exploration targeting: Lessons from  
458 two mines in the Tongling-Anqing district, eastern China. *Ore Geology Reviews*, 142, 104712.

459 Liu, X. X., Zhang, J., Huang, F., Cheng, J. W., Lu, K. X., Yang, J. F., Wang, M., Wang, Y. X., Qiu, J. W., & Zhang, X.  
460 Y. (2022b). Tungsten deposits in southern Jiangxi Province: Constraints on the origin of wolframite from in-situ  
461 U-Pb isotope dating. *Ore Geology Reviews*, 143.

462 Liu, Y., Carranza, E. J. M., & Xia, Q. (2022). Developments in quantitative assessment and modeling of mineral  
463 resource potential: an overview. *Natural Resources Research*, 31(4), 1825-1840.

464 Liu, Y., Cheng, Q. M., Xia, Q. L., & Wang, X. Q. (2013). Application of singularity analysis for mineral potential  
465 identification using geochemical data - A case study: Nanling W-Sn-Mo polymetallic metallogenic belt, South  
466 China. *Journal of Geochemical Exploration*, 134, 61-72.

467 Liu, Y., Cheng, Q. M., Xia, Q. L., & Wang, X. Q. (2014). Mineral potential mapping for tungsten polymetallic deposits  
468 in the Nanling metallogenic belt, South China. *Journal of Earth Science*, 25(4), 689-700.

469 Liu, Y., Cheng, Q. M., Xia, Q. L., & Wang, X. Q. (2015). The use of evidential belief functions for mineral potential  
470 mapping in the Nanling belt, South China. *Frontiers of Earth Science*, 9(2), 342-354.

471 Liu, Y., Zhou, K., & Xia, Q. (2018). A MaxEnt model for mineral prospectivity mapping. *Natural Resources Research*,  
472 27(3), 299-313.

473 Liu, Y., Xia, Q., & Carranza, E. J. M. (2019). Integrating sequential indicator simulation and singularity analysis to  
474 analyze uncertainty of geochemical anomaly for exploration targeting of tungsten polymetallic mineralization,  
475 Nanling belt, South China. *Journal of Geochemical Exploration*, 197, 143-158.

476 Maepa, F., Smith, R. S., & Tessema, A. (2021). Support vector machine and artificial neural network modelling of  
477 orogenic gold prospectivity mapping in the Swayze greenstone belt, Ontario, Canada. *Ore Geology Reviews*, 130.

478 Mao, J. W., Cheng, Y. B., Chen, M. H., & Pirajno, F. (2013a). Major types and time-space distribution of Mesozoic  
479 ore deposits in South China and their geodynamic settings. *Mineralium Deposita*, 48(3), 267-294.

480 Mao, J. W., Pirajno, F., & Cook, N. (2011). Mesozoic metallogeny in East China and corresponding geodynamic  
481 settings - An introduction to the special issue. *Ore Geology Reviews*, 43(1), 1-7.

482 Mao, Z. H., Cheng, Y. B., Liu, J. J., Yuan, S. D., Wu, S. H., Xiang, X. K., & Luo, X. H. (2013b). Geology and  
483 molybdenite Re-Os age of the Dahutang granite-related veinlets-disseminated tungsten ore field in the Jiangxin  
484 Province, China. *Ore Geology Reviews*, 53, 422-433.

485 Meng, E. H., Huang, S. Z., Huang, Q., Fang, W., Wu, L. Z., & Wang, L. (2019). A robust method for non-stationary  
486 streamflow prediction based on improved EMD-SVM model. *Journal of Hydrology*, 568, 462-478.

487 Nezhad, S. G., Mokhtari, A. R., & Rodsari, P. R. (2017). The true sample catchment basin approach in the analysis of  
488 stream sediment geochemical data. *Ore Geology Reviews*, 83, 127-134.

489 Ni, P., Wang, X. D., Wang, G. G., Huang, J. B., Pan, J. Y., & Wang, T. G. (2015). An infrared microthermometric study  
490 of fluid inclusions in coexisting quartz and wolframite from Late Mesozoic tungsten deposits in the Gannan  
491 metallogenic belt, South China. *Ore Geology Reviews*, 65, 1062-1077.

492 Ou, S. M., Lee, K. H., Tsai, M. T., Tseng, W. C., Chu, Y. C., & Tarn, D. C. (2022). Artificial Intelligence for Risk  
493 Prediction of Rehospitalization with Acute Kidney Injury in Sepsis Survivors. *Journal of Personalized Medicine*,  
494 12(1).

495 Parsa, M., & Maghsoudi, A. (2021). Assessing the effects of mineral systems-derived exploration targeting criteria for  
496 random Forests-based predictive mapping of mineral prospectivity in Ahar-Arasbaran area, Iran. *Ore Geology  
497 Reviews*, 138.

498 Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. (2015). Machine learning  
499 predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and  
500 support vector machines. *Ore Geology Reviews*, 71, 804-818.

501 Roshanravan, B., Aghajani, H., Yousefi, M., & Kreuzer, O. (2018). An Improved Prediction-Area Plot for  
502 Prospectivity Analysis of Mineral Deposits. *Natural Resources Research*, 28(3), 1089-1105.

503 Shu, X. J., Wang, X. L., Sun, T., Xu, X. S., & Dai, M. N. (2011). Trace elements, U-Pb ages and Hf isotopes of zircons

504 from Mesozoic granites in the western Nanling Range, South China: Implications for petrogenesis and W-Sn  
505 mineralization. *Lithos*, 127(3-4), 468-482.

506 Sinaice, B. B., Owada, N., Saadat, M., Toriya, H., Inagaki, F., Bagai, Z., & Kawamura, Y. (2021). Coupling NCA  
507 Dimensionality Reduction with Machine Learning in Multispectral Rock Classification Problems. *Minerals*, 11(8).

508 Sun, G. T., Zeng, Q. D., & Zhou, J. X. (2022). Machine learning coupled with mineral geochemistry reveals the origin  
509 of ore deposits. *Ore Geology Reviews*, 142.

510 Sun, T., Chen, F., Zhong, L. X., Liu, W. M., & Wang, Y. (2019). GIS-based mineral prospectivity mapping using  
511 machine learning methods: A case study from Tongling ore district, eastern China. *Ore Geology Reviews*, 109, 26-  
512 49.

513 Sun, T., Li, H., Wu, K. X., Chen, F., Zhu, Z., & Hu, Z. J. (2020a). Data-Driven Predictive Modelling of Mineral  
514 Prospectivity Using Machine Learning and Deep Learning Methods: A Case Study from Southern Jiangxi Province,  
515 China. *Minerals*, 10(2).

516 Sun, X. L., Liu, M. X., & Sima, Z. Q. (2020b). A novel cryptocurrency price trend forecasting model based on  
517 LightGBM. *Finance Research Letters*, 32.

518 Vaulet, T., Al-Memar, M., Fourie, H., Bobdiwala, S., Saso, S., Pipi, M., Stalder, C., Bennett, P., Timmerman, D.,  
519 Bourne, T., & De Moor, B. (2022). Gradient boosted trees with individual explanations: An alternative to logistic  
520 regression for viability prediction in the first trimester of pregnancy. *Computer Methods and Programs in*  
521 *Biomedicine*, 213.

522 Wang, K. J., Zheng, X. Q., Wang, G. W., Liu, D. Y., & Cui, N. (2020a). A Multi-model ensemble approach for gold  
523 mineral prospectivity mapping: A Case Study on the Beishan Region, Western China. *Minerals*, 10(12).

524 Wang, C. B., Pan, Y. P., Chen, J. G., Ouyang, Y. P., Rao, J. F., & Jiang, Q. B. (2020b). Indicator element selection and  
525 geochemical anomaly mapping using recursive feature elimination and random forest methods in the Jingdezhen  
526 region of Jiangxi Province, South China. *Applied Geochemistry*, 122.

527 Wang, D. H., Huang, F., Wang, Y., He, H. H., Li, X. M., Liu, X. X., Sheng, J. F., & Liang, T. (2020c). Regional  
528 metallogeny of Tungsten-tin-polymetallic deposits in Nanling region, South China. *Ore Geology Reviews*, 120.

529 Wang, X., Ren, M. H., & Chen, J. (2017). The muscovite granites: Parental rocks to the Nanling Range tungsten  
530 mineralization in South China. *Ore Geology Reviews*, 88, 702-717.

531 Wu, J. H., Kong, H., Li, H., Algeo, T. J., Yonezu, K., Liu, B. A., Wu, Q. H., Zhu, D. P., & Jiang, H. (2021). Multiple

532 metal sources of coupled Cu-Sn deposits: Insights from the Tongshanling polymetallic deposit in the Nanling Range,  
533 South China. *Ore Geology Reviews*, 139.

534 Xiang, J., Xiao, K. Y., Carranza, E. J. M., Chen, J. P., & Li, S. (2020). 3D Mineral Prospectivity Mapping with Random  
535 Forests: A Case Study of Tongling, Anhui, China. *Natural Resources Research*, 29(1), 395-414.

536 Xie, G., Mao, J., Li, W., Fu, B., & Zhang, Z. (2018). Granite-related Yangjiashan tungsten deposit, southern China.  
537 *Mineralium Deposita*, 54(1), 67-80.

538 Xiong, Y., & Zuo, R. (2018). GIS-based rare events logistic regression for mineral prospectivity mapping. *Computers  
539 & Geosciences*, 111, 18-25.

540 Xiong, Y. H., Zuo, R. G., Luo, Z. J., & Wang, X. Q. (2021). A Physically Constrained Variational Autoencoder for  
541 Geochemical Pattern Recognition. *Mathematical Geosciences*.

542 Xiong, Y. Q., Shao, Y. J., Cheng, Y. B., & Jiang, S. Y. (2020). Discrete Jurassic and Cretaceous Mineralization Events  
543 at the Xiangdong W(-Sn) Deposit, Nanling Range, South China. *Economic Geology*, 115(2), 385-413.

544 Xu, X. S., O'Reilly, S. Y., Griffin, W. L., Deng, P., & Pearson, N. J. (2005). Relict proterozoic basement in the Nanling  
545 Mountains (SE China) and its tectonothermal overprinting. *Tectonics*, 24(2).

546 Yousefi, M., & Carranza, E. J. M. (2015). Prediction-area (P-A) plot and C-A fractal analysis to classify and evaluate  
547 evidential maps for mineral prospectivity modeling. *Computers & Geosciences*, 79, 69-81.

548 Yousefi, M., Carranza, E. J. M., & Kamkar-Rouhani, A. (2013). Weighted drainage catchment basin mapping of  
549 geochemical anomalies using stream sediment data for mineral potential modeling. *Journal of Geochemical  
550 Exploration*, 128, 88-96.

551 Yousefi, M., Kamkar-Rouhani, A., & Carranza, E. J. M. (2012). Geochemical mineralization probability index (GMPI):  
552 A new approach to generate enhanced stream sediment geochemical evidential map for increasing probability of  
553 success in mineral potential mapping. *Journal of Geochemical Exploration*, 115, 24-35.

554 Youssef, A. M., Pourghasemi, H. R., Pourtaghi, Z. S., & Al-Katheeri, M. M. (2016). Landslide susceptibility mapping  
555 using random forest, boosted regression tree, classification and regression tree, and general linear models and  
556 comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. *Landslides*, 13(5), 839-856.

557 Yu, X. Q., Wu, G. G., Zhao, X., Gao, J. F., Di, Y. J., Zheng, Y., Dai, Y. P., Li, C. L., & Qiu, J. T. (2010). The Early  
558 Jurassic tectono-magmatic events in southern Jiangxi and northern Guangdong provinces, SE China: Constraints  
559 from the SHRIMP zircon U-Pb dating. *Journal of Asian Earth Sciences*, 39(5), 408-422.

560 Yuan, S. D., Williams-Jones, A. E., Romer, R. L., Zhao, P. L., & Mao, J. W. (2019). Protolith-related thermal controls  
561 on the decoupling of Sn and W in Sn-W metallogenic provinces: insights from the Nanling region, China. *Economic*  
562 *Geology*, 114(5), 1005-1012.

563 Zafari, M., Kumar, D., Umer, M., & Kim, K. S. (2020). Machine learning-based high throughput screening for nitrogen  
564 fixation on boron-doped single atom catalysts. *Journal of Materials Chemistry A*, 8(10), 5209-5216.

565 Zhang, J., Liu, X., Li, W., Zeng, Z., Hu, H., Peng, L., Cheng, J., Lu, K., & Li, P. (2021). The metallogenic epoch and  
566 geological implications of the tungsten-tin polymetallic deposits in southern Jiangxi Province, China: Constraints  
567 from cassiterite U-Pb and molybdenite Re-Os isotopic dating. *Ore Geology Reviews*, 134.

568 Zhang, R. Q., Lu, J. J., Lehmann, B., Li, C. Y., Li, G. L., Zhang, L. P., Guo, J., & Sun, W. D. (2017a). Combined  
569 zircon and cassiterite U-Pb dating of the Piaotang granite-related tungsten-tin deposit, southern Jiangxi tungsten  
570 district, China. *Ore Geology Reviews*, 82, 268-284.

571 Zhang, Y., Yang, J. H., Chen, J. Y., Wang, H., & Xiang, Y. X. (2017b). Petrogenesis of Jurassic tungsten-bearing  
572 granites in the Nanling Range, South China: Evidence from whole-rock geochemistry and zircon U-Pb and Hf-O  
573 isotopes. *Lithos*, 278, 166-180.

574 Zhou, T., Wang, F. X., & Yang, Z. (2017). Comparative Analysis of ANN and SVM Models Combined with Wavelet  
575 Preprocess for Groundwater Depth Prediction. *Water*, 9(10).

576

577

578

579 **Tables**

580

581 Table 1. Statistics of cumulative contribution rate of three variable types

	<b>LR</b>	<b>SVM</b>	<b>RF</b>	<b>CNN</b>	<b>LGBM</b>
<b>Geochemical element</b>	86.96%	78.56%	97.14%	90.24%	88.86%
<b>Granite mass</b>	8.07%	12.83%	1.35%	5.09%	5.16%
<b>Fault</b>	4.98%	8.61%	1.51%	4.67%	5.99%

582

583

584

585

586

587

Table 2. Model intersection data of the P-A plots

	<b>LR</b>	<b>SVM</b>	<b>RF</b>	<b>CNN</b>	<b>LGBM</b>
Percentage of study area	22%	27%	22%	19%	25%
Percentage of tungsten polymetallic deposit	78%	73%	78%	81%	75%
Likelihood of deposit occurrence	0.47%	0.36%	0.47%	0.57%	0.40%

588

589

590

591 **Figure captions**

592 Figure 1. Geographical location of the study area

593 Figure 2. Flow chart of the study showing modelling strategy

594 Figure 3. Biplot of F1 vs. F2 elements

595 Figure 4. IDW results of 11 geochemical elements: (a) Be; (b) Pb; (c) F; (d) Li; (e) W; (f) Sn; (g) Bi; (h) Cd; (I) Y; (j)  
596 U; (k) Zn

597 Figure 5. Fault-related elements: (a) buffering analysis of faults; (b) density of fault intersections

598 Figure 6. Schematic diagram of SVM hyperplane division principle

599 Figure 7. Schematic diagram of bagging algorithm

600 Figure 8. Schematic diagram of random forest algorithm

601 Figure 9. Local connection diagram: (a) Local connection diagram; (b) weight sharing schematic diagram; (c)  
602 Principle of convolution operation; (d) Principle of pool operation

603 Figure 10. Convolutional neural network structure diagram

604 Figure 11. Boosting algorithm schematics

605 Figure 12. (a) Level-wise strategy; (b) Leaf-wise strategy with depth constraint

606 Figure 13. Metallogenic prediction of five models: (a) LR, (b) SVM, (c) RF, (d) CNN, (e) LGBM

607 Figure 14. ROC curve of five models

608 Figure 15. K-fold\_AUC and original AUC value of five models

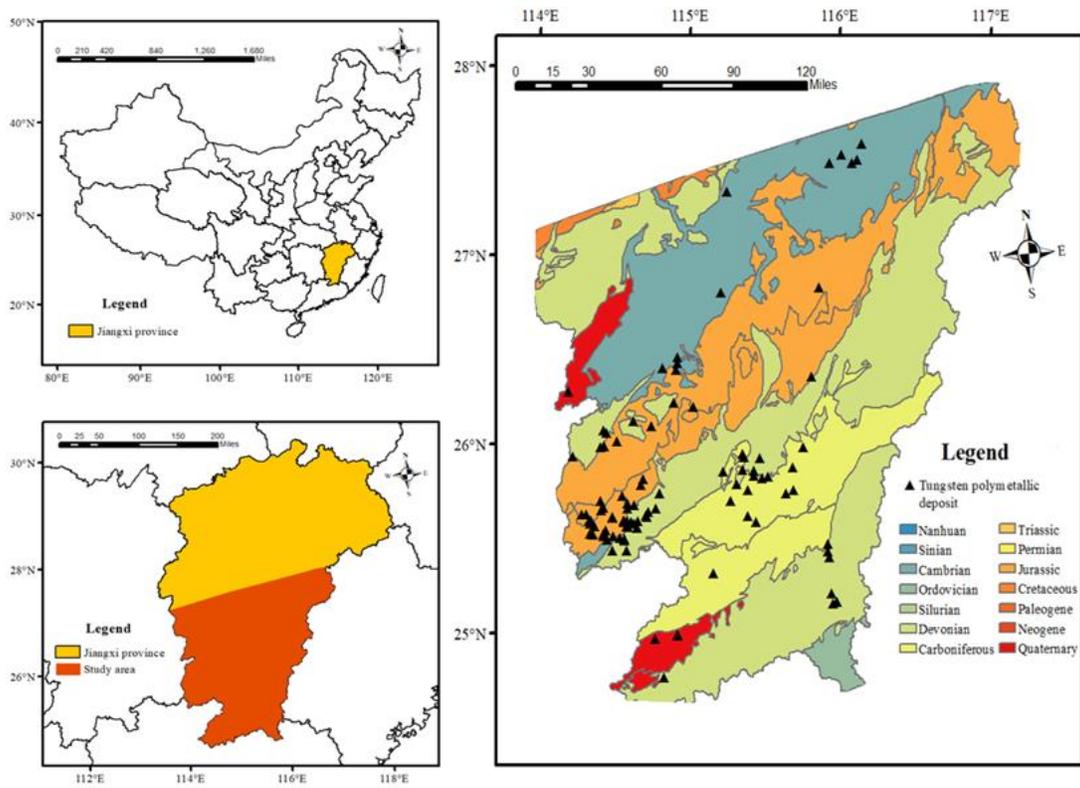
609 Figure 16. Contribution rate of five model variables: (a)LR, (b)SVM, (c)RF, (d)CNN, (e)LGBM

610 Figure 17. P-A Plot of five models: (a) LR, (b) SVM, (c) RF, (d) CNN, (e) LGBM

611 Figure 18. High potential areas for tungsten polymetallic exploration in the study area.

612 **Figures**

613

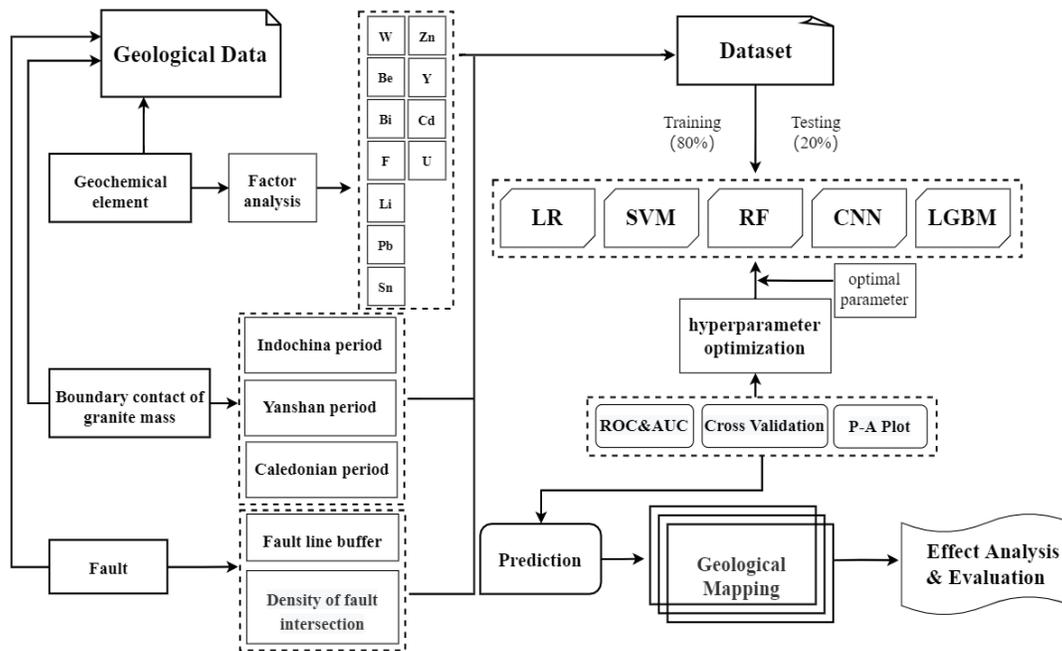


614

615

Figure 1. Geographical location of the study area

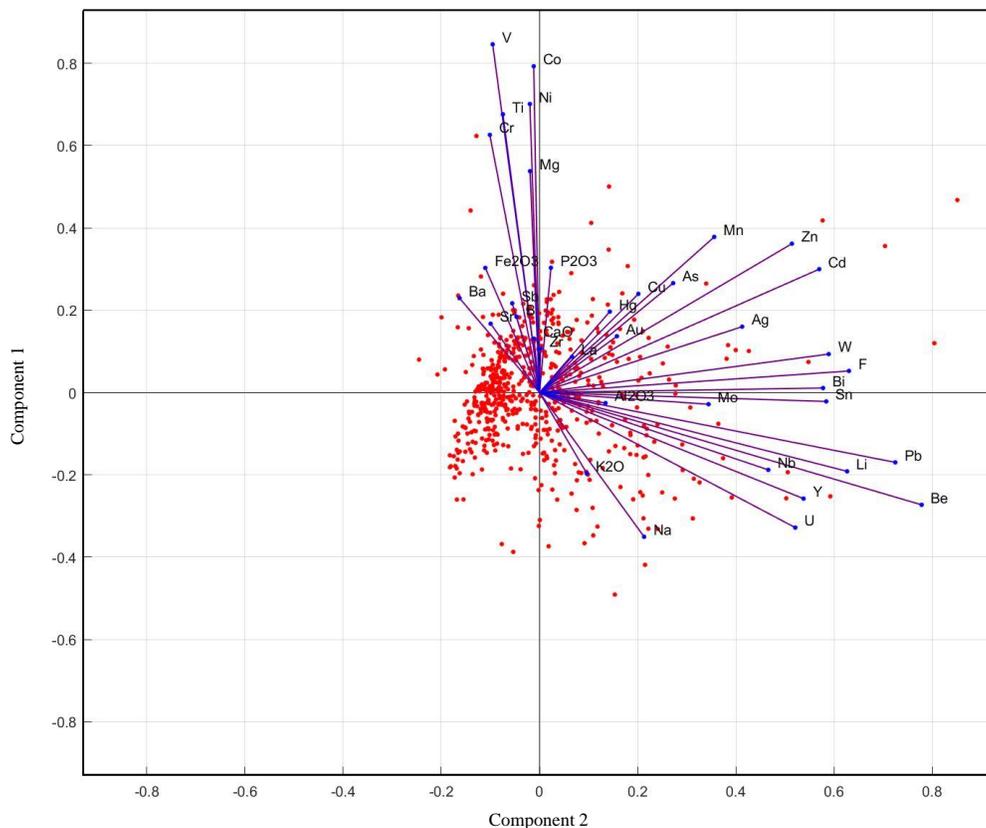
616  
617



618  
619

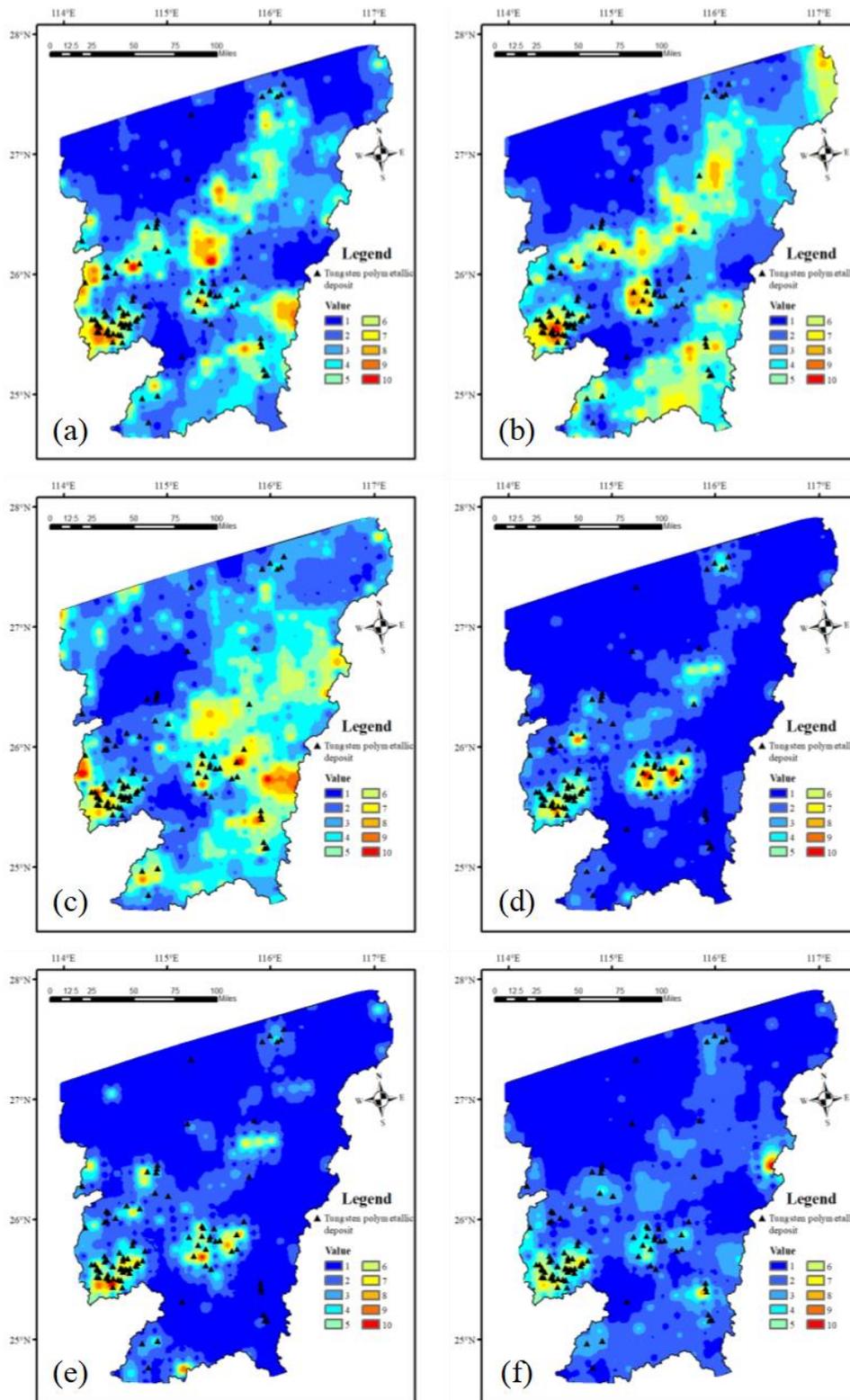
Figure 2. Flow chart of the study showing modelling strategy

620  
621



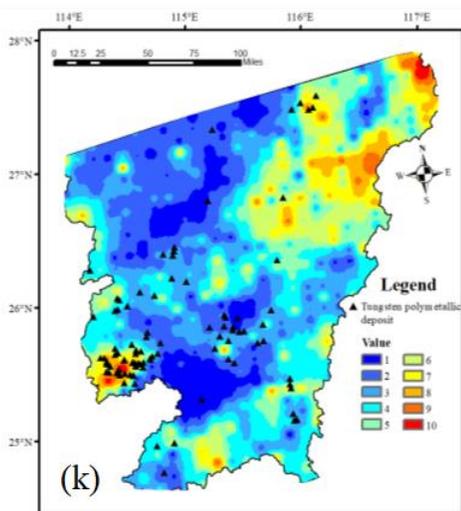
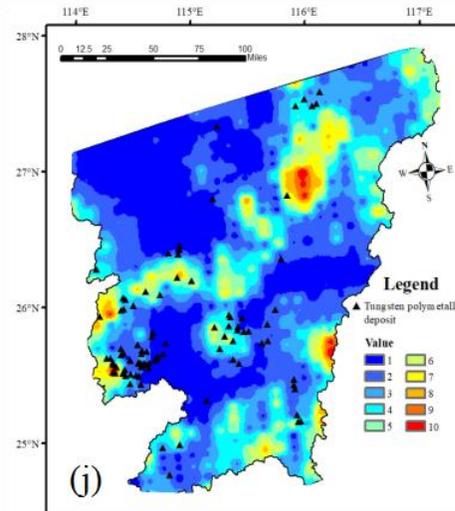
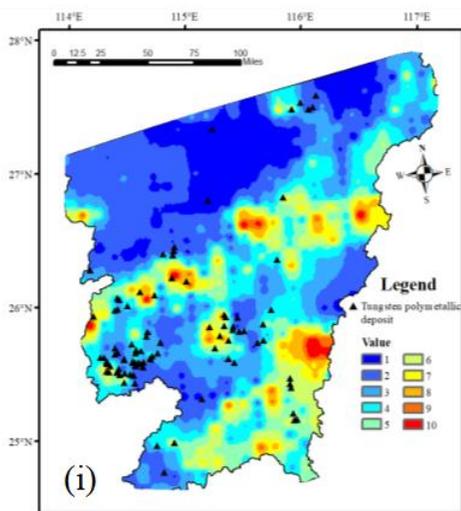
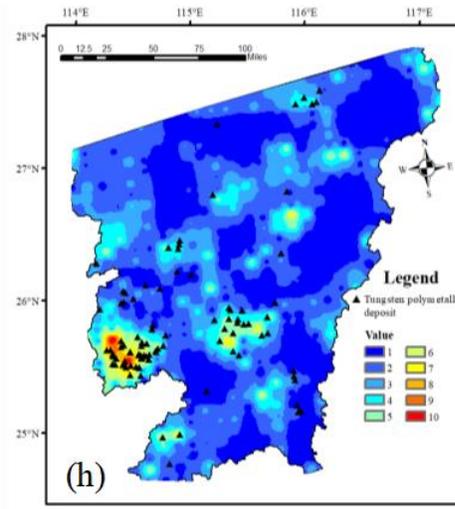
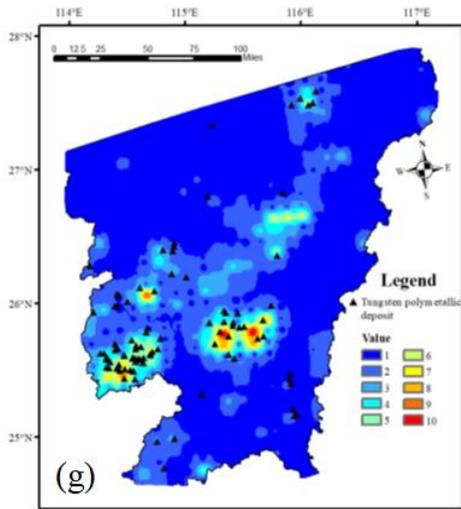
622  
623  
624

Figure 3. Biplot of F1 vs. F2 elements



625

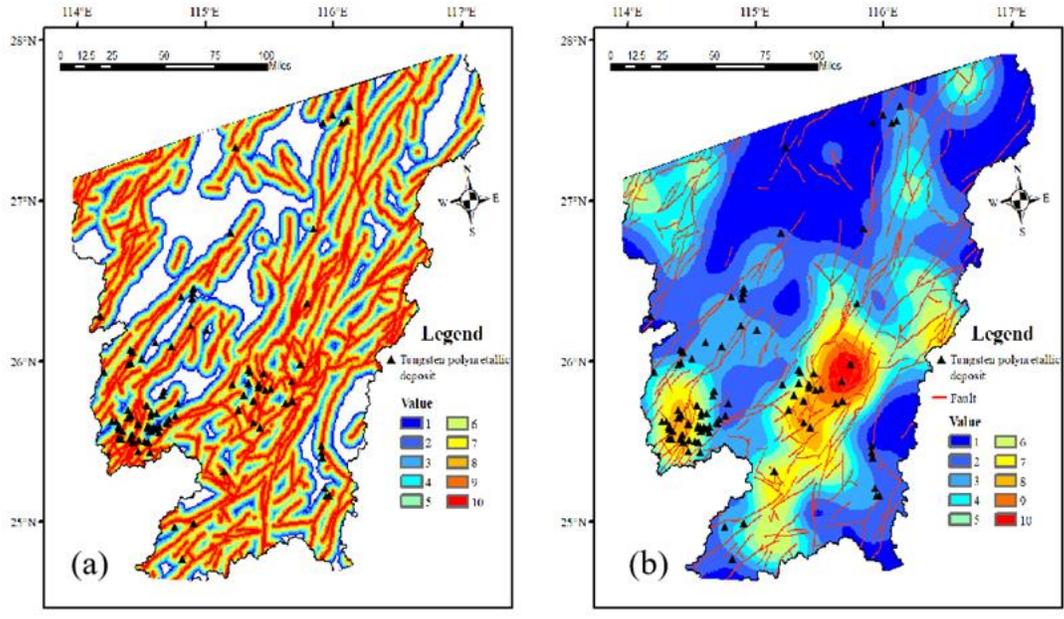
626 Figure 4. IDW results of 11 geochemical elements: (a) Be; (b) Pb; (c) F; (d) Li; (e) W; (f) Sn; (g) Bi; (h) Cd; (i) Y;  
 627 (j) U; (k) Zn  
 628



629

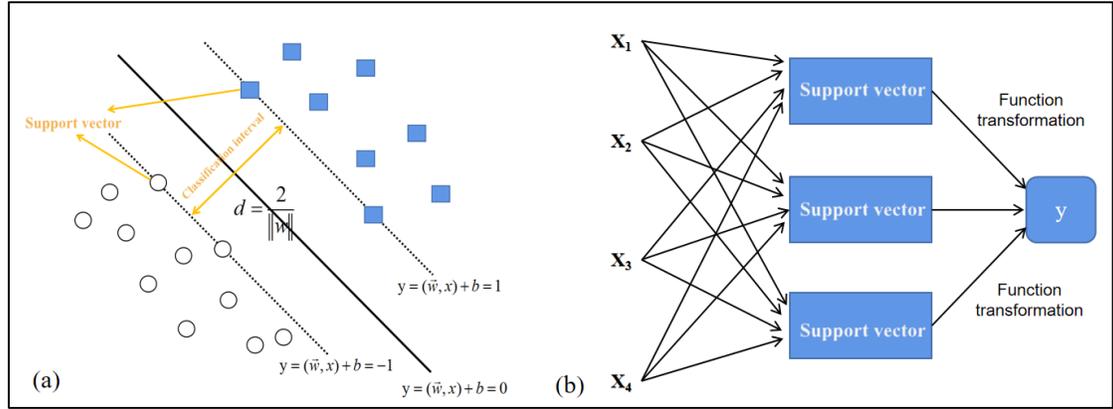
630

Figure 4 continued



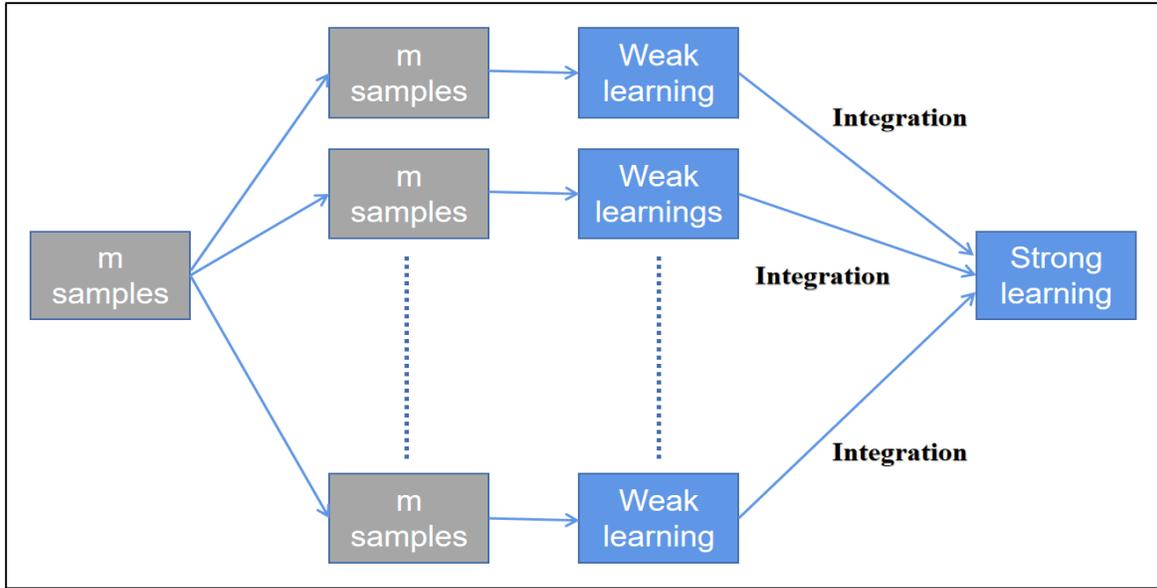
631  
632  
633  
634  
635

Figure 5. Fault-related elements: (a) buffering analysis of faults; (b) density of fault intersections



636  
637  
638

Figure 6. Schematic diagram of SVM hyperplane division principle

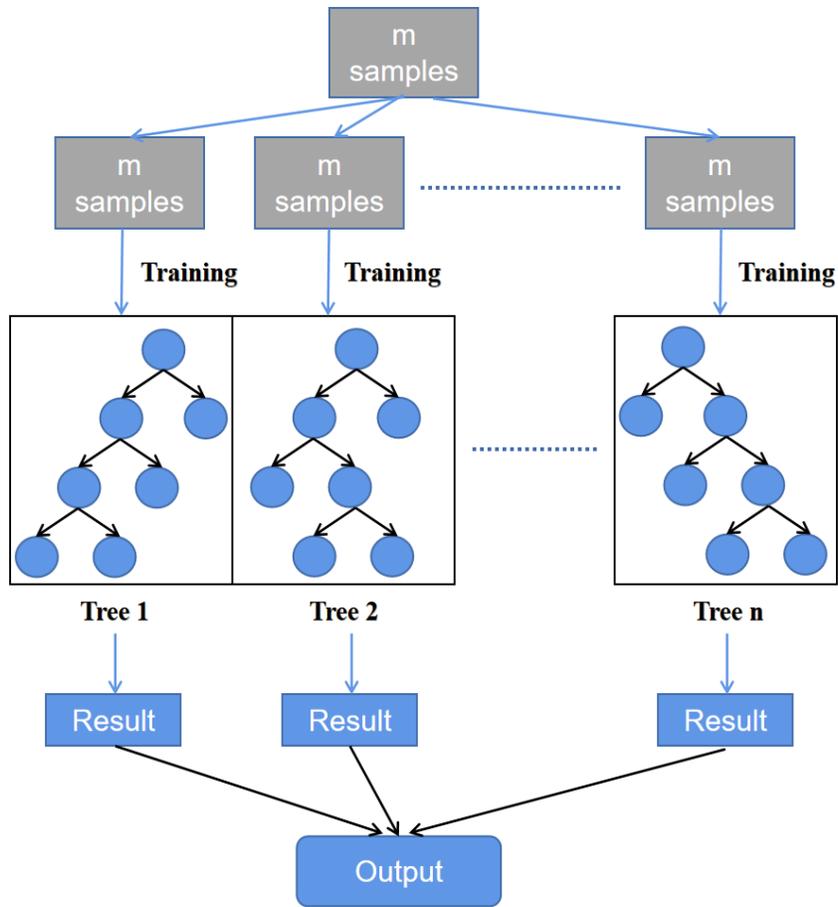


639

640

641

Figure 7. Schematic diagram of bagging algorithm

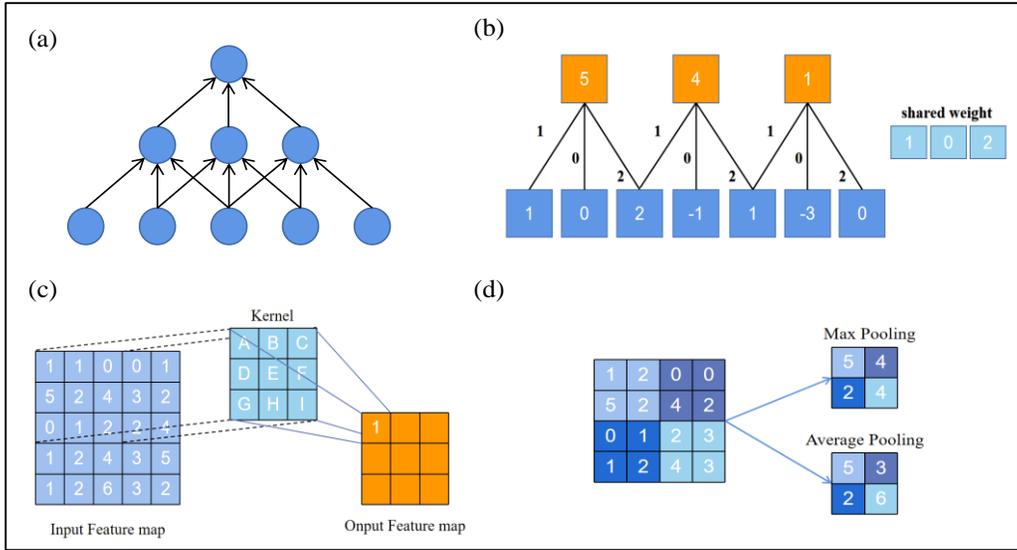


642

643

644

Figure 8. Schematic diagram of random forest algorithm



645

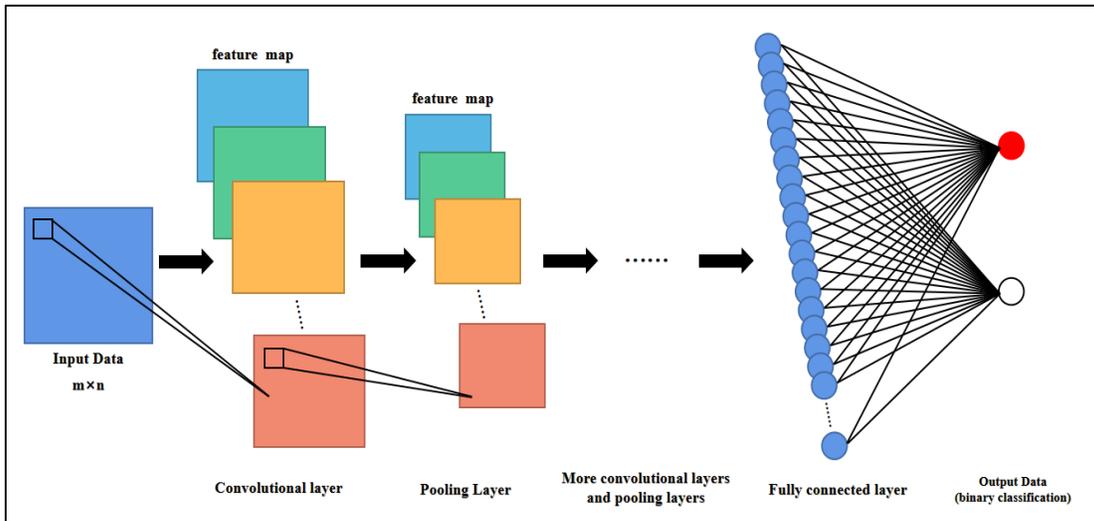
646

647

648

649

Figure 9. Local connection diagram: (a) Local connection diagram; (b) weight sharing schematic diagram; (c) Principle of convolution operation; (d) Principle of pool operation

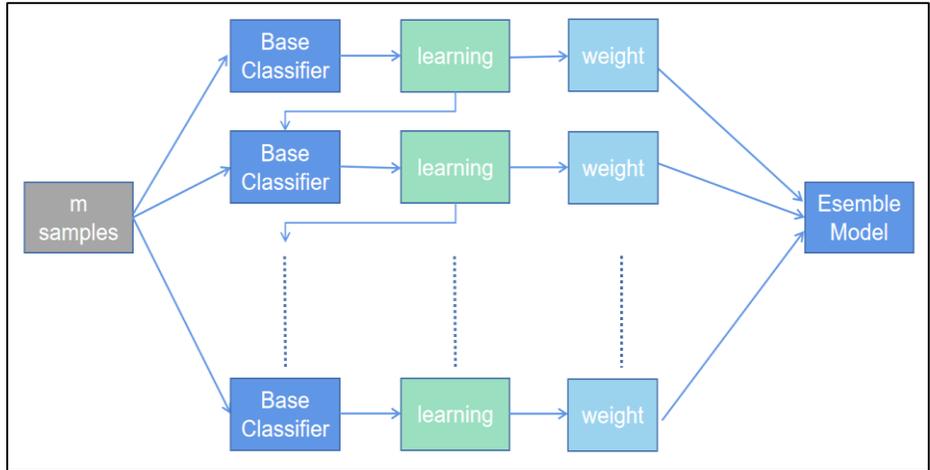


650

651

652

Figure 10. Convolutional neural network structure diagram

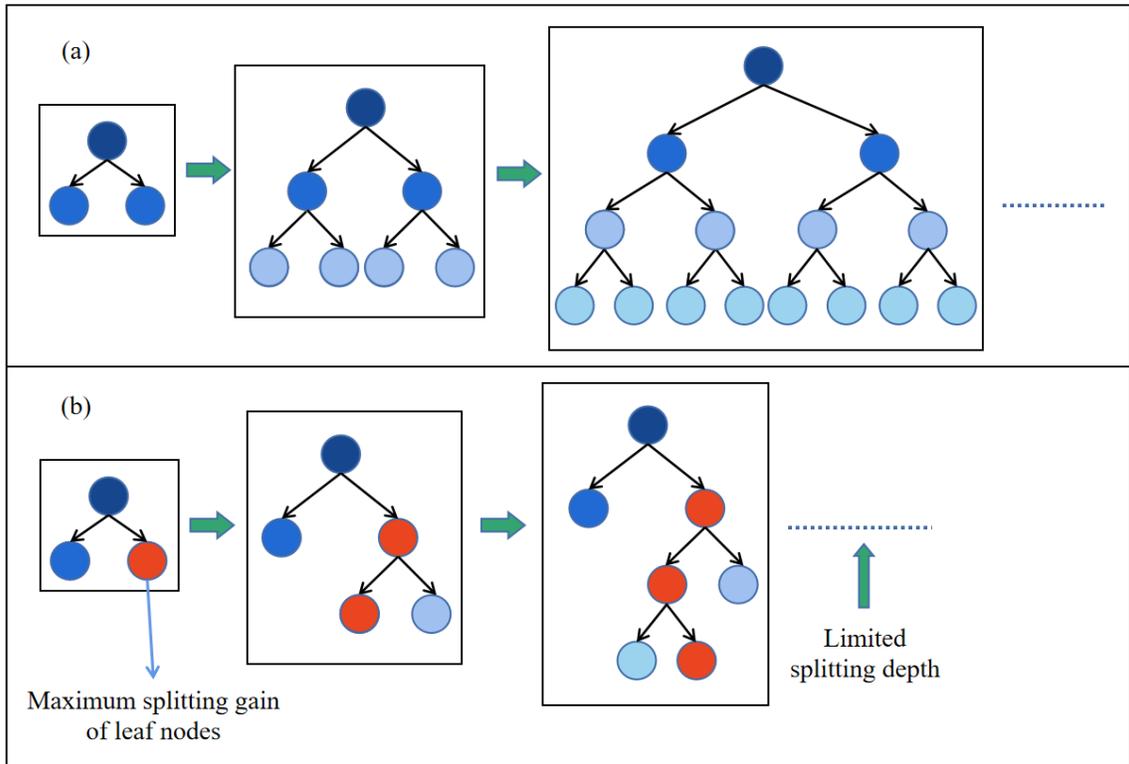


653

654

655

Figure 11. Boosting algorithm schematics

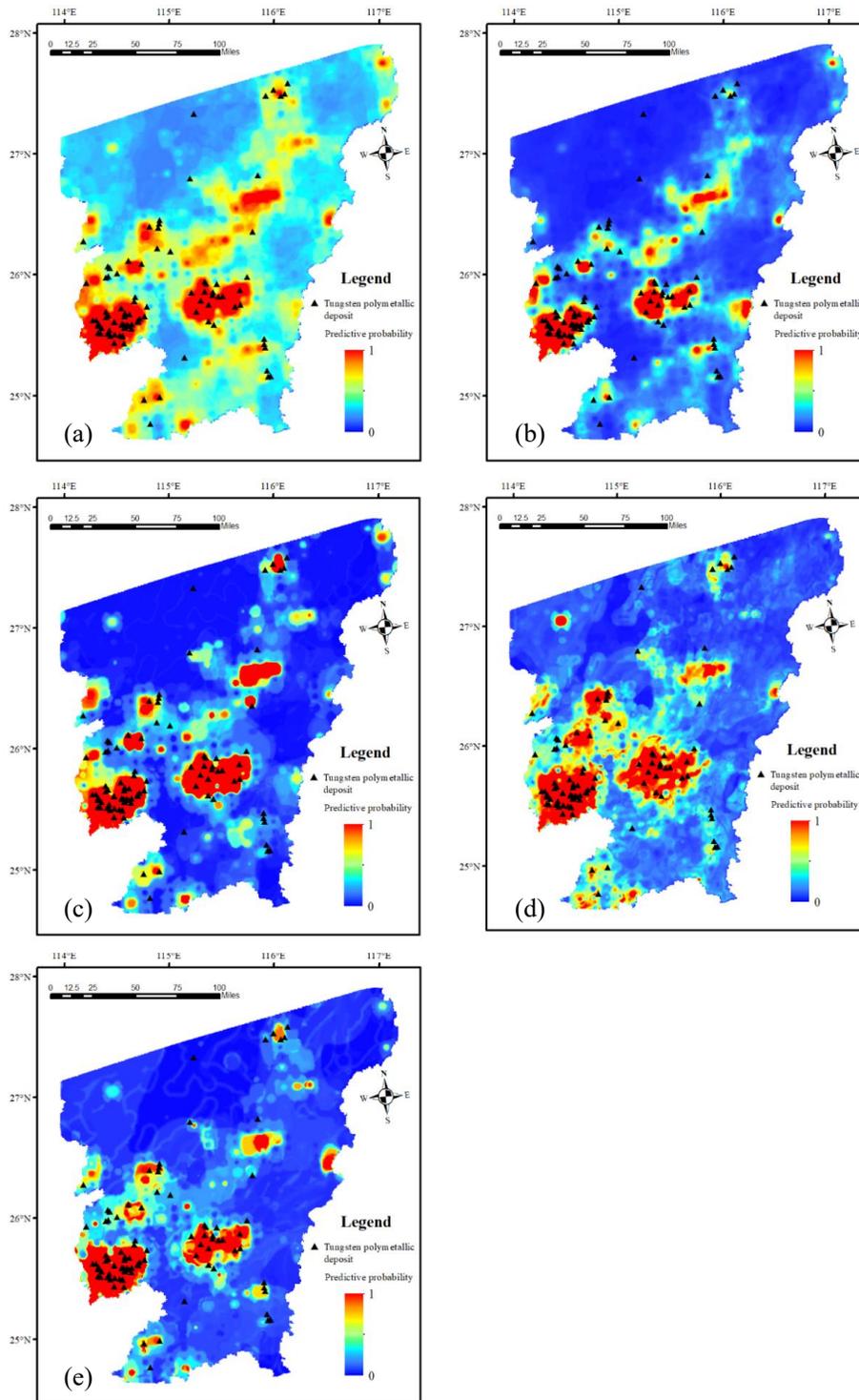


656

657

658

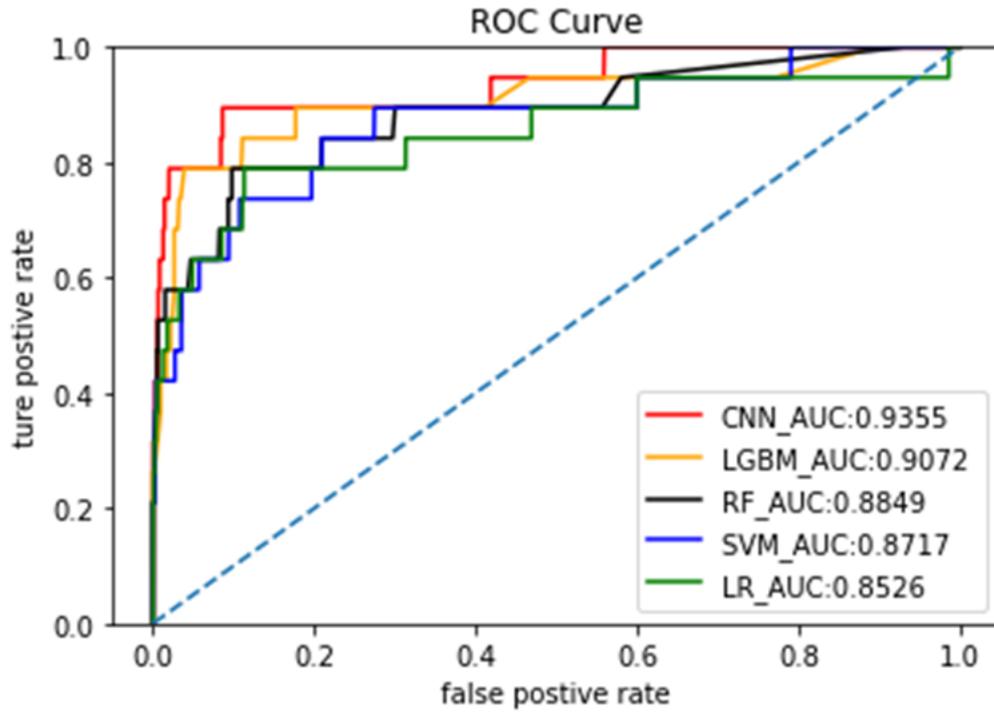
Figure 12. (a) Level-wise strategy; (b) Leaf-wise strategy with depth constraint



659

660

Figure 13. Metallogenetic prediction of five models: (a) LR, (b) SVM, (c) RF, (d) CNN, (e) LGBM

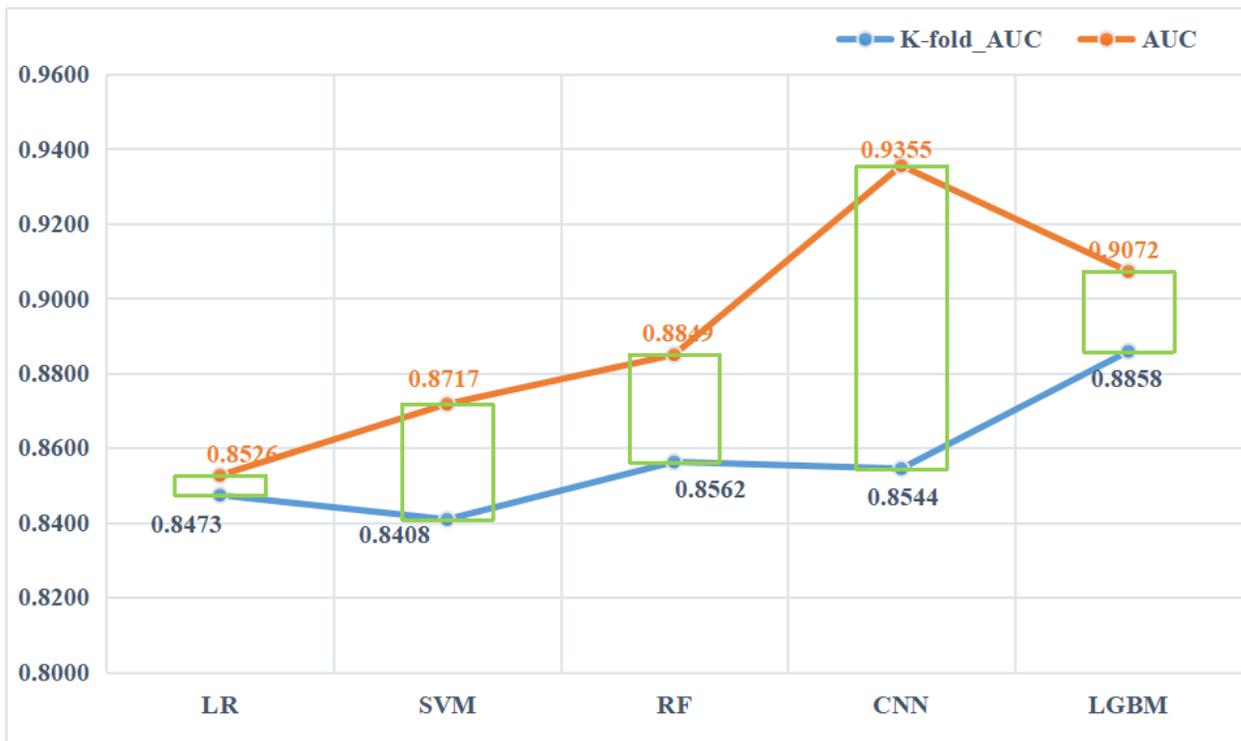


661

662

663

Figure 14. ROC curve of five models



664

665

666

Figure 15. K-fold\_AUC and original AUC value of five models

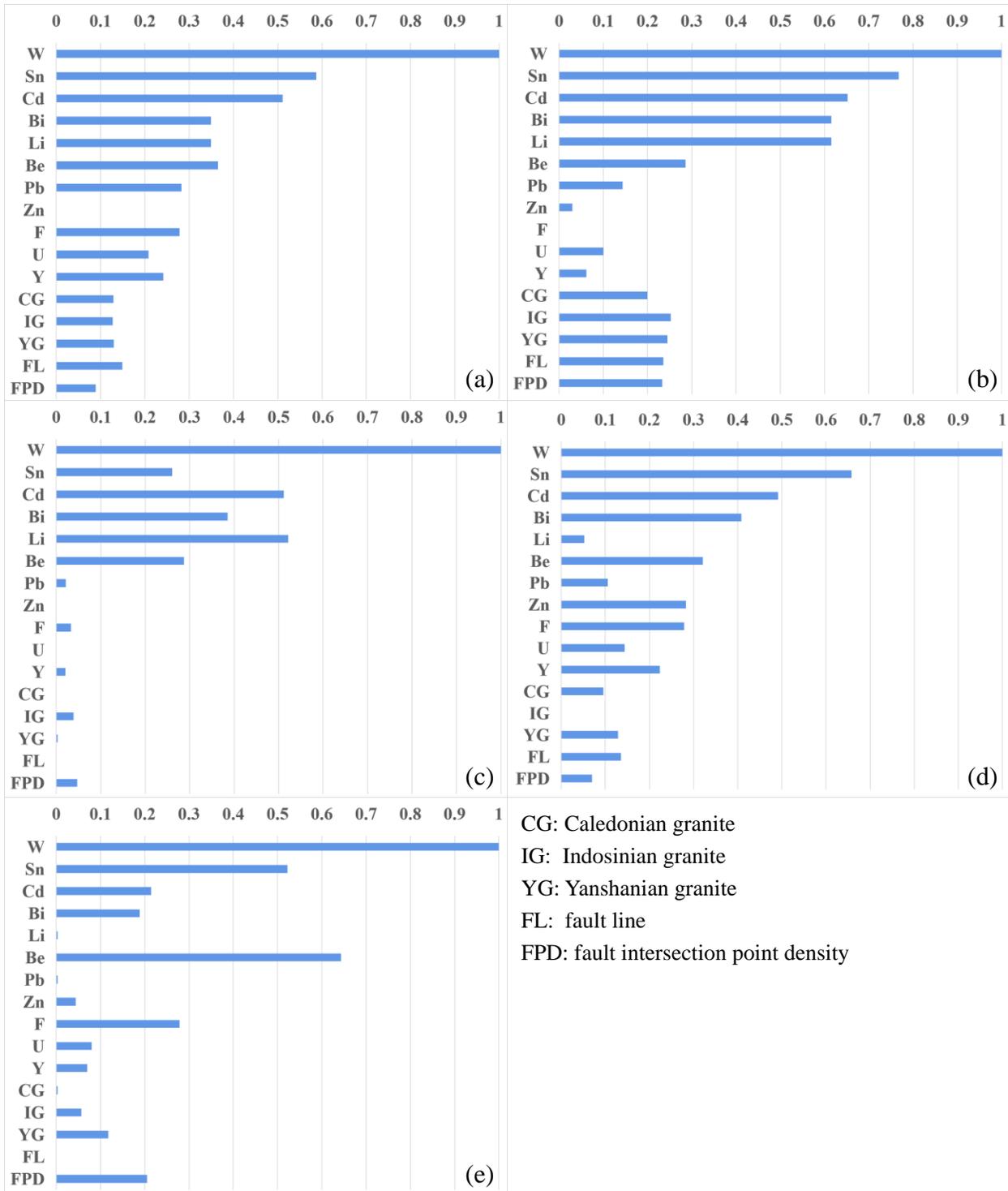
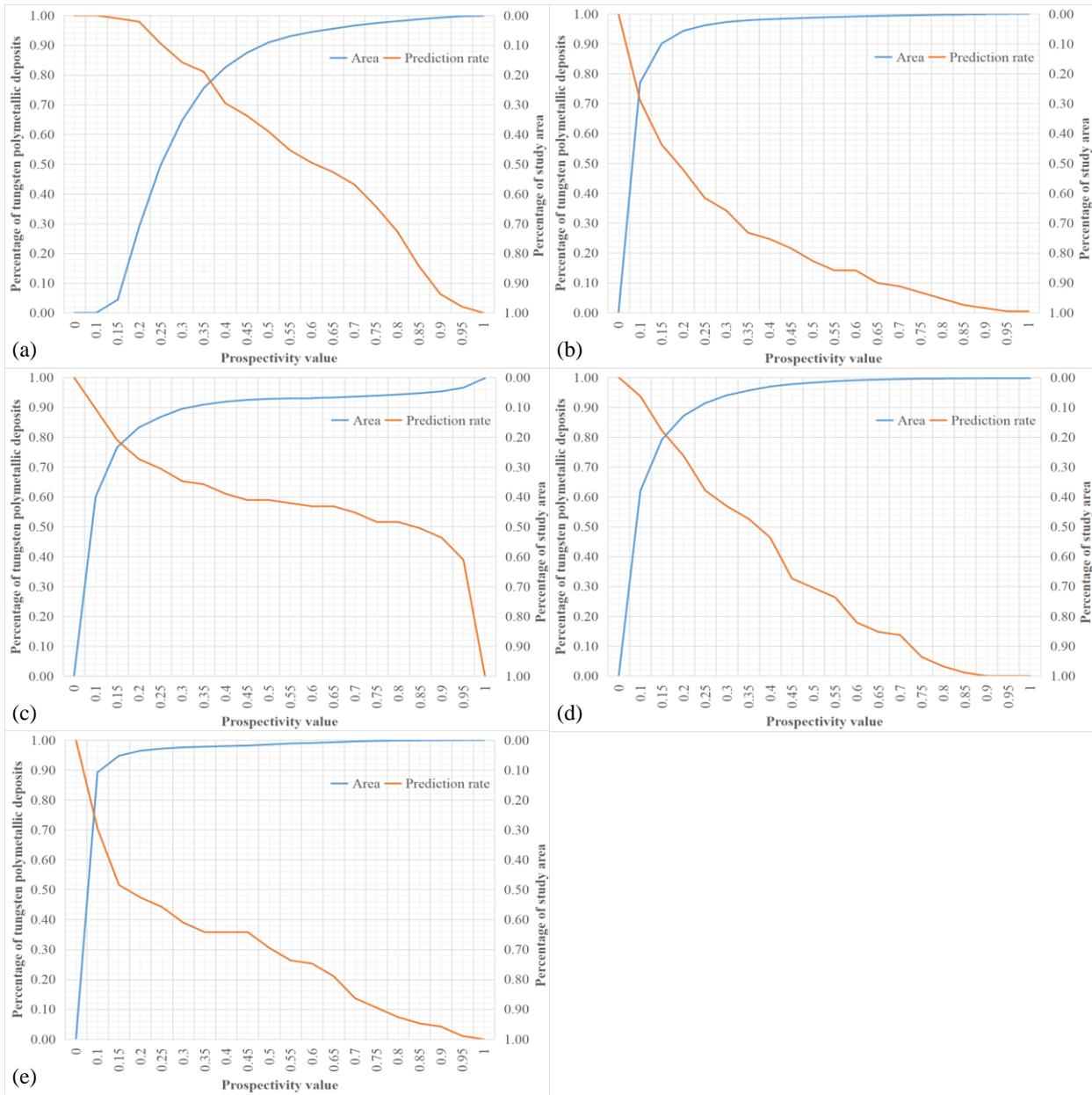


Figure 16. Contribution rate of five model variables: (a)LR, (b)SVM, (c)RF, (d)CNN, (e)LGBM

667

668

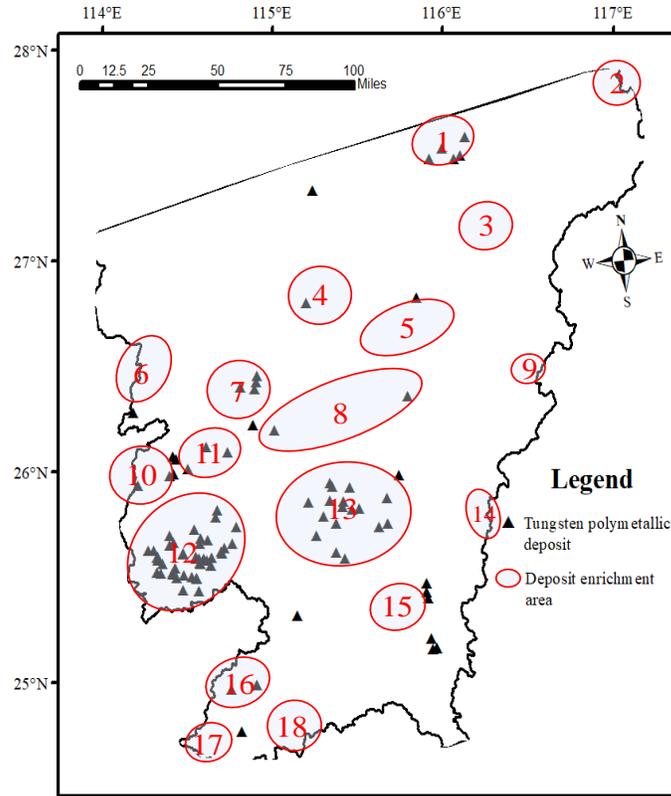
669



670

671

Figure 17. P-A Plot of five models: (a) LR, (b) SVM, (c) RF, (d) CNN, (e) LGBM



672

673

674

675

676 **End**

Figure 18. High potential areas for tungsten polymetallic exploration in the study area