remote sensing

MDPI

Article

Mineral Prospectivity Prediction in the Mayoumu Area, Tibet, Based on Multi-Source Exploration Information and Ensemble Learning Models

Kai Qiao, Tao Luo, Shihao Ding, Cong Han, Shisong Gong, Zhiwen Ren and Yong Huang *

Research Center of Applied Geology of China Geological Survey, Chengdu 610036, China; qk6173@126.com (K.Q.); 18537776173@163.com (T.L.); dingsh1998@126.com (S.D.); hac0802@163.com (C.H.); gongshisong2025@163.com (S.G.); zhiwen1994@126.com (Z.R.)
* Correspondence: huangy@mail.cgs.gov.cn

Highlights

What are the main findings?

- A unified multi-source, multi-scale predictor system is established for complex orogenic belts by synergistically integrating GF-5 hyperspectral MTMF sub-pixel alteration mapping with Sentinel-2 iCrosta hydroxyl/iron-oxide anomalies, together with ILR–RPCA-derived geochemical anomaly factors and Euclidean-distance-to-fault structural constraints.
- An integrated ensemble-learning prospectivity framework (GBDT, XGBoost, LightGBM) is developed and coupled with ROC/AUC evaluation and SHAP attribution to jointly quantify prediction performance and interpret key controls, demonstrating effectiveness in the study area and strong transferability to gold prospectivity mapping in analogous orogenic belts worldwide.

What is the implications of the main findings?

- The proposed workflow is transferable to plateau–orogenic belts worldwide and can support target prioritization and efficient follow-up exploration in complex, heterogeneous terrains.

Abstract

Plateau–orogenic belts host a substantial share of global gold resources, yet quantitative prospectivity mapping is challenged by complex mineralization and strongly heterogeneous, multi-scale datasets. Using the Mayoumu area (Tibet) as a representative orogenic gold district, we develop an integrated multi-source workflow that fuses remote-sensing alteration information with regional geochemical and structural constraints within an ensemble-learning framework. Alteration anomalies were mapped from GF-5 hyperspectral imagery using mixture-tuned matched filtering (MTMF) and from Sentinel-2 multispectral imagery using the iCrosta method to extend alteration signals across scales. Geochemical anomalies were extracted from 1:200,000 stream-sediment data through isometric log-ratio (ILR) transformation and robust principal component analysis (RPCA). At the same time, ore-controlling structures were quantified using Euclidean-distance-to-fault layers. Three Boosting-based ensemble models—gradient boosting decision tree (GBDT), extreme gradient boosting (XGBoost), and light gradient boosting machine (LightGBM)—were trained to predict mineral prospectivity. Performance was evaluated using confusion matrix metrics and ROC–AUC, and key predictors were interpreted using SHAP. All three models achieved AUC values > 0.90, with LightGBM performing

Check for updates

Academic Editor: Kevin Tansey

Received: 15 January 2026
Revised: 20 February 2026
Accepted: 25 February 2026
Published: 26 February 2026

Copyright: © 2026 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.

best (AUC = 0.94) and delineating high-prospectivity zones that coincide with known occurrences and highlight additional targets. The proposed workflow provides a practical, transferable reference for gold prospectivity mapping in complex orogenic belts worldwide.

Keywords: hyperspectral; multi-source; ensemble learning; mineral prospectivity prediction; Tibet

1. Introduction

As a key region in China's new round of strategic actions for major mineral exploration breakthroughs, Tibet possesses highly favorable metallogenic conditions resulting from its unique geological evolution and tectonic framework [1–6]. Although the harsh natural environment of Tibet imposes substantial technical constraints on conventional surface-based exploration methods, surface characteristics such as sparse vegetation cover provide distinct advantages for accurately identifying and extracting alteration anomalies using remote sensing techniques. Numerous studies [7–12] have applied multispectral and hyperspectral remote sensing data to investigate alteration information related to metallic mineral deposits in Tibet, yielding important results that have effectively supported the identification of indicative mineralization signatures and the expansion of prospective exploration targets. Since its discovery in 2001, extensive studies [13–18] have been conducted by both domestic and international researchers, focusing on the geological characteristics of this region. However, existing research has largely concentrated on the structural setting, ore genesis, geochronology, lithogeochemical characteristics, and evolution of the ore-forming fluids in the Mayoumu gold deposit, whereas systematic investigations addressing future exploration directions in the broader Mayoumu area remain limited. One of the primary reasons for this limitation is the insufficient development of regional-scale integrated mineral prospectivity prediction studies. The available multi-source exploration datasets have not yet been fully exploited or effectively integrated, particularly with respect to their combination with big data analytics and advanced information technologies. Consequently, there is a pressing need to introduce new techniques and methodologies that enable intelligent integration and analysis of multidisciplinary exploration data, including remote sensing, geological, and geochemical datasets. Such an approach is crucial for predicting deeply concealed mineralization and for providing a robust scientific foundation to inform future mineral exploration strategies.

Mineral prospectivity prediction (mineral potential mapping) is an incremental exploration process. Its core objective is to infer the likelihood of mineralization in prospective zones by analyzing ore-forming conditions and exploration evidence derived from geological investigations, thereby providing guidance for subsequent exploration activities. In recent years, machine learning algorithms have been widely adopted in mineral resource prediction because they can effectively capture complex, nonlinear, and high-dimensional spatial patterns. Previous studies have demonstrated that data-driven prospectivity models can yield robust predictive performance. Representative data-driven approaches include classical machine learning methods such as logistic regression [19], random forest (RF) [20,21], and support vector machines (SVM) [22,23], as well as deep learning methods represented by convolutional neural networks (CNNs) [24] and generative adversarial networks (GANs) [25]. Moreover, driven by the increasing demand for spatially explicit modeling of multi-source geoscience data and cross-scale data fusion, graph neural networks (GNNs) and attention-based variants [26,27], together with Transformer-based methods [28,29], have also been introduced into mineral prospectivity research over the past

two years and have shown encouraging progress. These methods can handle large-scale, high-dimensional, and non-Gaussian datasets with relatively high computational efficiency, enable more accurate estimation of key parameters, and reduce predictive uncertainty. Nevertheless, model performance can vary substantially across datasets; in particular, in structurally complex and highly heterogeneous regions, the predictive capability of a single classifier is often constrained. At the same time, data-driven approaches commonly suffer from a "black-box" effect, limiting the interpretability of prediction results [30,31]. Without reliable explanation tools, it is difficult to clarify feature importance and quantify each predictor's contribution to model outputs, which in turn limits model optimization and the incorporation of new predictors. To address this issue, several studies [32] have advocated integrating prior geological knowledge with data-driven methods to improve both predictive accuracy and interpretability, or seeking better solutions by comparing knowledge-driven and data-driven approaches [33].

Against this background, ensemble learning [34] has become an increasingly important approach in mineral prospectivity prediction. Ensemble learning enhances predictive accuracy, generalization capability, and robustness by combining multiple weak learners into a single, stronger learner through specific strategies. Among various ensemble strategies, Bagging and Boosting are the two most representative. Bagging generates multiple training subsets via bootstrap resampling and trains independent models, with final predictions obtained through a voting scheme [35], as exemplified by the random forest algorithm [36]. In contrast, Boosting assigns higher weights to misclassified samples during iterative training, thereby progressively enhancing overall model performance. Representative Boosting-based methods include gradient boosting decision trees (GBDT), extreme gradient boosting (XGBoost), and light gradient boosting machine (LightGBM). In recent years, these boosting-based algorithms [37–39] have demonstrated significant advantages in mineral prospectivity studies, exhibiting high predictive accuracy and robustness. Compared with other machine learning approaches, boosting algorithms perform well when handling imbalanced datasets, generally require less complex model optimization and parameter tuning, show low sensitivity to non-informative features, and possess strong resistance to overfitting. More importantly, with the rapid development of explainable artificial intelligence, the integration of ensemble learning methods with shaley additive explanations (SHAP) provides an effective solution to the "black-box" problem. By incorporating SHAP into ensemble models, the contribution of each feature to prediction results can be quantitatively assessed, and the internal decision logic of Stacking or Boosting structures can be systematically interpreted [40,41]. This integration not only substantially enhances model interpretability but also offers a reliable basis for feature selection and model optimization, thereby improving the scientific rigor and practical value of mineral prospectivity prediction while maintaining a high predictive performance.

Based on the aforementioned studies, this research takes the Mayoumu area as the study region and conducts mineral prospectivity prediction by integrating multi-source remote sensing data, geochemical information, and geological structural features. Three ensemble learning models—GBDT, XGBoost, and LightGBM—were constructed and applied to the Mayoumu area; notably, all three are boosting-based learning strategies that iteratively combine multiple weak learners to form a strong predictor. Furthermore, the SHAP method was employed to interpret key feature variables within the models, thereby elucidating the contribution mechanisms of different mineralization indicators to the prediction results. In addition, confusion matrices and multiple classification performance metrics were used to systematically evaluate the predictive capabilities of different machine learning models, enabling improved interpretability and the selection of the optimal model for the study area. On this basis, an integrated mineral prospectivity mapping framework

was developed, and a fused mineral prospectivity map (MPM) was generated, providing a new methodological approach and theoretical support for subsequent mineral exploration in the Mayoumu area (Figure 1).

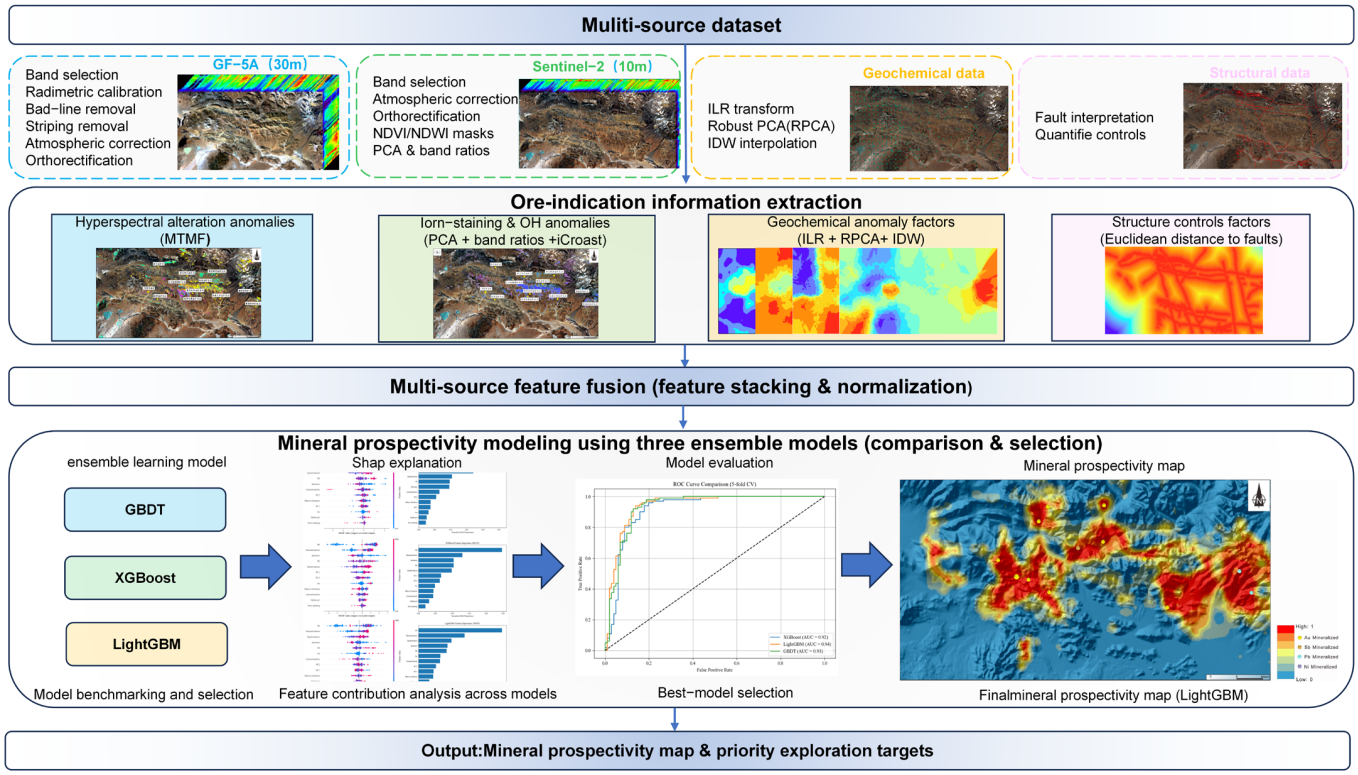


Figure 1. The technical process framework of this study.

2. Data Overview

2.1. Geological Profile

The Mayoumu area is located within the Zhongba microblock of the western segment of the Yarlung Zangbo suture zone in Tibet and has attracted considerable attention following the discovery of the Mayoumu gold deposit, the first orogenic gold deposit identified in Tibet. The region has undergone intense tectonic movements, accompanied by frequent magmatic activities and recent hydrothermal processes, which provide favorable conditions for the formation of large to super-large gold deposits. Recent exploration efforts have revealed multiple mineralized points and prospecting clues, indicating a favorable metallogenic setting and significant mineral exploration potential in the area (Figure 2).

The study area exhibits well-exposed and widespread stratigraphy, including Paleozoic, Mesozoic, and Cenozoic units (Figure 2). Paleozoic strata are predominantly distributed in the central part of the region; Mesozoic units are mainly concentrated in the southern part; and Cenozoic strata are extensively exposed throughout the area. The Paleozoic sequence comprises the Jiangmunong Formation ($S_{2-4}j$), Mayoumu Formation ($D_{1-3}m$), Kangtuo Formation (C_1k), and Quqa Formation (P_2q); the Mesozoic sequence includes the Qiongguo Formation ($T_{1-2}q$), Xiukang Formation (T_3x), and Gaxue Formation (K_1g); the Cenozoic sequence comprises the Liuqu Formation ($E_{1-2}l$), Dazhuoka Formation (E_3N_1d), and Quaternary alluvial deposits (Q) (Figure 2).

Intrusive rocks are widely distributed in the area, exhibiting diverse lithologies ranging from intermediate to acidic compositions, predominantly granite, with emplacement ages spanning the Late Cretaceous, Eocene, and Miocene. The alteration types of the mineralized

host rocks are varied, including epidotization, silicification, sericitization, hematitization, and limonitization.

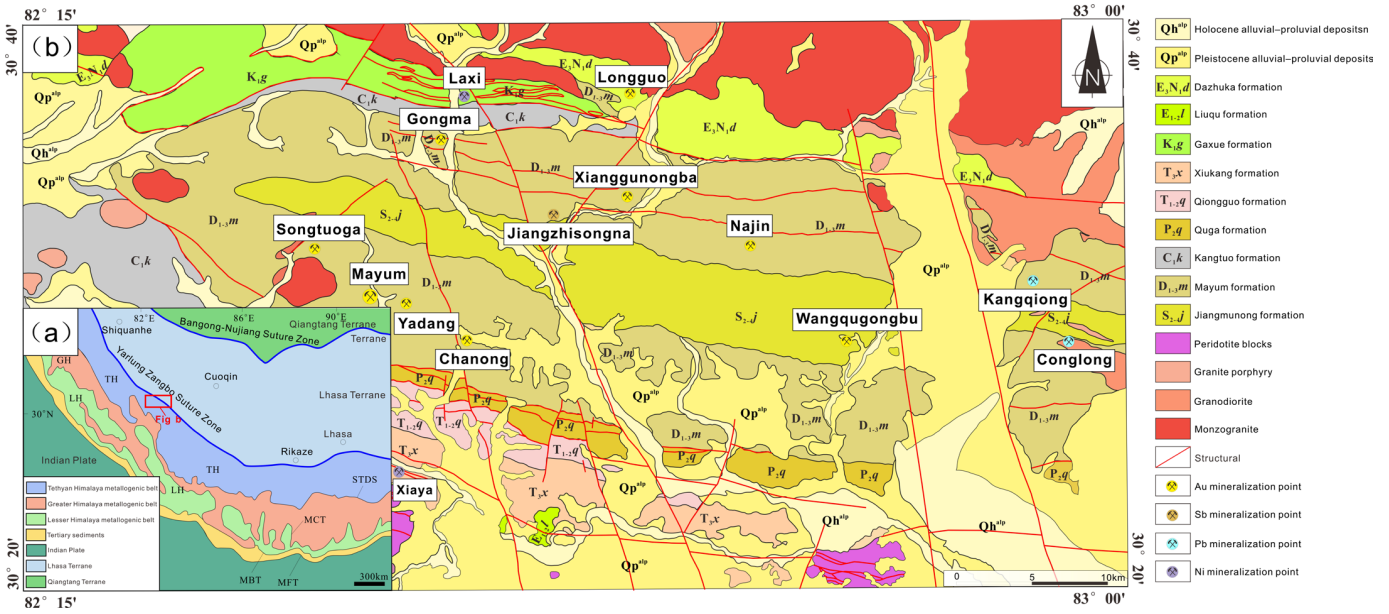


Figure 2. Tectonic map (**a**) and geological map (**b**) of the Mayoumu area [18].

2.2. Overview and Preprocessing of Multispectral Data

The Sentinel-2 satellite [42] is part of the European Union's Copernicus Earth Observation Program and carries a 13-band imager covering visible, near-infrared, and shortwave-infrared wavelengths, with spatial resolutions of 10 m, 20 m, and 60 m, respectively. In this study, Sentinel-2 Level-2A (L2A) imagery acquired on 7 June 2022, with cloud cover below 5%, was used, providing data that meet the requirements of this research.

The radiation information received by the sensor is typically influenced by terrain geometry, atmospheric effects, and sensor characteristics. Therefore, image preprocessing is necessary prior to extracting alteration anomalies. Standard preprocessing for multispectral remote sensing data includes radiometric calibration, atmospheric correction, and orthorectification. Sentinel-2 L2A products already provide surface reflectance data with these corrections applied. However, low-spatial-resolution bands still require spatial enhancement for subsequent analyses. To address this, the S2 Resampling tool in SNAP software (version 11.0.0) was used to perform unmixing-based reconstruction of low-resolution bands while preserving spectral integrity, improving their spatial resolution to 10 m to meet analytical requirements.

In multispectral remote sensing-based extraction of alteration anomalies, non-geological factors such as clouds/cloud shadows, water bodies, vegetation cover, snow, and topographic shadowing can substantially distort pixel spectral shapes and brightness, leading to spurious anomalies or artificial enhancement/attenuation in PCA outputs and thereby degrading alteration-detection accuracy. To minimize these effects, we constructed a hierarchical masking scheme based on Sentinel-2 L2A surface reflectance data and implemented it following the logic of "remove invalid observations first, then exclude non-geological targets, and finally mitigate low-illumination effects". First, cloud- and cloud shadow-contaminated pixels were identified and removed using the quality-control layers embedded in the Sentinel-2 product, specifically the scene classification layer (SCL). Next, the normalized difference water index (NDWI) and normalized difference vegetation index (NDVI) were calculated to generate water and vegetation masks, thereby reducing interference from low-reflectance water surfaces and vegetation cover on alteration

signals [43]. In addition, because snow and shadow in high-altitude environments often induce anomalous responses across the visible to shortwave-infrared bands, a density-slicing method was applied to delineate snow- and shadow-affected areas. A Quaternary (alluvial–proluvial) cover mask was further constructed based on the mapped extent of Quaternary deposits to reduce the confusion of alteration indicators caused by unconsolidated surficial cover. Finally, all masks were combined using a union operation and applied uniformly to the Sentinel-2 imagery with the ENVI masking tool, yielding multispectral data with non-geological disturbances effectively removed and providing a reliable input for subsequent alteration anomaly extraction and integrated analysis.

2.3. Overview and Preprocessing of Hyperspectral Data

The GF-5 satellite [44], successfully launched in 2018, is a key component of China's high-resolution Earth observation system. The satellite is equipped with six types of sensors, among which this study primarily utilizes the advanced hyperspectral imager (AHSI), which covers the visible to shortwave-infrared range. The AHSI sensor contains 330 spectral bands with a spectral resolution of 5–10 nm and a spatial resolution of 30 m. Multiple GF-5 scenes were mosaicked to generate the datasets used in this study, with acquisition dates of 7 June 2023, 7 June 2024, 7 November 2024, and 14 November 2024. Cloud cover was low, and the data met the requirements of this study. Preprocessing of the GF-5 hyperspectral data included band selection, radiometric calibration, bad line correction, stripe removal, atmospheric correction, orthorectification, and spectral smoothing, all conducted using ENVI 5.3 software.

Due to sensor limitations and atmospheric interference, certain spectral bands in the hyperspectral imagery provide insufficient information for subsequent analyses. Therefore, low-quality bands were excluded prior to extracting surface information. Specifically, bands 193–200 and 246–264 were strongly affected by water vapor absorption and thus removed. In the near-infrared (NIR) region, bands 145–150 overlap with shortwave-infrared (SWIR) bands 151–154. Considering the higher signal-to-noise ratio of the NIR bands, bands 145–150 were retained while bands 151–154 were discarded. Additionally, band 325 and subsequent bands exhibited low signal-to-noise ratios and were also removed. After band selection, a total of 283 effective bands were retained for subsequent processing.

Some bands in the GF-5 hyperspectral imagery contained noise in the form of bad lines and stripes, which degraded overall image quality. Bad lines in GF-5 primarily appear as column defects, where one or several columns of pixels exhibit abnormally low values or missing data. These were corrected by replacing the affected pixels with the mean value of adjacent columns, with pre- and post-correction results shown in Figure 3a,b. Unlike bad lines, stripes usually affect broader areas where pixel values remain within a reasonable digital number (DN) range but are systematically lower and exhibit irregular fluctuations relative to surrounding pixels, with variable stripe widths. To suppress stripe artifacts, we applied a global, column-wise destriping correction. The full column of pixels was used as the statistical window, and the gain and offset were estimated from the column-wise mean and standard deviation. A linear radiometric normalization was then performed column by column to enforce radiometric consistency and reduce stripe noise. The destriping results before and after correction are shown in Figure 3c,d.

Consistent with the preprocessing applied to the multispectral data, atmospheric correction of the GF-5 imagery was performed using the FLAASH module in ENVI (version 5.6) to mitigate absorption and scattering effects caused by atmospheric gases such as CO_2, O_2, and O_3. The atmospherically corrected imagery was then orthorectified using Copernicus digital elevation model (DEM) data at 12.5 m spatial resolution to correct image tilt and projection distortions induced by terrain relief and sensor geometric errors. Finally,

the imagery was smoothed using a Savitzky–Golay (SG) filter. Based on repeated tests, the optimal parameters were a half-window size of 5 points on each side of the filter kernel and a polynomial order of 4, which effectively removed jagged noise and completed the GF-5 preprocessing workflow.

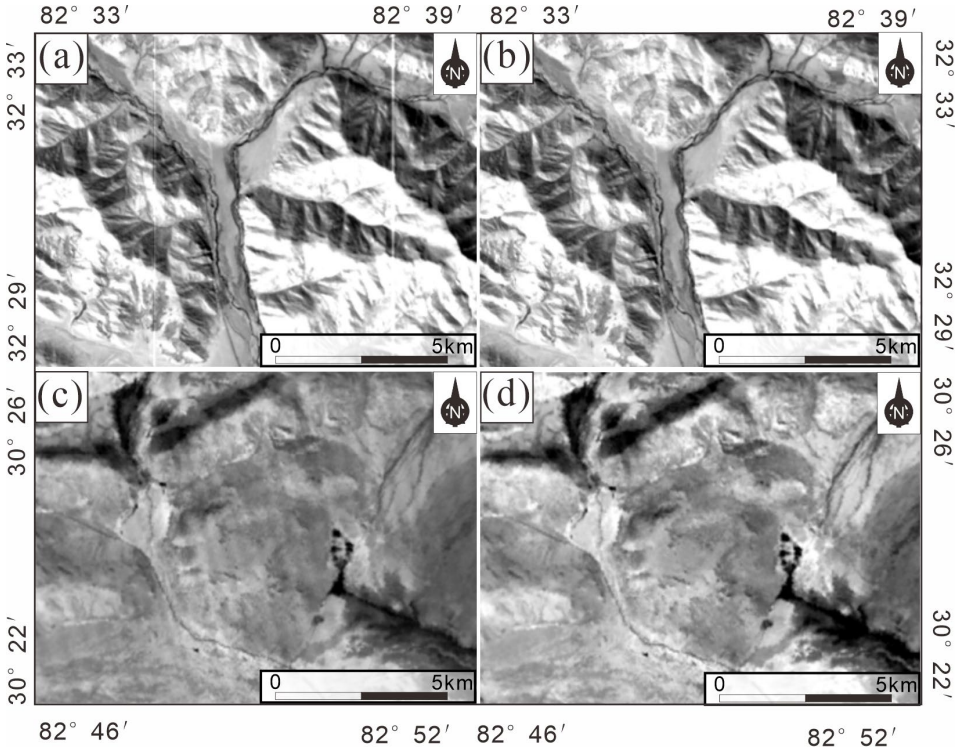


Figure 3. Comparison of GF-5A data from the Mayoumu area before (**a**) and after (**b**) bad line correction, and before (**c**) and after (**d**) stripe removal.

2.4. Overview of the Geochemical Dataset

The data used in this study were derived from regional geochemical surveys in Southwest China. The geochemical survey primarily focused on stream sediment sampling, supplemented by soil sampling. Sampling was conducted at a scale of 1:200,000, with a typical density of 1–2 points per km^2, which was reduced to 1 point per 4 km^2 in areas with complex terrain or difficult field conditions. All samples were collected and processed under contamination-free conditions and analyzed using a standardized multi-element, multi-method protocol with strict quality control procedures, covering a total of 32 elements and 7 oxides.

For this study, the regional dataset was clipped to the Mayoumu area in the Tibet Autonomous Region, yielding 698 stream sediment samples at a 1:200,000 scale. As the Mayoumu area is a typical hydrothermal metallogenic zone, nine hydrothermal mineralization indicator elements—Au, Ag, Cu, Pb, Zn, Sb, As, Hg, and Ti—were selected for detailed geochemical analysis based on regional geology and metallogenic characteristics.

3. Materials and Methods

Building on the multi-source data acquisition and the screening of geochemical indicators described above, this study further presents the workflow for constructing the mineral prospectivity prediction models, including the principles of the ensemble learning algorithms employed, the sample and feature input scheme, and the model training and validation strategies.

3.1. Ensemble Learning for Prospectivity Modeling

3.1.1. GBDT Modeling

GBDT [45] is a tree-based ensemble learning method that follows a Boosting strategy and is widely used for both classification and regression tasks. In Boosting, multiple weak learners (typically shallow decision trees) are added sequentially to form a strong learner by progressively correcting the errors made by the existing ensemble. Specifically, at each iteration, a new learner is trained along the negative gradient direction of the loss function with respect to the current model, thereby reducing the overall prediction error until convergence. The core objective of GBDT is to minimize the expected value of a specific loss function $L(y, F(x))$, obtaining an approximate representation $\widetilde{F}(x)$ of the true mapping function $F(x)$, thereby optimally fitting the relationship between input variables x and the target variable y [46].

For binary classification, GBDT uses the negative binomial log-likelihood as the loss function, which can be expressed as:

$$L(y, F(x)) = log(1 + exp(-2yF(x))) \tag{1}$$

The training process of GBDT can be summarized as follows:

(a) Initialize the model:

$$F_0 = \frac{1}{2} log \frac{1+\widetilde{y}}{1-y}, \ \overline{y} = \frac{1}{N}\sum_{i=1}^{N} y_i \tag{2}$$

(b) Compute the negative gradient residuals (pseudo-responses):

$$\widetilde{y}_i = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F=F_{m-1}(x)} = \frac{2y_i}{1 + exp(2y_i F_{m-1}(x_i))} \tag{3}$$

(c) Fit the pseudo-responses using regression trees as base learners and partition samples into J terminal regions:

$$\{R_{jm}\}_{j=1}^{J} = TerminalRegions\left(h_m; \left\{\widetilde{y}_i, x_i\right\}_{i=1}^{N}\right) \tag{4}$$

(d) Compute the output value for each terminal region:

$$\gamma_{jm} = \sum_{x_i \in R_{jm}} \widetilde{y}_i / \sum_{x_i \in R_{jm}} \left|\widetilde{y}_i\right| (2 - \left|\widetilde{y}_i\right|) \tag{5}$$

(e) Update the model:

$$F_m(x) = F_{m-1}(x) + \sum_{j=1}^{J} \gamma_{jm} I\ (x_i \in R_{jm}) \tag{6}$$

where I is the indicator function: $I = 1$ if the condition is true, otherwise $I = 0$. The final probability can be estimated using a log-odds transformation:

$$p_+(x) = \widehat{Pr}(y = 1|x) = 1/\left(1 + e^{-2F_m(x)}\right) \tag{7}$$

$$p_-(x) = \widehat{Pr}(y = -1|x) = 1 - p_+(x) \tag{8}$$

3.1.2. XGBoost Modeling

XGBoost was proposed by Chen & Guestrin [47] in 2016. It is an iterative Boosting ensemble algorithm developed on the framework of GBDT, progressively integrating multiple classification and regression tree (CART) trees in an additive manner. XGBoost constructs a strong learner by assigning different weights to multiple weak learners, offering advantages such as fast training, high prediction accuracy, and strong capability for handling large-scale sparse data. During training, each new tree fits the residuals between the target value and the sum of predictions from all previous trees. Tree growth is guided by the gain in the objective function to select optimal features and split points. The final prediction is obtained by a weighted summation of all trees' outputs.

Let K be the number of trees; the overall model can be formalized as:

$$\widehat{y_i} = \sum_{k=1}^{K} f_k(x_i), \quad f_k \in F \tag{9}$$

where F denotes the space of CART functions, and f_k represents the k-th base learner. Model training minimizes an objective function composed of empirical loss and a regularization term:

$$L(\varnothing) = \sum_i L\left(\widetilde{y_i}, y_i\right) + \sum_k \Omega(f_k) \tag{10}$$

$$\Omega(f) = \gamma\, T + \frac{1}{2}\lambda\ \ \omega^2 \tag{11}$$

Here, L measures the difference between the prediction $\widetilde{y_i}$ and the true label y_i; $\Omega(f)$ penalizes model complexity, with λ and γ as regularization parameters, T the number of leaf nodes, and ω the weight of each leaf. Following Chen & Guestrin [47], the core steps of the algorithm are:

(a) Additive Objective Formulation

At iteration t, the objective function is:

$$L^{(t)} \simeq \sum_{i=1}^{n} l\left(y_i, \widetilde{y_i}^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) \tag{12}$$

(b) Second-order Taylor Expansion

Expanding at $\widetilde{y_i}^{(t-1)}$ gives:

$$L^{(t)} \simeq \sum_{i=1}^{n}\left[l\left(y_i, \widetilde{y_i}^{(t-1)}\right) + g_i f_t(x_i) + \frac{1}{2}h_i f_t^2(x_i)\right] + \Omega(f_t) \tag{13}$$

(c) Objective Simplification

Removing constant terms yields:

$$\widetilde{L}^{(t)} = \sum_{i=1}^{n}\left[g_i f_t(x_i) + \frac{1}{2}h_i f_t^2(x_i)\right] + \Omega(f_t) \tag{14}$$

(d) Split Candidate Evaluation

For a candidate tree structure q with leaf node sample sets $I_j = i \mid q(x_i) = j$:

$$\widetilde{L}^{(t)}(q) = -\frac{1}{2}\sum_{j=1}^{T}\frac{\left(\sum_{i\in I_j} g_i\right)^2}{\sum_{i\in I_j} h_i + \lambda} + \gamma T \tag{15}$$

For a candidate split, dividing node I into left I_L and right I_R child nodes, the gain is:

$$L_{split} = -\frac{1}{2}\left[\frac{\left(\sum_{i\in I_L} g_i\right)^2}{\sum_{i\in I_L} h_i + \lambda} + \frac{\left(\sum_{i\in I_R} g_i\right)^2}{\sum_{i\in I_R} h_i + \lambda} - \frac{\left(\sum_{i\in I} g_i\right)^2}{\sum_{i\in I} h_i + \lambda}\right] - \gamma \quad (16)$$

These steps iterate until the pre-set number of rounds is reached or convergence is achieved. Detailed derivations are provided in Chen & Guestrin [47].

3.1.3. LightGBM Modeling

LightGBM is an efficient GBDT framework proposed by Microsoft in 2017 [48]. By introducing several improvements to traditional GBDT, LightGBM exhibits superior performance in large-scale data processing scenarios. As an efficient boosting algorithm, LightGBM adopts a histogram-based splitting method and a leaf-wise tree growth strategy, combined with a maximum depth constraint to accelerate training and reduce memory consumption. Consequently, LightGBM is well-suited for both classification and regression tasks, achieving high predictive accuracy while substantially improving training efficiency and resource utilization. The histogram algorithm discretizes continuous feature values into a finite number of bins and accumulates gradient statistics and sample counts within each bin in a single data scan, thereby reducing storage and computational costs during split searching.

In terms of objective function design, LightGBM follows a formulation similar to XGBoost by incorporating regularization terms to shrink the update step size and suppress overfitting [48]. The training procedure of LightGBM can be summarized as follows.

(a) Objective Function Definition

For the t-th tree, the objective function is defined as:

$$obj^t = \left[\sum_{i=1}^{N} L(y_i, \tilde{y}^{(t-1)} + p)\right] + \frac{1}{2}\lambda p^2 + \gamma T \quad (17)$$

where L denotes the loss function, N is the total number of samples, y_i is the true label of the i-th sample, $\tilde{y}^{(t-1)}$ represents the prediction from the previous iteration, P is the output of the t-th tree, λ is the regularization parameter controlling the step size, γ is a pruning-related regularization hyperparameter, and T denotes the number of leaf nodes in the t-th tree.

(b) Second-order Taylor Expansion

The loss function is approximated using a second-order Taylor expansion:

$$L\left(y_i, \tilde{y}^{(t-1)} + P\right) \approx L\left(y_i, \tilde{y}^{(t-1)}\right) + L'\left(y_i, \tilde{y}^{(t-1)}\right)P + \frac{1}{2}L''\left(y_i, \tilde{y}^{(t-1)}\right)P^2 \quad (18)$$

where $L'(\cdot)$ and $L''(\cdot)$ denote the first-order gradient and second-order Hessian, respectively.

(c) Objective Function Simplification

By aggregating the first-order gradients g_i and second-order derivatives h_i of all samples, the objective function can be rewritten as:

$$obj^t = \frac{1}{2}(H + \lambda)P^2 + GP \quad (19)$$

$$G = \sum_{i=1}^{N} g_i \quad (20)$$

$$H = \sum_{i=1}^{N} h_i \tag{21}$$

(d) Optimal Leaf Output and Model Update

By taking the derivative of Equation (19) with respect to P and setting it to zero, the optimal leaf output for the t-th tree is obtained as:

$$P = -\frac{G}{H + \lambda} \tag{22}$$

For regression tasks, the squared loss function is commonly adopted:

$$L(y, P) = \frac{1}{2}(y - P)^2 \tag{23}$$

3.2. Shapley Additive Explanation

SHAP is an additive feature attribution method grounded in Shapley values, which Lloyd Shapley [49,50] originally proposed within the framework of cooperative game theory to quantify the contribution of individual players to a collective outcome. In recent years, Lundberg and Lee [51] introduced this concept into the field of machine learning, developing a unified and efficient framework for feature attribution that enables the transparent interpretation of model predictions. Thus far, SHAP has been widely applied to the interpretability analysis of various models, including both regression and classification tasks. Prior to the widespread adoption of SHAP, the interpretation of ensemble learning models such as GBDT, XGBoost, and LightGBM mainly relied on feature importance measures and partial dependence plots (PDPs). Although feature importance can intuitively quantify the overall contribution of each feature and identify those that exert a strong influence on model performance, it fails to reveal the specific relationships between individual features and the prediction outcome at the single-sample level. In contrast, SHAP assigns an attribution value to each feature for every prediction, thereby enabling fine-grained, instance-level explanations and quantifying both the magnitude and direction (positive or negative) of each feature's contribution to a specific prediction. Consequently, SHAP provides a more detailed and explicit interpretation of complex machine learning models.

In explainable machine learning, Shapley values serve as an additive feature attribution approach that expresses a model's prediction as a linear combination of simplified binary input features, formulated as:

$$g(x') = \varnothing_0 + \sum_{i=1}^{M} \varnothing_i x_i' \tag{24}$$

where x_i' denotes the binary representation of the original input feature x, M is the total number of simplified input features, ϕ represents the Shapley values, ϕ_0 is a constant term corresponding to the expected (mean) prediction of the original model, and ϕ_i denotes the contribution of the i-th feature to the prediction for a given sample x (i.e., the SHAP value). Accordingly, the SHAP-based decomposition of the prediction for an individual sample can be expressed as:

$$y_i = y_{base} + f(x_{i1}) + f(x_{i2}) + \cdots + f(x_{ik}) \tag{25}$$

where x_i denotes the i-th sample, x_{ik} represents the k-th feature of the i-th sample, and y_{base} is the baseline prediction of the model, defined as the average prediction across all samples. The term $f(x_{ik})$ corresponds to the SHAP value of the feature x_{ik}, which quantifies the contribution of the k-th feature to the prediction of the i-th sample. Specifically, each SHAP

value reflects the difference between the model prediction and the baseline value when the corresponding feature is taken into account. A positive SHAP value ($f(x_{ik}) > 0$) indicates that the feature contributes positively to the prediction, whereas a negative SHAP value implies a negative contribution.

4. Results

4.1. Extraction of Multi-Source Exploration Information

4.1.1. Extraction of Alteration Anomalies from Sentinel-2 Data

Previous studies have demonstrated that principal component analysis (PCA) [52–54] is one of the most effective and widely used techniques for extracting alteration information from multispectral remote sensing data. By applying an orthogonal transformation, PCA compresses highly correlated spectral information into a limited number of principal components, thereby enhancing subtle spectral features related to geological alteration. In addition, band–ratio-based approaches can further emphasize weak alteration signals by exploiting reflectance and absorption contrasts among specific spectral bands.

In this study, Sentinel-2 multispectral imagery was used to extract alteration anomalies by integrating PCA with band combinations based on the characteristic spectral responses of iron-stained and hydroxyl-bearing minerals (Figure 4). Iron-stained minerals exhibit absorption features at 0.48–0.51 μm and 0.85–0.89 μm (Bands 2 and 8A), and strong reflectance at 0.65–0.70 μm and 1.52–1.70 μm (Bands 4 and 11). Accordingly, Bands 2, 4, 8A, and 11 were selected to construct the band set for iron-staining alteration. PCA results indicate that the corresponding alteration information is mainly concentrated in the inverted fourth principal component (−PC4), which was therefore used as the indicator for iron-staining alteration anomalies (Table 1).

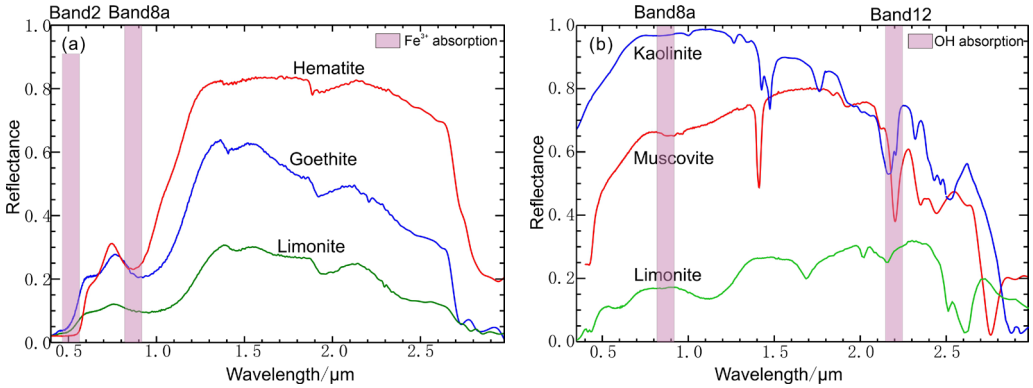


Figure 4. Spectral reflectance curves of USGS spectral library [55] minerals: (**a**) iron-stained; (**b**) Hydroxyl.

Table 1. Eigenvectors of principal component analysis for Sentinel-2 imagery in the Mayoumu area.

Sentinel-2	Eigenvector	Band 2	Band 4	Band 8A	Band 11
Iron staining	PC1	−0.286269	−0.441432	−0.525123	−0.668905
	PC2	−0.504497	−0.491926	−0.184289	0.685222
	PC3	0.813534	−0.429239	−0.337488	0.200047
	PC4	0.041172	−0.615549	0.759200	−0.207409
Sentinel-2	Eigenvector	Band 2	Band 8A	Band 11	Band 12
Hydroxyl	PC1	0.264576	0.487298	0.624152	0.550431
	PC2	0.71592	0.477385	−0.412515	−0.298988
	PC3	0.502718	−0.606154	−0.240346	0.567525
	PC4	−0.405873	0.408927	−0.618465	0.534366

Hydroxyl-bearing minerals are characterized by absorption features at 0.86–0.96 μm and 2.0–2.5 μm (Bands 8A and 12), and reflectance features at approximately 0.50 μm and 1.60–1.70 μm (Bands 2 and 11). Thus, Bands 2, 8A, 11, and 12 were selected to construct the hydroxyl alteration band combination. PCA results similarly show that hydroxyl alteration information is predominantly captured by the inverted fourth principal component (−PC4).

For anomaly identification and grading, the iCrosta method was applied to the PCA outputs to determine thresholds and extract alteration anomalies. First, histogram statistics were computed for the principal-component pixel values representing the hydroxyl- and ferric iron-related anomalies, from which the mean (μ) and standard deviation (σ) were derived. A dynamic thresholding range was then defined using $\mu + n\sigma$. Following the standard deviation-based classification scheme, anomaly intensity was divided into three levels, with thresholds of $\mu + 3\sigma$ (Level I), $\mu + 2.5\sigma$ (Level II), and $\mu + 2\sigma$ (Level III). The extracted anomaly maps were subsequently refined using a median filter to remove isolated speckle pixels and enhance spatial continuity. This procedure ultimately produced the Sentinel-2 hydroxyl and ferric iron alteration anomaly maps for the study area (Figure 5).

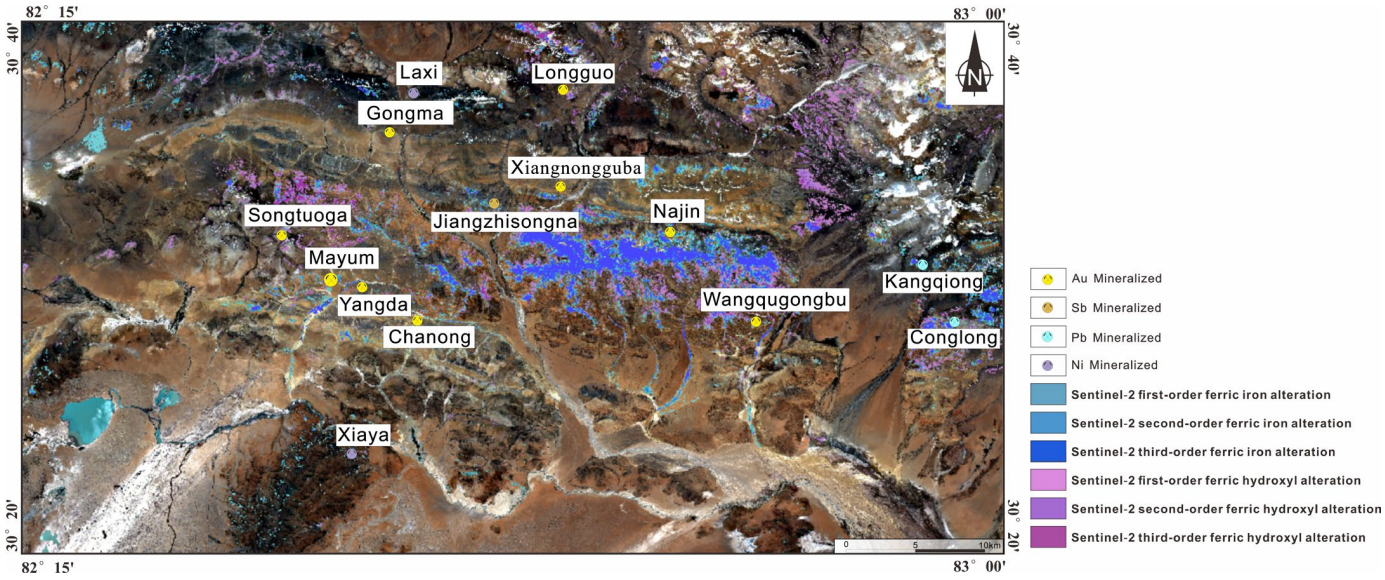


Figure 5. Remote sensing alteration information extraction results from Sentinel-2 data in the Mayoumu area.

4.1.2. Extraction of Alteration Anomalies from GF-5 Hyperspectral Data

The spectral signature of a remote sensing pixel represents a mixture of the spectral characteristics of multiple endmembers within that pixel. When different surface materials coexist within a single pixel, mixed pixels are inevitably formed [12,56]. Mixed pixels are particularly prevalent in hyperspectral imagery and can significantly reduce the accuracy and reliability of mineral identification during alteration mapping. Given the 30 m spatial resolution of GF-5 hyperspectral data, mixed pixels are unavoidable in the study area. Therefore, spectral unmixing is required to extract meaningful endmember information, suppress interference from non-target materials, and enhance alteration signal detection.

Among various spectral unmixing approaches, linear spectral mixture analysis (LSMA) is one of the most widely used and easily implemented models [57]. This model assumes that the spectral response of a pixel in a given band is a linear combination of the reflectance contributions from individual endmembers, proportional to their areal fractions within the pixel. The unmixing results are typically expressed as abundance maps, in which pixel values indicate the relative proportion of each endmember. The selection, number,

and representativeness of endmembers are critical factors controlling the accuracy and effectiveness of linear unmixing.

After data preprocessing, the minimum noise fraction (MNF) transformation was first applied to the GF-5 hyperspectral imagery to reduce noise and dimensionality, thereby minimizing spectral redundancy and extracting the dominant spectral information. Based on the MNF results, the pure pixel index (PPI) algorithm was then employed to identify spectrally pure pixels by setting appropriate iteration numbers and threshold values. The extracted endmember spectra were subsequently compared with reference spectra from the USGS spectral library in ENVI to determine the corresponding mineral types.

Following endmember identification, mixture-tuned matched filtering (MTMF) [11] was applied to the GF-5 hyperspectral data to map alteration minerals. MTMF integrates linear spectral unmixing with matched filtering, thereby enforcing physical feasibility while efficiently matching known endmember spectra, and enabling sub-pixel estimation of alteration-mineral abundances. The MTMF score images and infeasibility indices were jointly evaluated, and pixels with high scores and low infeasibility were selected as reliable regions of interest (ROIs). Based on these ROIs, spatial distribution maps of the alteration minerals were generated. Ultimately, maps of the major alteration minerals—limonite, hematite, muscovite, and epidote—were produced for the study area (Figure 6).

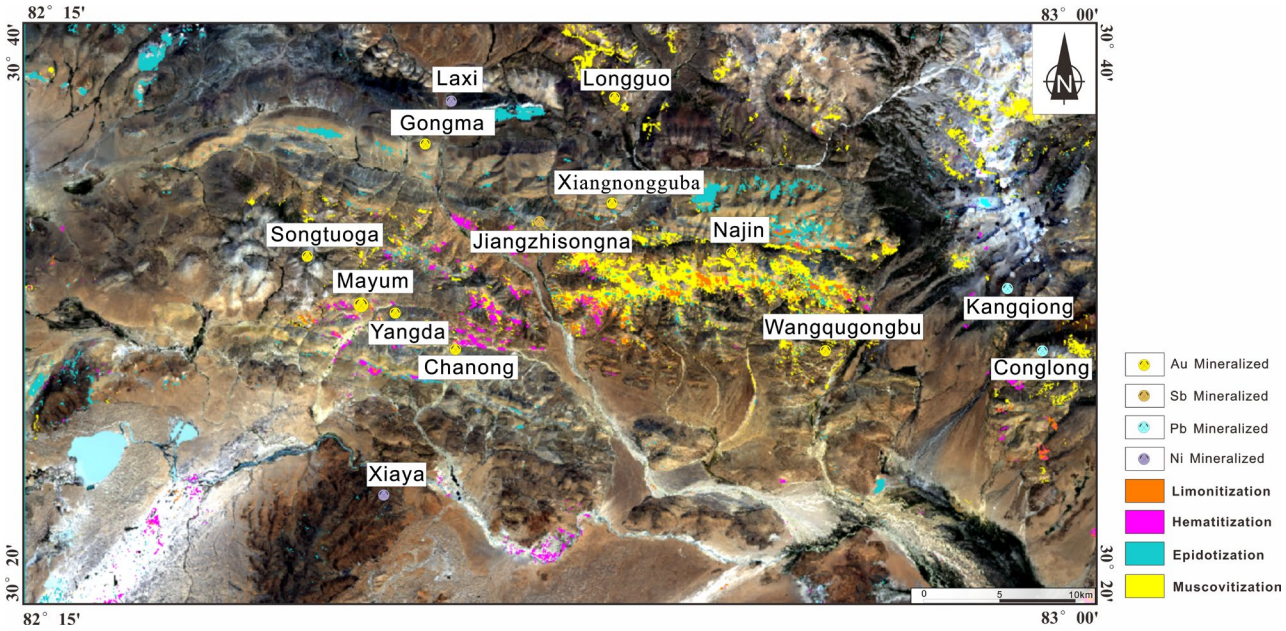


Figure 6. Remote sensing alteration information extraction results from GF-5 data in the Mayoumu area.

4.1.3. Extraction of 1:200,000-Scale Geochemical Anomalies

Geochemical data are typical compositional data [58,59], for which most elemental concentrations do not follow a normal distribution. In conventional geochemical studies, logarithmic transformation is commonly applied to approximate normality; however, log-transformed data are still affected by the closure effect, which hampers the identification of geochemical element associations with clear metallogenic significance. To overcome this limitation, this study adopts a combined approach of isometric log-ratio (ILR) transformation and robust principal component analysis (RPCA) [60], which effectively removes closure effects while extracting principal components related to mineralization.

In this study, the geochemical dataset comprising 698 samples from the study area was analyzed in R using ILR transformation and RPCA. A biplot was employed to visualize

the PCA results, enabling rapid and reliable characterization of the internal data structure and clustering–dispersion patterns. Figure 7 shows the biplot of PC1 and PC2 derived from RPCA after ILR transformation. The results indicate that the ILR–RPCA framework effectively alleviates the compositional (closure) constraint, clarifying inter-element relationships and improving interpretability. The first two principal components (PC1 and PC2) have the largest eigenvalues and, together, explain 72.19% of the cumulative variance, representing the dominant geochemical variability in the original dataset. In terms of loading patterns, PC1 exhibits positive loadings for Au, Ag, Cu, Pb, Zn, and Ti, reflecting a geochemical association characterized by their co-enrichment, whereas PC2 shows positive loadings for Hg, Zn, Cu, and Ti, indicating another composite enrichment pattern involving Hg. Considering the known mineralization information and the regional metallogenic setting, Au, Sb, and Pb display strong spatial consistency with mineralized occurrences and thus provide effective indicators of mineralization. Based on these results, single-element anomalies of Au, Sb, and Pb, together with the PC1 and PC2 scores, were interpolated using inverse distance weighting (IDW) to generate continuous raster layers, which were subsequently used as inputs to the mineral prospectivity prediction models (Figure 8). For cartographic presentation and spatial comparison, the thematic classes in Figure 8 were defined using the Jenks natural breaks method with seven classes.

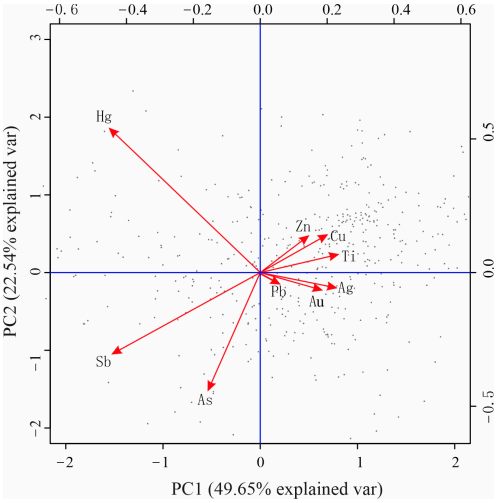


Figure 7. Biplot of geochemical RPCA (1:200,000 scale) in the Mayoum area.

4.1.4. Extraction of Structural Information

The study area is located in a key segment of the active tectonic belt in southern Tibet and has experienced intense deformation. The regional fault network not only governs fluid pathways and zones of stress concentration, but also generates highly permeable damage zones through geometric discontinuities such as fault intersections, bends, and step-overs, thereby providing favorable space for the migration, focusing, and precipitation of hydrothermal fluids. Faults in the Mayoumu area are well developed and display clear hierarchical and orientational patterns: E–W-trending master fault zones dominate, N–S-trending faults are secondary, and a smaller number of oblique NE–SW and NW–SE subsidiary faults are also present. Faults with different orientations commonly intersect, branch, and locally change strike, forming fault-junction areas and widened fracture corridors. These structural domains typically exhibit enhanced structural permeability and fluid-focusing effects and thus constitute favorable sites for Au–Sb–Pb hydrothermal mineralization.

Considering the regional tectonic framework, major thrust/nappe-related faults and their associated shear structures likely served as primary conduits for deep-sourced fluids,

whereas transpressional fractures, step-over zones, and intersection nodes generated by the superposition of oblique subsidiary faults and strike–slip components may further enhance permeability and promote ore localization. This interpretation is consistent with the observation that known mineralized occurrences in the study area are preferentially distributed in the vicinity of faults. To incorporate structural information into the mineral prospectivity models, we parameterized the fault vector data using Euclidean distance analysis to construct a "fault distance" predictor. The spatial distribution of fault distance is shown in Figure 8f, and this raster layer was subsequently used as an input variable for the mineral prospectivity prediction models.

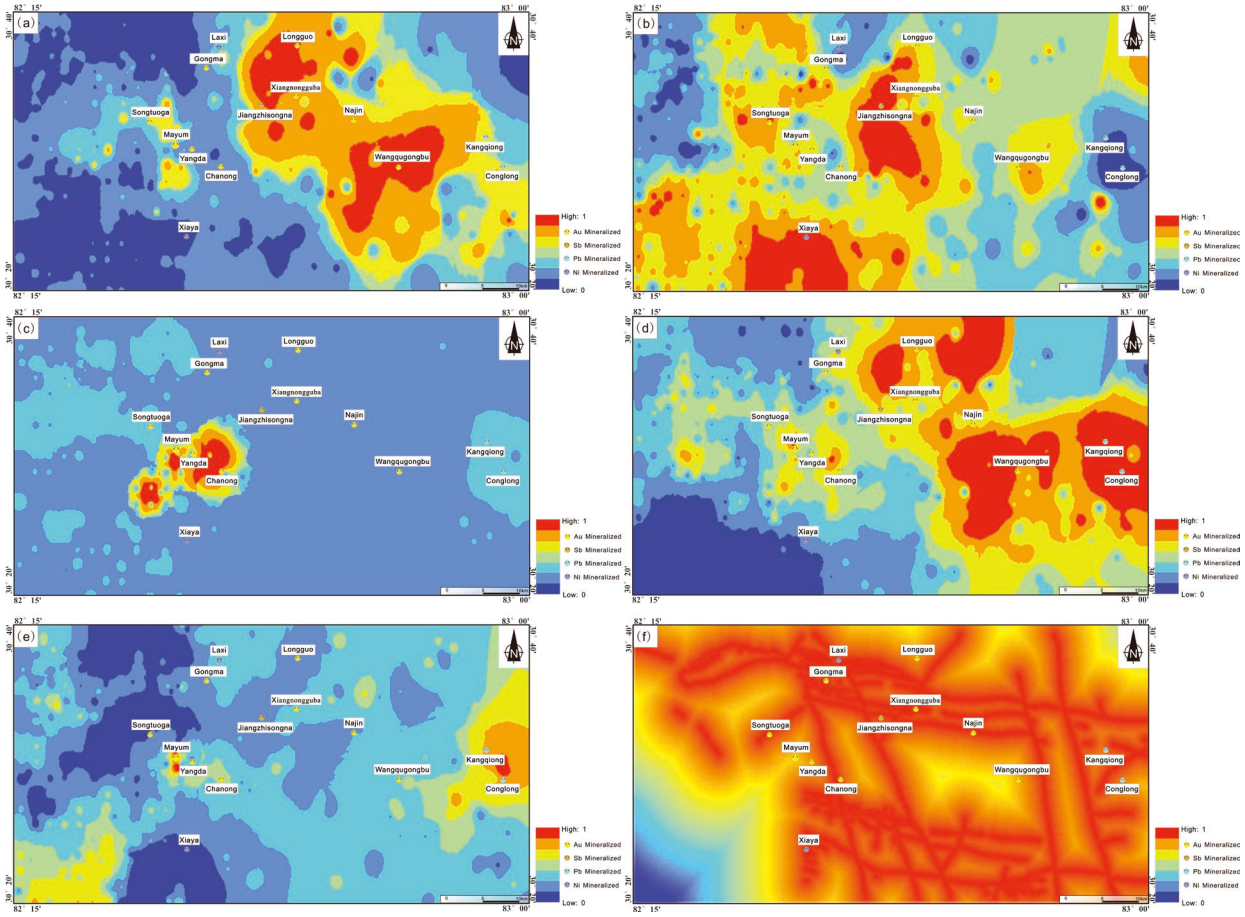


Figure 8. Geochemical analysis and structural interpretation of the Mayoumu area: (**a**) PC1, (**b**) PC2, (**c**) Au, (**d**) Sb, (**e**) Pb, and (**f**) interpreted structural features.

4.1.5. Sample Dataset Preparation

ArcGIS Pro (version 3.4) was used to generate a regular fishnet grid over the study area. To enable unified modeling of multi-scale datasets while balancing spatial representation and data stability, a 500 m × 500 m grid was adopted as the basic prediction unit, yielding 11,211 independent cells. This grid size is sufficient to capture regional spatial variations in anomalies, while avoiding extensive no-data values and interpolation instability that may arise from overly fine grids. During gridding, all predictor layers were first converted to raster format and the 10 predictors were harmonized in terms of spatial resolution and map projection. Predictor values were then extracted at the representative point of each grid cell (i.e., the cell centroid) to assign values to all 11,211 prediction units. This workflow unifies the spatial support scale while preserving the geological meaning and spatial heterogeneity of each predictor, resulting in a dataset comprising 11,211 samples and 10 quantitative predictors. To improve comparability among different types of predictors and to reduce

potential scale effects on model convergence and feature-contribution interpretation, all continuous predictors were standardized prior to machine-learning modeling to enhance training robustness and ensure consistency in interpretability analyses.

In mineral prospectivity prediction, the objective of machine-learning models is to determine whether a spatial unit is mineralized based on a combination of ore-controlling factors, which constitutes a typical binary classification problem. Model inference provides the probability distribution of mineralized versus non-mineralized units conditioned on the predictor variables. To date, a total of 14 polymetallic deposits/occurrences dominated by Au, Sb, and Pb have been identified in the Mayoumu area. Because mineralization is a typical spatial "rare event", the number of known mineralized occurrences is far smaller than that of non-mineralized units. Therefore, the selection of non-mineralized samples followed three principles: (1) prioritizing units with a relatively low mineralization potential based on regional metallogenic conditions; (2) ensuring a random and spatially uniform distribution; and (3) maintaining a sample size broadly comparable to that of mineralized samples to achieve class balance.

Given the limited number of known deposits/occurrences, to mitigate bias induced by class imbalance while avoiding an excessively large buffer that would overly restrict the candidate space for negative samples—and to keep the buffer scale consistent with the 500 m grid (approximately corresponding to the influence range of 1–2 grid cells)—a 750 m buffer was delineated around each occurrence, and the grid units within the buffer were treated as positive samples. Areas outside the buffers, together with the negative-anomaly zones on the geochemical anomaly maps, were defined as candidate regions for non-mineralized samples. Non-mineralized samples were then generated by random sampling within these candidate regions; moreover, to further reduce potential spatial autocorrelation caused by clustering of negative samples, a minimum spacing of >1000 m was enforced between any two randomly generated points. Ultimately, a balanced dataset comprising 101 mineralized units and 101 non-mineralized units was constructed for subsequent model training and evaluation.

4.2. Mineral Prospectivity Mapping Using Ensemble Learning Models

In this study, three ensemble learning algorithms—GBDT, XGBoost, and LightGBM —were employed for mineral prospectivity prediction, and binary classification models were implemented using relevant packages in the Python (version 3.10) ecosystem. To mitigate potential effects of heterogeneous predictor scales on model convergence and feature-contribution interpretation, all continuous predictors were standardized prior to model training using StandardScaler. Given the limited sample size and the need to preserve class proportions, stratified five-fold cross-validation was adopted for model training and generalization assessment (StratifiedKFold, n_splits = 5, shuffle = True, random_state = 42). Specifically, approximately 80% of the samples were used for training and 20% for validation in each fold, thereby maintaining consistent positive-to-negative class ratios across folds and reducing evaluation bias associated with class imbalance.

Regarding hyperparameter settings, we followed recommended default configurations and applied key constraints based on sample size and model-complexity control principles to reduce overfitting risk and ensure reproducibility. For LightGBM, tree complexity was regularized by limiting model depth and leaf growth (max_depth = 4, num_leaves = 20, min_child_samples = 5, n_estimators = 100, random_state = 42). For XGBoost, the log-loss function was used as the evaluation metric and a fixed random seed was set to ensure reproducible results (eval_metric = "logloss", random_state = 42). GBDT was trained and evaluated under the same cross-validation framework to enable consistent comparison, yielding stable generalization performance for all models under unified evaluation conditions.

4.2.1. Variable Importance Assessment

The SHAP method was applied to quantitatively evaluate the relative importance of predictive variables and to interpret their contributions to mineral prospectivity modeling. This approach not only generates prediction outputs for individual samples but also quantifies the positive or negative influence of each feature on model predictions, thereby providing insights into the roles of different geological and geochemical factors in mineralization processes.

SHAP values were calculated based on the standardized model outputs and visualized using SHAP beeswarm plots and feature importance ranking diagrams. As shown in Figure 9, the SHAP-based interpretations derived from the GBDT, XGBoost, and LightGBM models exhibit a high degree of overall consistency. All three models indicate that Sb anomalies represent the most influential predictive factor for mineral prospectivity in the study area. The SHAP values associated with both high and low Sb concentrations exert a pronounced effect on model outputs, suggesting a strong linkage between Sb enrichment and gold mineralization, and highlighting its significance as a key component of the regional hydrothermal mineralization system.

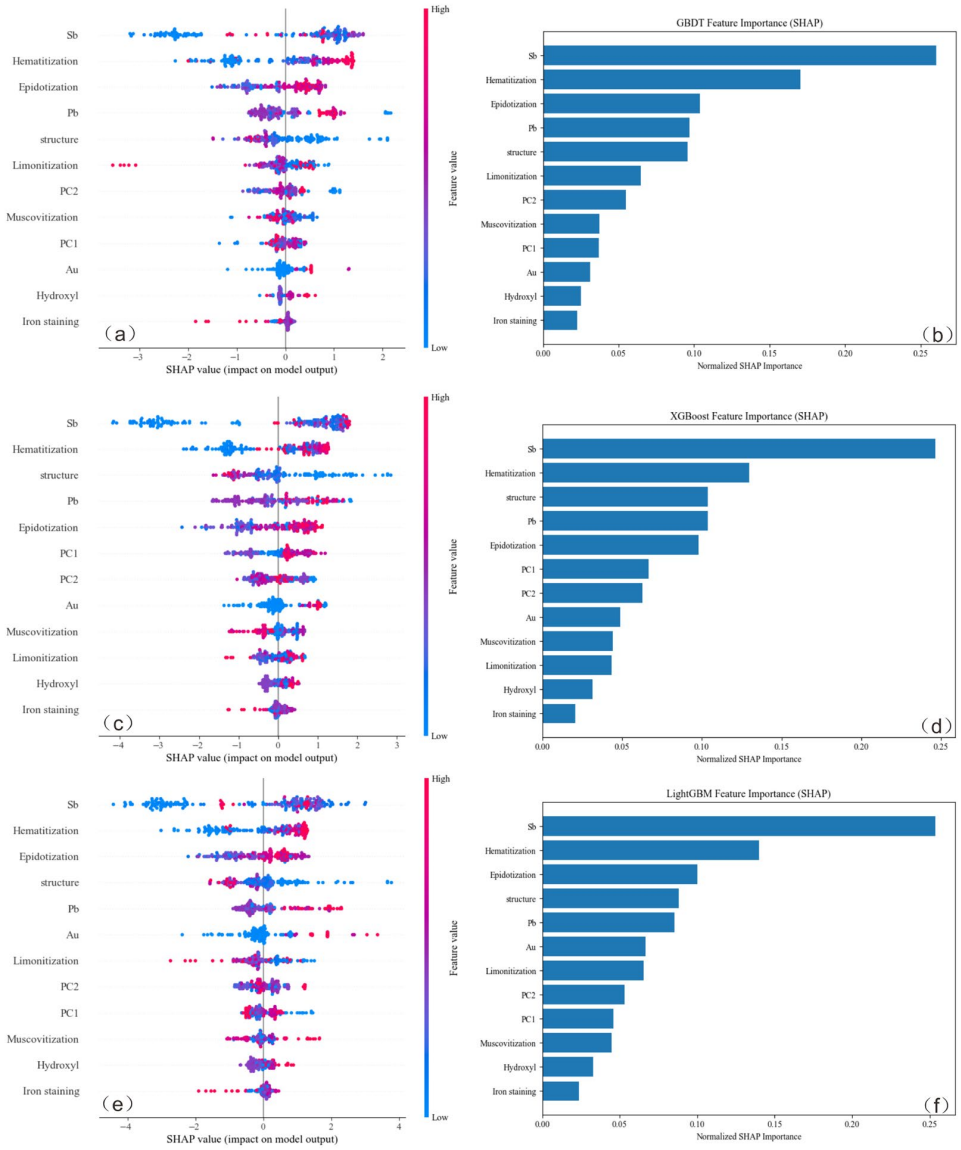


Figure 9. SHAP feature importance analysis results for GBDT (**a**,**b**), XGBoost (**c**,**d**), and LightGBM (**e**,**f**) models.

In addition to Sb, alteration mineral indicators such as hematite and epidote, together with fault-related structural factors, also display high importance across all three models. This observation suggests that the spatial distribution of mineral deposits in the study area is strongly controlled by tectonic activity and closely associated with low- to medium-temperature hydrothermal alteration processes. Although minor differences exist in the ranking of individual features among the models, their overall trends remain consistent, indicating robust and reliable identification of key ore-controlling factors. Moreover, predictive variables including Pb, Au, and principal component factors (PC1 and PC2) also contribute to mineral prospectivity prediction, underscoring the auxiliary indicative role of multi-element geochemical anomalies in integrated mineral prospectivity assessment.

Overall, the predicted probability of mineralization increases markedly with the number and combined contribution of key predictive factors. Based on the SHAP analysis, future mineral exploration in the Mayoumu area should prioritize geological settings characterized by pronounced Sb anomalies, in conjunction with hematite and epidote alteration signatures, and well-developed fault structures, to enhance target delineation accuracy and exploration efficiency.

4.2.2. Model Performance Evaluation

To further characterize the classification behavior of the models in mineral prospectivity prediction, confusion matrices together with multiple performance metrics were used to systematically evaluate the predictive capability of the different models. The confusion matrix provides an intuitive summary of the classification outcomes, enabling explicit quantification of false negatives (FNs) and false positives (FPs). In exploration practice, FNs are particularly undesirable because they may cause prospective targets to be missed, whereas FPs may increase follow-up field verification workload and associated costs.

The 5-fold combined confusion-matrix results indicate consistently strong performance across the three models. For GBDT, the matrix yields TN = 82, FP = 19, FN = 3, and TP = 98. For XGBoost, the corresponding values are TN = 81, FP = 20, FN = 4, and TP = 97. LightGBM shows the same overall pattern as GBDT (TN = 82, FP = 19, FN = 3, TP = 98). Overall, all models produced very low FN counts, suggesting a strong ability to identify mineralized units and thus reducing the risk of overlooking potential targets—an important consideration for mineral exploration. Meanwhile, the FP counts remain moderate, implying that the additional verification burden is controllable under practical exploration constraints (Figure 10).

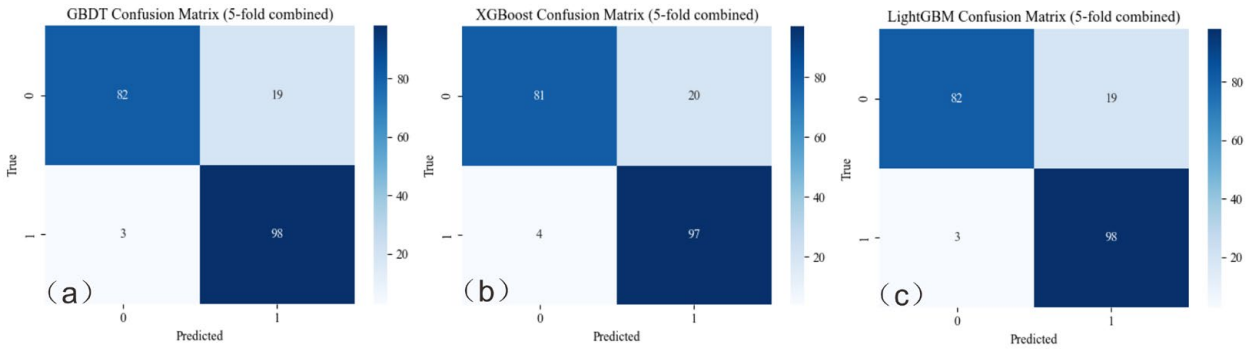


Figure 10. Confusion-matrix results for GBDT (**a**), XGBoost (**b**), and LightGBM (**c**) models.

Based on the confusion matrix, model performance was quantitatively evaluated using accuracy, precision, recall, and the F1-score. Recall reflects the model's ability to identify mineralized samples and is a key indicator for reducing the risk of missed mineralization. The F1-score balances precision and recall and is particularly suitable for application-oriented mineral prospectivity studies with limited sample sizes. Overall, the ensemble

learning models developed in this study achieved good predictive performance. Among them, XGBoost and LightGBM showed particularly strong results in key metrics such as accuracy and the F1-score, indicating superior discriminative capability (Table 2). In addition, all models achieved relatively high recall values, indicating the effective identification of mineralized samples, which is of practical significance for minimizing missed targets and enhancing exploration efficiency.

Table 2. Performance comparison of different machine-learning models based on five-fold cross-validation.

Model	Accuracy	Precision	Recall	F1-Score
XGBoost	0.88	0.87	0.9	0.88
LightGBM	0.89	0.88	0.92	0.9
GBDT	0.89	0.86	0.97	0.91

Accurate delineation of high-confidence areas and the reliability of model predictions are critical for mineral prospectivity assessment in the study area; therefore, it is necessary to evaluate model performance in a scientific and effective manner. The receiver operating characteristic (ROC) curve and the area under the ROC curve (AUC) are among the most widely used performance evaluation tools in machine learning and data mining and are commonly adopted to assess the effectiveness of mineral prospectivity prediction models. A ROC curve closer to the upper-left corner indicates better classification performance. The AUC provides a threshold-independent summary metric, typically ranging from 0.5 to 1.0, where values closer to 1.0 indicate stronger discriminative ability. An AUC of 1.0 represents perfect classification, whereas an AUC of 0.5 suggests a performance equivalent to random guessing.

In this study, ROC curves were generated for mineral prospectivity predictions in the Mayoumu area, and AUC values were calculated for different models (Figure 11). The results show that all three models (GBDT, XGBoost, and LightGBM) achieved AUC values greater than 0.9, indicating high predictive accuracy and performance substantially better than random prediction. Among them, LightGBM yielded the highest AUC (0.94) and thus provided the best overall performance.

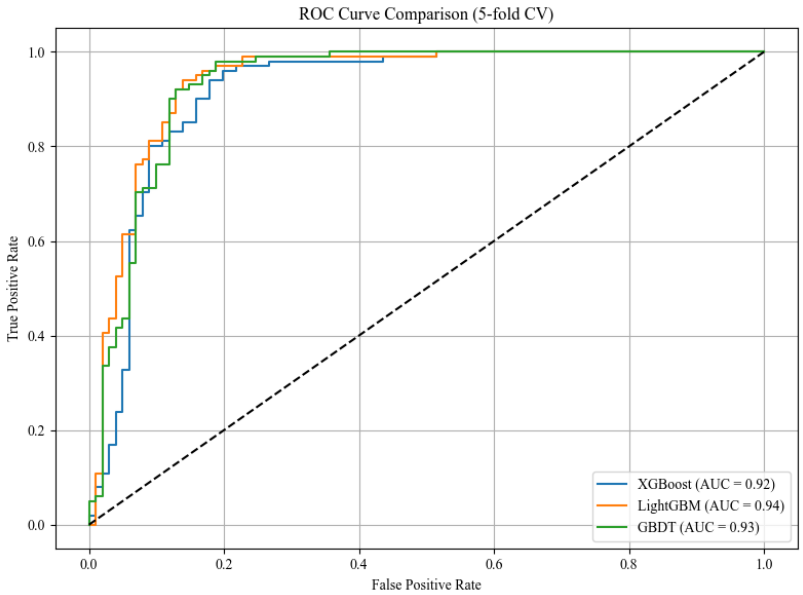


Figure 11. ROC curves and AUC values of the GBDT, XGBoost, and LightGBM models.

Overall, the three boosting-based ensemble models (GBDT, XGBoost, and LightGBM) all exhibited strong discriminative capability. The confusion matrices show consistently low false-negative (FN) counts, indicating that the models can effectively identify mineralized units and thereby reduce the risk of missing prospective targets. A comparison of key metrics further suggests that LightGBM provides a more balanced performance in terms of precision and F1-score (Precision = 0.88; F1 = 0.90) and achieves the highest AUC (0.94), reflecting more stable overall separability. Accordingly, the LightGBM model was selected for subsequent mineral prospectivity prediction.

4.2.3. Mineral Prospectivity Assessment

The mineral prospectivity prediction by the LightGBM model is a floating-point value ranging from 0 to 1, representing the likelihood of mineral deposit occurrence within each spatial unit. Higher probability values indicate a greater potential for undiscovered mineralization. In this study, the predicted mineralization probabilities for the Mayoumu area were classified into seven levels using the standard deviation method and visualized using a color scheme, where red denotes high prospectivity and blue denotes low prospectivity.

As shown in the mineral prospectivity map (Figure 12), high-probability zones are mainly concentrated in the central part of the study area. Notably, almost all known deposits/occurrences exhibit good spatial correspondence with the predicted high-prospectivity zones. In addition, several high-prospective targets have been identified outside the currently known deposits/occurrences, representing prospective areas that have not yet been systematically explored. These results indicate substantial remaining exploration potential in the Mayoumu area and further validate the reliability and effectiveness of the LightGBM-based predictions. The prospectivity map provides practical guidance and technical support for delineating exploration targets and prioritizing follow-up verification in future exploration programs.

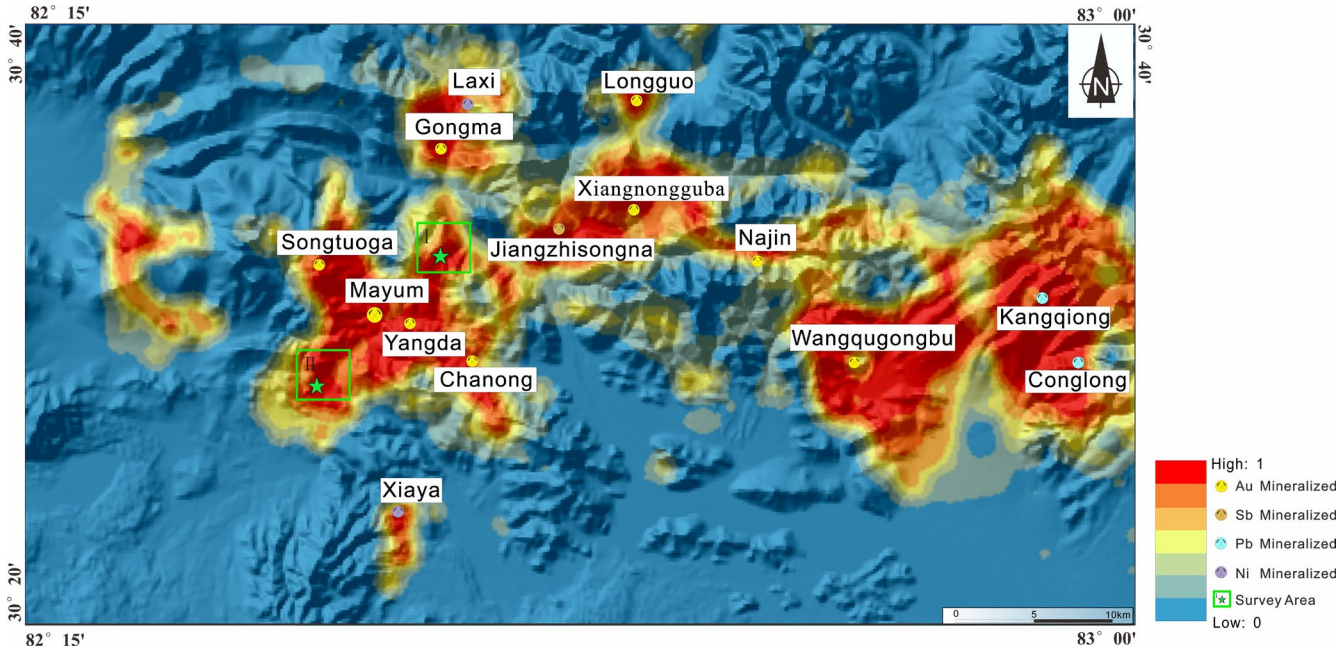


Figure 12. Mineralization potential distribution map of the Mayoumu area.

5. Discussion

Plateau–orogenic belts commonly record multi-stage tectono–magmatic and hydrothermal overprints, resulting in strongly nonlinear mineralization responses and pronounced cross-scale heterogeneity in exploration indicators. Under such conditions, single-source

prospectivity mapping often captures only a partial expression of the mineral system and may yield unstable targets at the regional scale. Accordingly, this study focuses on the unified quantification and fusion of multi-scale remote-sensing alteration signals, regional geochemical anomalies, and structural controls to better represent the multi-factor coupling inherent to orogenic mineral systems and to improve the reliability and interpretability of quantitative prospectivity prediction.

In terms of remote-sensing alteration extraction, GF-5 hyperspectral data were processed using MTMF to identify alteration minerals at the sub-pixel scale, whereas Sentinel-2 multispectral imagery was analyzed using the iCrosta method to delineate hydroxyl- and iron oxide-related anomalies. This forms a complementary workflow of "hyperspectral fine discrimination–multispectral scale extension", thereby enhancing the spatial continuity and scale adaptability of alteration signals. For geochemistry, to address the compositional nature and closure effect of regional geochemical survey data, an ILR transformation combined with RPCA was employed to derive robust anomaly components (PC1 and PC2), together with key elemental anomaly layers (Au–Sb–Pb), so as to reduce biases induced by outliers and closure. For structural controls, Euclidean distance to faults was used to convert ore-controlling structures into a continuous and computable distance variable. The resulting unified predictor system provides a more consistent input basis for ensemble-learning models to characterize nonlinear mineralization responses in complex orogenic belts.

Based on the established predictor system, GBDT, XGBoost, and LightGBM all achieved strong predictive performance (AUC > 0.90). This indicates that Boosting-based ensemble learning can automatically learn the nonlinear interactions and synergistic contributions among multi-source ore-controlling factors during training, and—when combined with SHAP—can further interpret the marginal contributions and directional effects of key predictors, thereby improving the consistency and interpretability between "statistical results" and "geological processes." Among them, LightGBM performed best (AUC = 0.94) and was consistently supported by the confusion matrix and multiple evaluation metrics (Accuracy, Precision, Recall, and F1-score). Given that mineralized occurrences represent typical "rare events", false negatives may lead to missing prospective targets, whereas false positives can substantially increase verification costs. Therefore, multi-metric evaluation facilitates a more exploration-relevant balance between "detecting more" and "detecting accurately", enabling a robust assessment of model stability and application fitness.

To mitigate the "black-box" effect of data-driven models, SHAP were further introduced to interpret the contribution and directionality of predictors. The results indicate that Sb geochemical anomalies are the most influential predictor. Meanwhile, iron oxide-related alteration information (e.g., goethite/hematite indicators) and fault proximity show stable and significant contributions across different models, suggesting a tight coupling among gold mineralization, hydrothermal alteration processes, and structurally controlled fluid pathways in the study area. To strengthen the credibility of the "statistical results–geological processes" correspondence, the predicted high-prospectivity zones were compared with the geological map. High-value zones not only show good spatial consistency with known mineralized occurrences (Au–Sb–Pb), but also tend to occur regionally in areas characterized by stronger deformation, more complex lithological contacts, and near cover boundaries. This pattern is consistent with the metallogenic understanding that orogenic hydrothermal systems are focused along faults and that metal accumulation is enhanced where favorable wall rocks overlap with alteration zones. Based on these interpretations and comparisons, target prioritization in this study emphasizes the combined criteria of "strong Sb anomalies + cross-scale superposition of alteration signals + proximity to faults (especially intersec-

tions, bends, or zones with dense secondary fractures)", thereby improving geological interpretability and operational applicability.

The prospectivity map compiled from LightGBM-derived mineralization probabilities using standard deviation-based classification shows that high-potential zones are mainly concentrated in the central part of the study area. In addition to known mineralization, several new high-prospectivity areas were delineated, providing clear guidance for subsequent verification and exploration deployment. From a broader perspective, this modular framework (multi-scale remote-sensing alteration extraction–ILR–RPCA geochemical anomaly factorization–structural distance quantification–ensemble-learning prediction–SHAP interpretation) is broadly applicable to common challenges in plateau–orogenic belts, including data heterogeneity, scale inconsistency, and nonlinear responses, and it has potential for transfer to regions with similar tectonic settings—particularly in areas with low vegetation cover and strong alteration detectability. Nevertheless, cross-region application should still consider sample representativeness and spatial autocorrelation, resolution- and scale-related effects among multi-source datasets, and the expression of predictive uncertainty. Future improvements may be achieved through spatial block cross-validation, scale-sensitivity analysis, and uncertainty quantification to further enhance model generalization and decision-support value for exploration.

To validate the model outputs from an application-oriented perspective, route-based geological surveys and representative outcrop verification were conducted in Survey Area I and Survey Area II based on the predicted results (Figure 12). Field investigations indicate that both areas are dominated by quartz sandstone exposures. Locally, the rocks are strongly fractured due to compressional and shear deformation, and quartz veins/lenses are well developed, providing favorable space for fluid migration and trapping. In the regional geological context, the study area occupies a key position within the active tectonic belt of southern Tibet, where faults are well developed and have undergone multiple stages of reactivation. Fault damage zones—especially structurally complex segments such as intersections and bends—tend to form high-permeability pathways for hydrothermal fluids. When superimposed with the thermal effects of magmatic intrusion, hydrothermal circulation and metal activation, migration, and precipitation may be further promoted, resulting in hydrothermal mineralization–alteration assemblages structured by fault–fracture systems.

In Survey Area I, distinct galena mineralization was observed in multiple quartz veins and lenses near high-prospectivity zones (Figure 13a,b). Mineralization is commonly hosted along fractures and bedding-parallel weak planes and locally forms vein- to lens-shaped concentrations, indicating strong hydrothermal infilling and replacement and favorable conditions for Pb mineralization (potentially accompanied by Au–Sb). Iron staining and silicification are commonly developed around mineralized zones, consistent with oxidative hydrothermal stages and/or supergene oxidation overprinting of sulfides, further supporting the correspondence between predicted high-value zones and actual mineralized outcrops. In Survey Area II, structurally shattered and altered rocks are widely developed, with pronounced argillization and brecciation, suggesting proximity to a fault damage zone and strong structural control. Stibiconite (Figure 13c) and massive stibnite grains (field observation) were locally identified, reflecting intense hydrothermal activity, substantial Sb input, and subsequent oxidation overprinting. In addition, epidotization, malachitization, and copper blue were observed around some quartz veins and lenses (Figure 13d), forming an alteration–mineralization assemblage of "silicification–epidotization (±chloritization)–oxidized/secondary Cu mineral enrichment", implying that, in addition to Pb–Sb mineralization, Cu-related hydrothermal–supergene modification may also be superimposed locally.

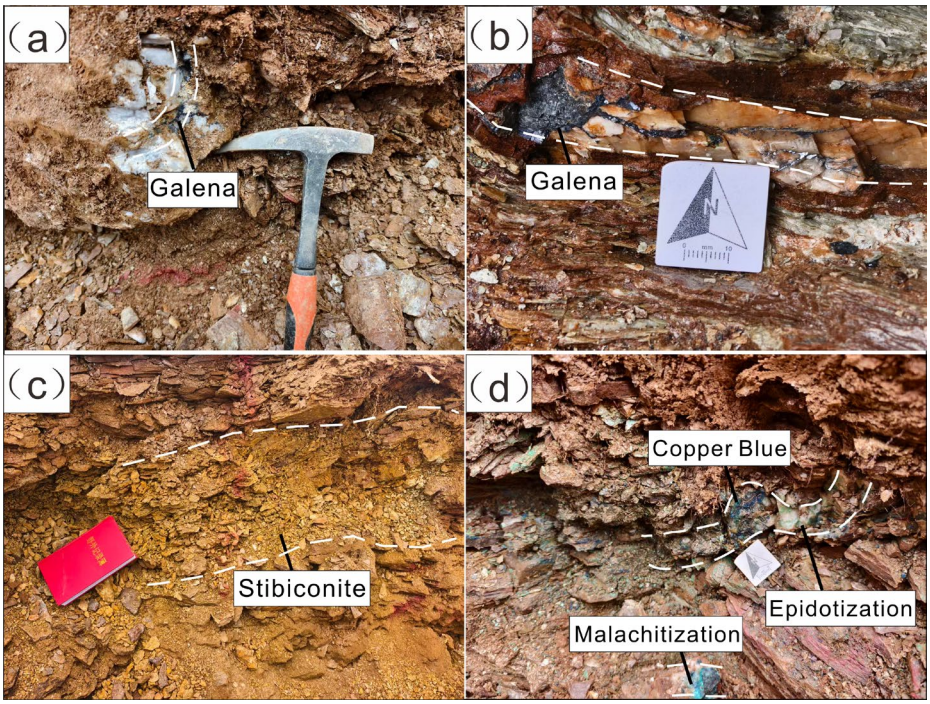


Figure 13. Field verification photographs from the survey area, showing mineralization phenomena in Survey Area I (**a**,**b**) and Survey Area II (**c**,**d**).

In summary, Survey Areas I and II exhibit favorable ore-forming geological conditions constrained jointly by well-developed fault damage and fracture systems, quartz veins/lenses providing precipitation space, metal mineral indications (galena and Sb mineralization), and alteration assemblages including epidotization, iron staining/oxidation, and secondary Cu minerals. These field observations show good spatial correspondence with the high-value zones delineated on the prospectivity map, providing direct geological and mineralization evidence supporting the prediction results and offering an important basis for zone-based verification and target prioritization in other high-prospectivity areas across the study region. Future work will conduct systematic checks of high-value anomalies along major faults and their intersections, combined with trenching and geochemical assays, to further constrain ore-controlling factors and metallogenic processes and to improve verification efficiency and exploration success rates.

6. Conclusions

In recent years, the application of remote sensing and machine learning to mineral prospectivity prediction has expanded substantially. However, systematic research is still needed on (i) the deep integration of multi-source remote-sensing, geochemical, and structural information, and (ii) the synergistic use of multi-scale remote-sensing datasets in complex orogenic belt settings. To address these issues, this study integrates GF-5 hyperspectral data (sub-pixel alteration mineral extraction based on Mixture-Tuned Matched Filtering, MTMF) with Sentinel-2 multispectral imagery (hydroxyl- and iron oxide-related anomaly extraction based on the iCrosta method), thereby achieving a complementary advantage of "hyperspectral fine discrimination–multispectral scale extension". In addition, 1:200,000 regional geochemical anomaly information and a fault-distance factor were incorporated to construct a unified predictor system that can be directly used in ensemble-learning models, providing a reproducible workflow for multi-source data fusion in complex orogenic belts. The main findings are as follows:

(1) Under a comprehensive multi-metric evaluation framework including the confusion matrix, Accuracy, Precision, Recall, F1-score, and ROC-AUC, all three models exhibited strong predictive capability (AUC > 0.90), indicating that ensemble learning integrating multi-source exploration predictors can effectively represent the nonlinear mineralization response of the study area. Among them, LightGBM achieved the best overall performance on key metrics (AUC = 0.94) and was therefore identified as the optimal model for mineral prospectivity prediction in the Mayoumu area.

(2) The prospectivity map compiled by classifying LightGBM-derived mineralization probabilities using the standard deviation method shows that high-potential zones are mainly concentrated in the central part of the study area and display good spatial consistency with known deposits/occurrences. Meanwhile, several previously untested high-prospectivity zones were delineated outside known mineralized sites, suggesting substantial remaining exploration potential. SHAP-based interpretation further indicates that Sb anomalies constitute the most influential predictor, while alteration-mineral information (e.g., goethite and epidote indicators) and fault-related predictors maintain stable and significant contributions across models, revealing a tight coupling among gold mineralization, hydrothermal alteration processes, and structurally controlled ore formation. Future work will prioritize field verification and exploration deployment in these high-potential targets to validate the predictions and improve exploration efficiency.

(3) Overall, the proposed workflow—"multi-source exploration information fusion–ensemble-learning prediction–SHAP-based interpretability analysis"—effectively addresses the prediction challenges posed by strong geological heterogeneity and nonlinear mineralization responses in complex orogenic belts, and it demonstrates good applicability and scalability. This framework not only supports target prioritization and exploration decision-making in the Mayoumu area, but also provides a transferable methodological reference for gold prospectivity mapping and exploration planning in regions with similar tectonic settings worldwide.

Author Contributions: Conceptualization, K.Q.; methodology, K.Q.; software, K.Q. and S.D.; validation, K.Q. and C.H.; formal analysis, K.Q. and T.L.; investigation, K.Q., S.G. and Z.R.; resources, T.L. and Y.H.; data curation, K.Q.; writing—original draft preparation, C.H.; writing—review and editing, K.Q. and S.D.; visualization, K.Q., T.L. and S.D.; supervision, T.L. and Y.H.; project administration, T.L.; funding acquisition, Y.H. and C.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Science and Technology Innovation Foundation of the Survey Center of Comprehensive Natural Resources (KC20250007), the China Geological Survey Projects (DD20240014, DD20243088), the Ministry of Natural Resources' New Round of Scientific and Technological Support Project for Mineral Exploration Breakthroughs (ZKKJ202427-03), and the Tibet Science and Technology Program Projects (XZ202401YD0006-07).

Data Availability Statement: The data that support the findings of this study are available from the first author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Huang, Y.; Li, G.M.; Ding, J.; Dai, J.; Yan, G.Q.; Dong, S.L.; Huang, H.X. Origin of the newly discovered Zhunuo porphyry Cu–Mo–Au deposit in the western part of the Gangdese porphyry copper belt in the southern Tibetan Plateau, SW China. *Acta Geol. Sin.-Engl. Ed.* **2017**, *91*, 109–134. [CrossRef]
2. Zhao, Q.; Yan, Y.; Tonai, S.; Dilek, Y.; Zhu, Z.F. Timing of India–Asia collision and significant coupling between them around 51 Ma: Insights from the activation history of the Zhongba–Gyangze thrust in southern Tibet. *Geology* **2024**, *52*, 61–66. [CrossRef]

3. Lin, B.; Tang, J.X.; Tang, P.; Beaudoin, G.; Laflamme, C.; Li, F.Q.; Zheng, W.B.; Song, Y.; Qi, J.; Sun, M.; et al. Multipulsed magmatism and duration of the hydrothermal system of the giant Jiama porphyry Cu system, Tibet, China. *Econ. Geol.* **2024**, *119*, 201–217. [CrossRef]
4. Huang, Y.; Ren, M.H.; Jowitt, S.M.; Li, G.M.; Fu, J.G.; Zhang, Z.; Lang, X.H.; Liu, H.; Zhang, L. Middle Triassic arc magmatism in the southern Lhasa terrane: Geochronology, petrogenesis and tectonic setting. *Lithos* **2021**, *380*, 105857. [CrossRef]
5. Huang, Y.; Wang, X.N.; Wang, Y.Y.; Yuan, H.Y.; Han, C.; Xia, L.Z.; Yang, J.F. Geological characteristics and genesis of the Xialonggang Pb–Zn–Sb deposit in South Tibet. *Acta Geol. Sin.-Engl. Ed.* **2025**, *99*, 1738–1755. [CrossRef]
6. Huang, M.L.; Bi, X.W.; Hu, R.Z.; Chiaradia, M.; Zhu, J.J.; Xu, L.L.; Yang, Z.Y. Linking porphyry Cu formation to tectonic change in post-subduction settings: A case study from the giant Yulong belt, eastern Tibet. *Econ. Geol.* **2024**, *119*, 279–304. [CrossRef]
7. Dai, J.; Qu, X.; Song, Y. Porphyry copper deposit prognosis in the middle region of the Bangonghu–Nujiang metallogenic belt, Tibet, using ASTER remote sensing data. *Resour. Geol.* **2018**, *68*, 65–82. [CrossRef]
8. Guo, N.; Guo, W.; Shi, W.; Huang, Y.; Guo, Y.; Lian, D. Characterization of illite clays associated with the Sinongduo low sulfidation epithermal deposit, Central Tibet using field SWIR spectrometry. *Ore Geol. Rev.* **2020**, *120*, 103228. [CrossRef]
9. Bai, L.Y.; Dai, J.J.; Song, Y.; Liu, Z.B.; Chen, W.; Wang, N.; Wu, C.Y. Predictive prospecting using remote sensing in a mountainous terrestrial volcanic area, western Bangongco–Nujiang mineralization belt, Tibet. *Remote Sens.* **2023**, *15*, 4851. [CrossRef]
10. Fu, Y.F.; Cheng, Q.M.; Jing, L.H.; Ye, B.; Fu, H.Z. Mineral prospectivity mapping of porphyry copper deposits based on remote sensing imagery and geochemical data in the Duolong ore district, Tibet. *Remote Sens.* **2023**, *15*, 439. [CrossRef]
11. Chen, Q.; Cai, D.Y.; Xia, J.S.; Zeng, M.; Yang, H.Y.; Zhang, R.S.; He, Y.; Zhang, X.X.; Cheng, Y.; Xu, X.H.; et al. Remote sensing identification of hydrothermal alteration minerals in the Duobuza porphyry copper mining area in Tibet using WorldView-3 and GF-5 data: The impact of spatial and spectral resolution. *Ore Geol. Rev.* **2025**, *180*, 106573. [CrossRef]
12. Qiao, K.; Luo, T.; Ding, S.H.; Quan, L.C.; Kong, J.G.; Liu, Y.W.; Huang, Y. Extraction of remote sensing alteration information based on integrated spectral mixture analysis and fractal analysis. *Minerals* **2025**, *15*, 1047. [CrossRef]
13. Zhai, W.; Sun, X.M.; Zhou, F.; Mo, R.W.; Wang, C. Helium and argon isotope systematics of orogenic gold–antimony deposits in the Himalayan orogen, southern Tibet. *Acta Geol. Sin.-Engl. Ed.* 2014, *in press*. [CrossRef] [PubMed]
14. Zhai, W.; Sun, X.M.; Yi, J.Z.; Zhang, X.G.; Mo, R.W.; Zhou, F.; Wei, H.X.; Zeng, Q.G. Geology, geochemistry, and genesis of orogenic gold–antimony mineralization in the Himalayan orogen, South Tibet, China. *Ore Geol. Rev.* **2014**, *58*, 68–90. [CrossRef]
15. Lin, C.; Zhang, J.J.; Huang, T.L.; Wang, X.X.; Zhang, B. Tectonic affinity of the Zhongba microterrane: Constraints from Silurian–Carboniferous detrital zircon geochronology in the Mayoumu area, southern Tibet. *Geol. Sci.* **2020**, *55*, 574–597. (In Chinese)
16. Liu, Q.; Liu, F.; Li, H.L.; Zong, K.Q.; Xiang, S.Y. Remnants of a Middle Triassic oceanic lithosphere in the western Yarlung Zangbo suture zone, southern Tibet. *Acta Geol. Sin.-Engl. Ed.* **2020**, *94*, 34. [CrossRef]
17. Ai, K.; Zhang, K.X.; Song, B.W.; Shen, T.Y.; Ji, J.L. Upper Oligocene–Lower Miocene Gangdese conglomerate along the Yarlung–Zangbo suture zone and its implications for Palaeo-Yarlung–Zangbo initiation. *Front. Earth Sci.* **2022**, *10*, 808843. [CrossRef]
18. Qiao, K.; Huang, Y.; Quan, L.C.; Shao, J.C.; Ai, Z.B. Multi-source remote sensing alteration information extraction and mineral prospectivity prediction in the Mayoumu area, Tibet. *Geol. Bull. China* **2025**, 1–17.
19. Kost, S.; Rheinbach, O.; Schaeben, H. Using logistic regression model selection towards interpretable machine learning in mineral prospectivity modeling. *Geochemistry* **2021**, *81*, 125826. [CrossRef]
20. Qaderi, S.; Maghsoudi, A.; Yousefi, M.; Pour, A.B. Translation of mineral system components into time step-based ore-forming events and evidence maps for mineral exploration: Intelligent mineral prospectivity mapping through adaptation of recurrent neural networks and random forest algorithm. *Ore Geol. Rev.* **2025**, *179*, 106537. [CrossRef]
21. Zheng, C.H.; Yuan, F.; Luo, X.R.; Li, X.H.; Liu, P.F.; Wen, M.L.; Chen, Z.S.; Albanese, S. Mineral prospectivity mapping based on support vector machine and random forest algorithm: A case study from the Ashele copper–zinc deposit, Xinjiang, NW China. *Ore Geol. Rev.* **2023**, *159*, 105567. [CrossRef]
22. Zerai, F.T.; Gorsevski, P.V.; Panter, K.S.; Farver, J.; Tangestani, M.H. Integration of ASTER and Soil Survey Data by Principal Components Analysis and One-Class Support Vector Machine for Mineral Prospectivity Mapping in Kerkasha, Southwestern Eritrea. *Nat. Resour. Res.* **2023**, *32*, 2463–2493. [CrossRef]
23. Lachaud, A.; Adam, M.; Mišković, I. Comparative Study of Random Forest and Support Vector Machine Algorithms in Mineral Prospectivity Mapping with Limited Training Data. *Minerals* **2023**, *13*, 1073. [CrossRef]
24. Li, Q.K.; Chen, G.X.; Luo, L. Mineral prospectivity mapping using attention-based convolutional neural network. *Ore Geol. Rev.* **2023**, *156*, 105381. [CrossRef]
25. Kong, C.F.; Zhao, J.; Li, B.Y.; Wu, C.L.; Xu, K. Manganese mineral prospectivity mapping based on semi-supervised learning and multi-source geoscientific-sample Wasserstein generative adversarial network (Geo-WGAN) in Songtao of Guizhou, South China. *Ore Geol. Rev.* **2025**, *186*, 106933. [CrossRef]
26. Lou, Y.; Liu, Y. Mineral prospectivity mapping based on a novel self-ensembling graph convolutional network. *Math. Geosci.* **2025**, *57*, 629–656. [CrossRef]

27. Xu, Y.; Zuo, R. An interpretable graph attention network for mineral prospectivity mapping. *Math. Geosci.* **2024**, *56*, 169–190. [CrossRef]
28. Ning, Y.; Wang, Y.; Lu, J.; Tian, J.T.; Wang, C.; Sheng, S.T.; Wen, S.B.; Wang, S.H.; Dong, Y.H. Mineral prospectivity mapping for multi-source geoscience data: A novel unsupervised deep learning method. *Ore Geol. Rev.* **2025**, *186*, 106866. [CrossRef]
29. Gao, L.; Gopalakrishnan, G.; Nasri, A.; Li, Y.; Zhang, Y.; Ou, X.; Xia, K. Transformer–GCN fusion framework for mineral prospectivity mapping: A geospatial deep learning approach. *Minerals* **2025**, *15*, 711. [CrossRef]
30. Koh, P.W.; Liang, P. Understanding black-box predictions via influence functions. In Proceedings of the International Conference on Machine Learning (ICML), Sydney, Australia, 6–11 August 2017; pp. 1885–1894.
31. Doshi-Velez, F.; Kim, B. Considerations for evaluation and generalization in interpretable machine learning. In *Explainable and Interpretable Models in Computer Vision and Machine Learning*; Springer: Cham, Switzerland, 2018; pp. 3–17.
32. Chudasama, B. Fuzzy inference systems for mineral prospectivity modeling—Optimized using Monte Carlo simulations. *MethodsX* **2022**, *9*, 101629. [CrossRef]
33. Skirrow, R.G.; Murr, J.; Schofield, A.; Huston, D.L.; van der Wielen, S.; Czarnota, K.; Duan, J. Mapping iron oxide Cu–Au (IOCG) mineral potential in Australia using a knowledge-driven mineral systems-based approach. *Ore Geol. Rev.* **2019**, *113*, 103011. [CrossRef]
34. Dietterich, T.G. Ensemble learning. *Handb. Brain Theory Neural Netw.* **2002**, *2*, 110–125.
35. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [CrossRef]
36. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
37. Bigdeli, A.; Maghsoudi, A.; Ghezelbash, R. A comparative study of the XGBoost ensemble learning and multilayer perceptron in mineral prospectivity modeling: A case study of the Torud–Chahshirin belt, NE Iran. *Earth Sci. Inform.* **2024**, *17*, 483–499. [CrossRef]
38. He, H.J.; Zhu, H.L.; Yang, X.K.; Zhang, W.W.; Wang, J.H. Mineral prospectivity prediction based on convolutional neural network and ensemble learning. *Sci. Rep.* **2024**, *14*, 22654. [CrossRef]
39. Yu, Z.B.; Li, B.B.; Wang, X.J. Mineral prospectivity mapping susceptibility evaluation based on interpretable ensemble learning. *Ore Geol. Rev.* **2024**, *173*, 106248. [CrossRef]
40. Chen, Y.L.; Chen, B.W.; Shayilan, A. Combining categorical boosting and Shapley additive explanations for building an interpretable ensemble classifier for identifying mineralization-related geochemical anomalies. *Ore Geol. Rev.* **2024**, *173*, 106263. [CrossRef]
41. Zhou, S.Y.; Zhang, Z.C. Machine learning-driven skarn deposit classification with SHAP-enhanced garnet geochemistry analysis. *J. Asian Earth Sci.* **2025**, *294*, 106780. [CrossRef]
42. Zhang, G.; Zhao, Z.F.; Zhang, X.L.; Wu, X.T.; Zheng, Y.F.; Feng, L.X.; Huang, Z.Q. Comprehensive multi-source remote sensing data integration for enhanced mineralization alteration extraction and geological structure interpretation in the Lala region of Sichuan Province. *Ore Geol. Rev.* **2024**, *168*, 106032. [CrossRef]
43. Chen, Q.; Zhao, Z.F.; Zhou, J.X.; Zhu, R.F.; Xia, J.S.; Sun, T.; Zhao, X.; Chao, J.Q. ASTER and GF-5 satellite data for mapping hydrothermal alteration minerals in the Longtoushan Pb–Zn deposit, SW China. *Remote Sens.* **2022**, *14*, 1253. [CrossRef]
44. Qin, Y.H.; Zhang, X.L.; Zhao, Z.F.; Li, Z.Y.; Yang, C.B.; Huang, Q.Y. Coupling relationship analysis of gold content using Gaofen-5 (GF-5) satellite hyperspectral remote sensing data: A potential method in Chahuazhai gold mining area, Qiubei County, SW China. *Remote Sens.* **2022**, *14*, 109. [CrossRef]
45. Son, J.; Jung, I.; Park, K.; Han, B. Tracking-by-segmentation with online gradient boosting decision tree. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 11–18 December 2015; pp. 3056–3064.
46. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
47. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16), San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
48. Ke, G.L.; Meng, Q.; Finley, T.; Wang, T.F.; Chen, W.; Ma, W.D.; Ye, Q.W.; Liu, T.Y. LightGBM: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*; NeurIPS Foundation: La Jolla, CA, USA, 2017; p. 30.
49. Shapley, L.S.; Snow, R.N. Basic solutions of discrete games. In *Contributions to the Theory of Games*; Princeton University Press: Princeton, NJ, USA, 1952; Volume 1, pp. 27–35.
50. Štrumbelj, E.; Kononenko, I. Explaining prediction models and individual predictions with feature contributions. *Knowl. Inf. Syst.* **2014**, *41*, 647–665. [CrossRef]
51. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*; NeurIPS Foundation: La Jolla, CA, USA, 2017; p. 30.
52. Abdelkader, M.A.; Watanabe, Y.; Shebl, A.; Badawi, M.; Dawoud, M.; El-Dokouny, H.A.; Csámer, Á.; Abdellatif, M. Advanced exploration of rare metal mineralization through integrated remote sensing and geophysical analysis of structurally-controlled hydrothermal alterations. *J. Geochem. Explor.* **2024**, *267*, 107598. [CrossRef]

53. Carvalho, M.; Cardoso-Fernandes, J.; González, F.J.; Teodoro, A.C. Comparative performance of Sentinel-2 and Landsat-9 data for raw materials' exploration onshore and in coastal areas. *Remote Sens.* **2025**, *17*, 305. [CrossRef]
54. Yang, C.; Jia, H.K.; Dong, L.F.; Zhao, H.S.; Zhao, M.B. Selection of Landsat-8 operational land imager (OLI) optimal band combinations for mapping alteration zones. *Remote Sens.* **2024**, *16*, 392. [CrossRef]
55. Kokaly, R.F.; Clark, R.N.; Swayze, G.A.; Livo, K.E.; Hoefen, T.M.; Pearson, N.C.; Wise, R.A.; Benzel, W.; Lowers, H.A.; Driscoll, R.L.; et al. *USGS Spectral Library Version 7*; US Geological Survey; USGS: Reston, VA, USA, 2017; pp. 1–60.
56. Yuan, C.J.; Zhao, J.; Liu, Z.B.; Liu, X.F.; Tang, J. Detecting mineralization via white mica spectral footprints: From field-based sampling to satellite hyperspectral remote sensing. *Ore Geol. Rev.* **2025**, *181*, 106588. [CrossRef]
57. Quintano, C.; Fernández-Manso, A.; Shimabukuro, Y.E.; Pereira, G. Spectral unmixing. *Int. J. Remote Sens.* **2012**, *33*, 5307–5340. [CrossRef]
58. Zuo, R.G.; Xia, Q.L.; Wang, H.C. Compositional data analysis in the study of integrated geochemical anomalies associated with mineralization. *Appl. Geochem.* **2013**, *28*, 202–211. [CrossRef]
59. Zhao, Z.H.; Zhao, X.; Yin, Y.C.; Li, C.L.; Yang, Y.J.; Wang, Y.P. Identification of geochemical anomalies based on RPCA and multifractal theory: A case study of the Sidaowanzi area, Chifeng, Inner Mongolia. *ACS Omega* **2024**, *9*, 24998–25013. [CrossRef]
60. Zhao, Z.H.; Qiao, K.; Liu, Y.W.; Chen, J.; Li, C.L. Geochemical data mining by integrated multivariate component data analysis: The Heilongjiang Duobaoshan area (China) case study. *Minerals* **2022**, *12*, 1035. [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.