

Article

Integrating Multi-Source Geoscientific Data via Geologically Constrained Feature Engineering for Gold Prospectivity Mapping: A Case Study of Jiaoxibei, China

Yajie Feng ¹, Yongzhi Wang ^{1,2,*} , Yigao Cheng ³, Jiahui Zheng ¹, Shaohui Wang ¹, Zhaofeng An ¹ and Zheng Ji ¹

¹ College of Geoexploration Science and Technology, Jilin University, Changchun 130061, China; yjfeng24@mails.jlu.edu.cn (Y.F.); zhengjh456@163.com (J.Z.); shaohui24@mails.jlu.edu.cn (S.W.); azf24@mails.jlu.edu.cn (Z.A.); jizheng@jlu.edu.cn (Z.J.)

² Institute of Integrated Information for Mineral Resources Prediction, Jilin University, Changchun 130061, China

³ Institute of Marine Geology and Resources, Ocean College, Zhejiang University, Zhoushan 316021, China; yigaocheng@zju.edu.cn

* Correspondence: wangyongzhi@jlu.edu.cn

Highlights

What are the main findings?

- Geological feature engineering reduced 32 factors to nine core ore controls.
- The model captured 84.4% of known gold occurrences within 25.356% of the area.

What are the implications of the main findings?

- Feature screening improved the capture–area balance relative to the all-feature fuzzy model.
- The framework supports efficient and interpretable targeting in complex districts.

Abstract

The Jiaoxibei gold cluster is one of the most significant gold-producing regions in China and retains substantial regional prospecting potential. However, the superposition of multiple mineralization events has resulted in strong spatial coupling, multi-scale variability, and substantial redundancy among structural, alteration, geophysical, and geochemical information, limiting the effective extraction of key ore-controlling features. This study developed a geologically constrained feature-engineering framework for regional mineral potential evaluation. An initial indicator system comprising 32 geologically meaningful factors was constructed from structural geometry, remote-sensing alteration, gravity, magnetic, and geochemical information. The previously developed SOMML method was extended by introducing borehole-derived geological constraints to construct G-SOMML and generate the comprehensive geochemical anomaly factor Chem_F. Correlation-based redundancy reduction, Random Forest importance evaluation, SHAP interpretation, and geological screening were then combined to identify nine core factors. The selected factors were subsequently transformed according to their mineralization-response directions and integrated through category-balanced fuzzy synthetic evaluation. By linking geology-guided feature construction, borehole-constrained geochemical anomaly extraction, data-driven feature diagnosis, and balanced evidence integration, the framework provides a reproducible and interpretable workflow for organizing heterogeneous geoscientific information in regional mineral potential evaluation. The resulting high-potential zones captured 21 of the 22 known gold occurrences in the Sanshandao and Jiaojia areas and 17 of the 23 occurrences in the other parts of the study area, yielding an overall deposit capture rate of 84.4%. Compared with the all-feature fuzzy synthetic evaluation model (AT-Fuzzy), feature engineering



Academic Editors: Hengqian Zhao, Zhifang Zhao, Joana Cardoso-Fernandes and Deshuai Yuan

Received: 24 June 2026

Revised: 31 July 2026

Accepted: 3 August 2026

Published: 5 August 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

reduced the high-potential area ratio from 50.951% to 25.356% while increasing the deposit capture rate from 51.1% to 84.4%. At the map level, the proposed framework delineated a substantially smaller high-potential area than the Random Forest model while achieving higher deposit capture rates than both the Random Forest and Weights of Evidence models under the common evaluation domain and statistical thresholding criterion. These results support the applicability of the framework for interpretable regional mineral potential evaluation and target prioritization in complex metallogenic districts.

Keywords: feature engineering; fuzzy logic; potential evaluation; gold deposits; Jiaoxibei

1. Introduction

The Jiaodong metallogenic belt is located in the eastern part of the North China Craton, with the Jiaoxibei area situated in its northern segment, representing one of the most significant gold concentration areas in China [1,2]. A series of gold deposits controlled by large-scale fault zones have developed within the region, forming a typical broken-belt altered-rock type gold deposit system represented by the Jiaojia-style [3]. The region possesses enormous proven gold resources and hosts several large to giant gold deposits, demonstrating prominent resource advantages and prospecting potential. As a vital component of the Jiaodong metallogenic belt, the Jiaoxibei area serves as a classic representative for studying regional metallogenesis, deep fluid evolution, and structural ore-controlling mechanisms, and has long been a key area for research on gold mineralization regularities and prediction [4–6]. With the continuous deepening of domestic and international research, the degree of exploration for shallow resources in this area has become relatively high, but the difficulty of exploring deep and concealed orebodies has significantly increased [7,8]. Simultaneously, the superposition of multiple mineralization events has led to a high degree of spatial coupling among multi-source information such as structures, alterations, and geophysics, exhibiting significant redundancy and non-linear characteristics across different scales and physical properties [9]. Consequently, traditional methods face limitations in the effective representation of multi-source information, extraction of key ore-controlling factors, and control of information redundancy, which imposes higher requirements for the refined representation of mineralization information and feature construction methods.

Numerous studies have evaluated gold mineral potential in the Jiaoxibei area. Traditional prospecting methods, primarily based on geological experience and statistical analysis, perform mineralization prediction through the summarization of structural ore-controlling laws, superposition of favorable metallogenic factors, and weights of evidence [10–12]. While these methods can effectively characterize major ore-controlling factors at a regional scale, they typically rely on manual experience for factor weighting and struggle to quantitatively represent the complex relationships between multi-source information [13]. With the improvement of spatial data acquisition capabilities, data-driven machine learning methods (e.g., Random Forest, Support Vector Machines) have been gradually introduced into mineral potential evaluation research. These methods demonstrate certain advantages in handling non-linear relationships and multi-dimensional data fusion, achieving good predictive results in various study areas [14–17]. Recent studies have increasingly adopted explainable and ensemble machine-learning approaches to improve model transparency and predictive robustness [15,18,19]. The workflow developed in this study follows this trend by combining Random Forest feature importance and SHAP interpretation with geological screening, while retaining knowledge-driven fuzzy integration for regional mineral potential evaluation. However, existing studies mostly focus on model

construction and algorithm optimization and pay insufficient attention to the systematic construction of input features. On one hand, the direct use of raw or simply processed geoscientific data as model inputs leads to significant redundancy and scale-mixing issues among features [20]; on the other hand, in complex metallogenic systems, multi-source information is highly coupled in space. Collinearity and noise interference among features further weaken the discriminative power of the models and reduce the geological interpretability of the results [18,19]. Especially under conditions of uneven sample distribution or limited labeled data, relying solely on data-driven methods makes it difficult to stably characterize regional mineralization laws [21], indicating that current research still has clear deficiencies in the effective representation of mineralization information and the extraction of key features.

In response to these limitations, mineral potential evaluation tasks urgently require the effective representation of multi-source information within complex geological systems [13]. The gold mineralization process in the Jiaoxibei area involves various geological processes such as structural evolution, hydrothermal activity, and deep physical field responses. The associated information exhibits multi-scale superposition and strong coupling characteristics in space, where different data sources provide complementarity but are also accompanied by significant redundancy and noise interference [9,22–24]. In this context, relying solely on original variables or simple statistical transformations makes it difficult to fully characterize the essential attributes of ore-controlling factors, thereby limiting the effective identification of mineralization laws [25,26]. Feature engineering, as a crucial link connecting raw data and predictive models, provides a new pathway for the reconstruction and optimized expression of multi-source geoscientific information [27,28]. By introducing processes such as feature construction with explicit geological significance, redundant information elimination, and multi-scale information decoupling, feature engineering can preserve key mineralization information while reducing noise interference, thus enhancing the model's response to ore-controlling factors and the geological interpretability of the results [29,30]. However, the application of feature engineering in existing research mostly remains at the level of general data processing, lacking systematic design tailored to the complexity of metallogenic systems, particularly in areas such as multi-source information fusion, scale coordination, and the introduction of geological constraints. Accordingly, the specific research gap is the lack of a unified workflow that constructs geologically meaningful multi-source indicators, coordinates information across scales, removes redundant features, and retains geological constraints throughout feature screening and evidence integration.

To address the insufficient representation of mineralization information caused by the strong coupling, high redundancy, and multi-scale superposition of multi-source geoscientific information, this study developed a geologically constrained feature-engineering framework for regional mineral potential evaluation in the Jiaoxibei gold cluster (Figure 1). Specifically, we (1) reconstructed structural geometry, remote-sensing alteration, gravity, magnetic, and geochemical information to establish an initial mineralization indicator system with explicit geological meanings; (2) extended the SOMML method by introducing borehole-derived geological constraints to construct G-SOMML for comprehensive geochemical anomaly extraction [31]; (3) combined correlation-based redundancy reduction, Random Forest importance evaluation, SHAP interpretation, and geological screening to identify the core ore-controlling factors; and (4) transformed the selected factors according to their mineralization-response directions and integrated them through category-balanced fuzzy synthetic evaluation [32]. We evaluated the proposed framework based on deposit capture rates, prospective-area ratios, Prediction–Area curves, success-rate curves and corresponding SR-AUC values, together with comparisons against the all-feature fuzzy syn-

thetic evaluation (AT-Fuzzy), Weights of Evidence (WoE), and Random Forest (RF) models. Unlike conventional workflows that rely primarily on either expert-guided evidence integration or direct data-driven prediction, the proposed framework coordinates geological constraints and data-driven feature analysis throughout the construction, screening, interpretation, and balanced integration of the multi-source feature space, thereby improving the spatial focusing and geological interpretability of regional mineral potential evaluation.

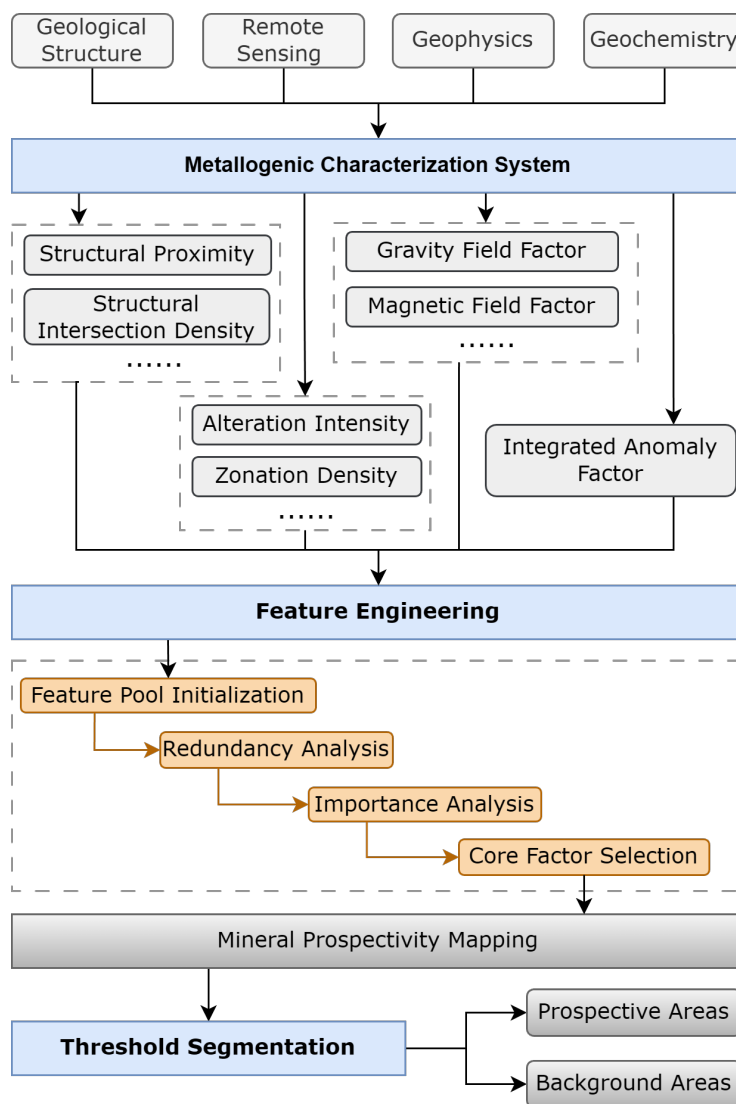


Figure 1. Flowchart of mineral potential evaluation based on feature engineering.

2. Geological Setting

The Jiaoxibei area is situated on the southeastern margin of the North China Craton (Figure 2). Since the Mesozoic, influenced by the subduction of the Paleo-Pacific Plate and the dextral shearing of the Tan-Lu fault zone, a structural framework centered on NE–NNE-trending faults has been formed [33,34]. Three main trunk fault zones—Sanshandao, Jiaojia, and Zhaoping—are distributed within the study area, arranged in a nearly parallel fashion with SE dips and dip angles ranging from 30° to 70° [35,36]. These major NE–NNE-trending faults are interpreted as pre-existing structures that were reactivated during gold mineralization and served as the principal ore-controlling structures, whereas post-mineralization fault activity locally modified the established mineralized fault zones [37,38]. These faults, along with their derived secondary fractures, bends, and intersections, serve as the primary spaces for the migration and accumulation of metallogenic fluids [39].

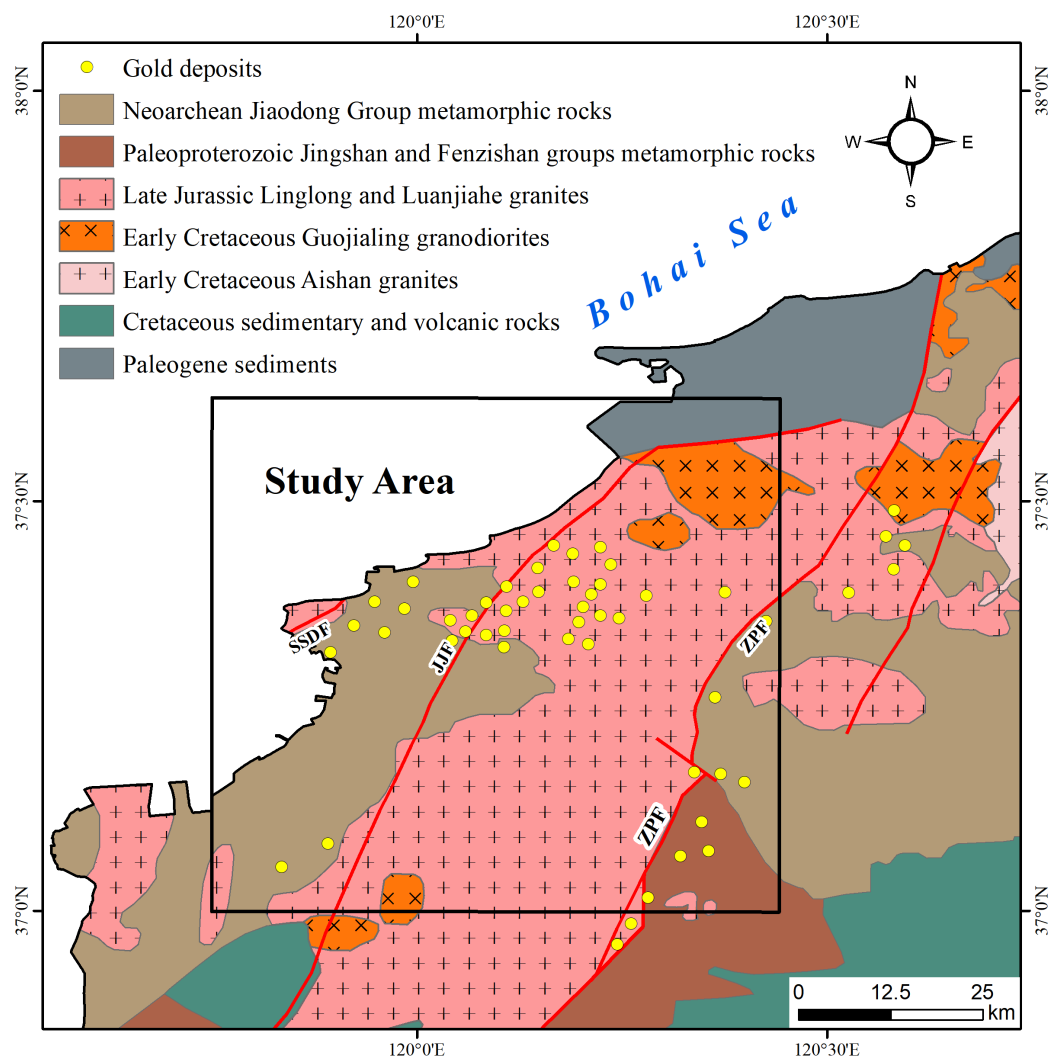


Figure 2. Geological map of the study area. Red lines indicate regional faults; SSDF: Sanshandao Fault; JIF: Jiaojia Fault; ZPF: Zhaoping Fault. Yellow circles indicate known gold deposits.

The stratigraphic units exposed in the area span from the Neoproterozoic to the Cenozoic [40]. Among them, the oldest basement is the Neoproterozoic Jiaodong Group metamorphic complex, primarily composed of amphibolite and leptyne. Since the Mesozoic, the Jiaoxibe area has undergone multiple stages of magmatic intrusion, mainly developing Late Jurassic Linglong-type granites (e.g., monzogranite) and Early Cretaceous Guojialing-type granites (e.g., K-feldspar porphyritic granodiorite) [1]. These two types of granitic plutons are the most important host rocks for gold deposits in the region.

The regional metallogenic model can be summarized as structurally controlled hydrothermal gold mineralization hosted mainly by the Linglong- and Guojialing-type granitoids. The dominant mineralization styles are altered-rock type in fracture zones and quartz-vein type [41]. Hydrothermal fluids migrated primarily through the reactivated major faults and associated secondary structures, and fluid–rock interaction produced distinct alteration zoning [2]. Within the fractured alteration zones in the footwall of the main fault planes, pyrite–sericite–quartz alteration zones, sericitized cataclasite zones, and potassic–sericitized granite zones are developed sequentially from the interior to the exterior [42]. Metallic minerals in the ores are dominated by pyrite, accompanied by chalcopyrite, galena, and sphalerite.

At the regional scale, the spatial distribution of gold mineralization reflects the combined influence of fault architecture, hydrothermal alteration, intrusive activity, and geo-

chemical dispersion. Major faults and their associated secondary structures control fluid migration and ore emplacement, whereas alteration zoning records the intensity and spatial extent of fluid–rock interaction. Gravity and magnetic variations provide indirect information on intrusive contacts, concealed structures, and subsurface physical-property contrasts, while multi-element geochemical anomalies characterize the geochemical response associated with mineralization. These complementary geological expressions provide the basis for constructing the structural, alteration, geophysical, and geochemical features used in the subsequent analysis.

3. Mineralization Indicator System

Following the translation logic of the mineral systems approach [43], the ore-forming processes and geological expressions summarized in Section 2 were converted into measurable structural, alteration, geophysical, and geochemical predictor layers.

3.1. Quantification of Structural Geometric Features

Structures are the primary controlling factors for the migration and accumulation of metallogenic fluids in the Jiaoxibei gold deposits. Accordingly, the structural geometric indicators constructed in this study characterize the mapped regional fault network, whose principal NE–NNE-trending components correspond to the major pre-existing faults reactivated during mineralization and their associated secondary structures [37,38]. To characterize the ore-controlling role of structures from multiple dimensions—including spatial proximity, structural complexity, and connectivity—this study designed and extracted four structural geometric feature indicators. We quantified these features using structural data vectorized from 1:50,000 geological maps of the Jiaoxibei area. We spatially calibrated the vectorized structures with reference to geomorphological linear features and image textures from high-resolution remote sensing imagery and uniformly projected them into the WGS 1984 coordinate system. We calculated the relevant indicators using ArcGIS 10.5 software and exported the results in TIFF format. The spatial distribution of the extracted structural geometric features is shown in Figure 3. The methodological parameters were defined a priori according to the geological meaning of each indicator, the 1:50,000 source-map scale, and the characteristic spatial dimensions of the mapped fault network; they were not optimized against the locations of known gold deposits.

3.1.1. Structural Proximity

Structural proximity reflects the spatial constraint range of structures on metallogenic fluid migration by quantifying the distance relationship between spatial units and faults (Figure 3a). This study utilized the Euclidean distance algorithm to construct a continuous structural proximity field covering the entire area. Based on the vectorized fault data, a parameterized expression of spatial constraint intensity was achieved by calculating the straight-line distance from the center point of each pixel in the grid to the nearest fault. The calculation formula is given in Equation (1).

$$d_{\text{fault}} = \sqrt{(x_i - x_f)^2 + (y_i - y_f)^2} \quad (1)$$

where (x_i, y_i) are the spatial coordinates of the calculation point, and (x_f, y_f) are the coordinates of the nearest fault point.

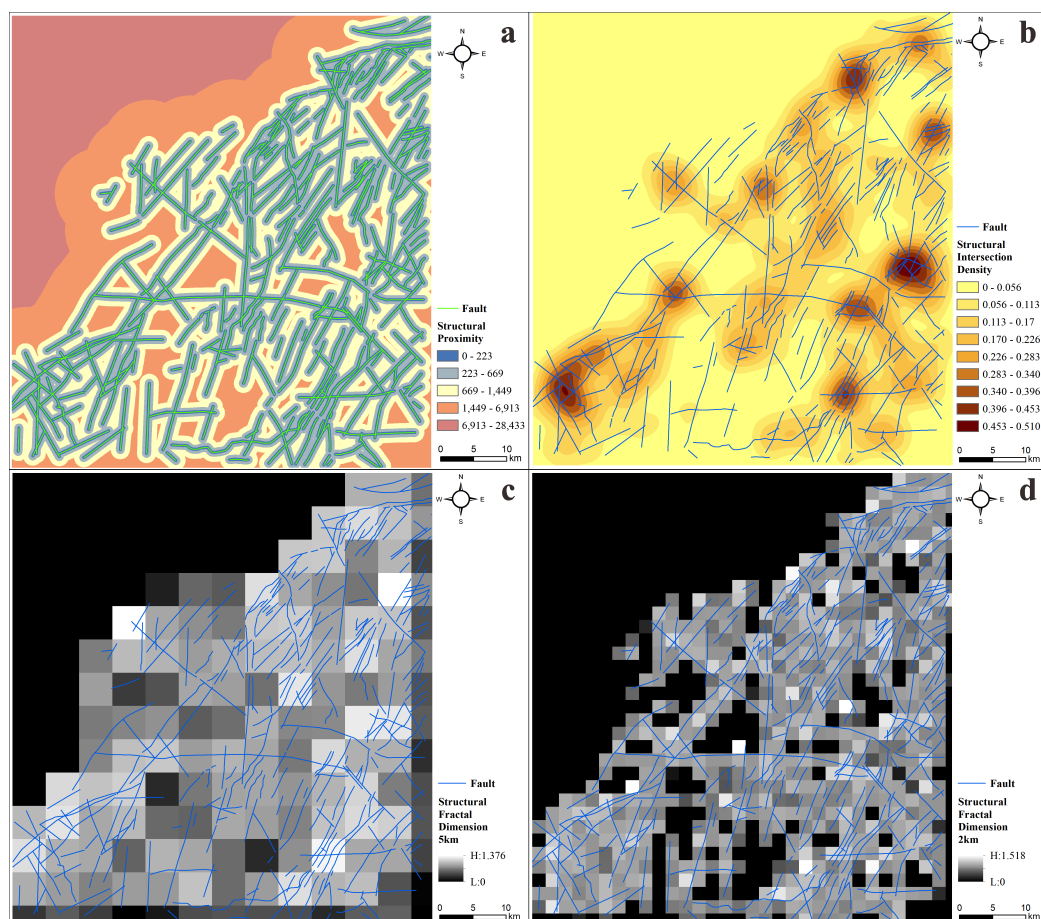


Figure 3. Spatial distribution maps of structural geometric features: (a) structural proximity; (b) structural intersection density; (c) structural fractal dimension calculated using a 5 km spatial window; and (d) structural fractal dimension calculated using a 2 km spatial window.

3.1.2. Structural Intersection Density

The intersection sites of faults are typically sensitive zones for stress concentration and abrupt changes in fluid permeability. This study adopts a method combining multi-level tolerance weighting with Kernel Density Estimation (KDE) to quantitatively characterize the connectivity of the fault system (Figure 3b). To account for both exact intersections and near-intersections among discontinuous or closely spaced fault segments, three nested extraction tolerances of 0 m, 300 m, and 500 m were defined a priori, with corresponding weights of 1, 0.7, and 0.5. The weighting strategy accounts for both the core contribution of high-precision intersections and the relevance of nearby structural nodes. Finally, the kernel density algorithm was used to calculate the spatial distribution frequency of weighted intersections within the regional grid, using a search radius of 5000 m. The kernel density results reflect the development intensity of structural nodes per unit area, effectively identifying the convergence locations of metallogenic fluids.

3.1.3. Multi-Scale Structural Fractal Dimension

The structural fractal dimension quantitatively describes the degree to which a fault network occupies space and its geometric irregularity. To simultaneously characterize the heterogeneity of the regional metallogenic background and local structural states, this study employed the Box-counting Method to perform calculations at spatial window scales

of 5 km (Figure 3c) and 2 km (Figure 3d). The formula for the box-counting dimension is given in Equation (2).

$$D = \lim_{\epsilon \rightarrow 0} \frac{\log N(\epsilon)}{\log(1/\epsilon)} \quad (2)$$

where ϵ is the side length of the covering boxes, and $N(\epsilon)$ is the total number of boxes required to cover the fault system.

The 5 km window corresponds to the spatial distribution scale of the main fault zones in the study area, comprehensively reflecting the superposition of multiple fault systems and the degree of regional structural fragmentation. The 2 km window is used to characterize the local complexity differences within the regional structural background, reflecting the concentration and nodal characteristics of faults within smaller spatial units. To ensure the stability of the regression calculation, the maximum value of the box scale set for local scales was restricted to within one-fourth of the window scale.

3.2. Surface Alteration Information Extraction

Surface hydrothermal alteration zones provide observable spectral expressions of hydrothermal fluid–rock interaction. This study extracted information for three core alteration indicators: sericitization, chloritization, and iron staining anomalies. The extracted indicators include point alteration intensity, alteration influence intensity, and alteration zoning density. Hyperspectral data from the AHSI sensor of the Ziyuan-1 02D (ZY-1 02D) satellite were utilized, featuring a spatial resolution of 30 m, a maximum spectral resolution of 10 nm, and a spectral range from 400 nm to 2500 nm across 166 bands. The preprocessing workflow included radiometric calibration, atmospheric correction, orthorectification, and mask clipping. To minimize interference from sea areas, buildings, and vegetation on spectral features, high-precision masks for vegetation, buildings, and shadows were constructed using Ziyuan-3 (ZY-3) optical imagery with 2 m spatial resolution through band math classification. These masks were resampled to 30 m to match the hyperspectral grid and reduce residual interference from masked surface features. Relevant indicator calculations were completed in ENVI 5.6.2 and ArcGIS software, uniformly clipped using a land-area mask, and output in TIFF format. The resulting alteration maps represent surface spectral expressions rather than direct observations of subsurface mineralization. Residual vegetation and mixed pixels may weaken diagnostic absorption features, anthropogenic surfaces may produce locally similar spectral responses, and weathering or oxidation may enhance iron-staining signals or obscure primary alteration minerals. The masking procedure reduces but cannot completely eliminate these effects; therefore, the alteration layers were treated as candidate surface indicators and evaluated jointly with the other evidence during subsequent feature screening and integration. The 30 m hyperspectral products provide finer raster detail, whereas the 1:50,000 structural maps and 1:200,000 geophysical and geochemical datasets represent information at broader effective spatial scales. Each dataset was processed according to its source scale and geological meaning before the derived layers were co-registered for comparative analysis. Grid harmonization ensured spatial alignment but did not increase the effective resolution of the coarser datasets.

3.2.1. Point Alteration Intensity

Point alteration intensity reflects the relative strength of hydrothermal activity by quantifying the absorption characteristics of various minerals in specific bands. We calculated the point alteration intensity (Figure 4) using index operations to extract the original abundance values of sericite, chlorite, and iron staining anomalies. For sericitization, we constructed a ratio index using its absorption pit at 2200 nm combined with flanking bands. For chlorite, considering noise interference at the 2400 nm end, we used the left-shoulder

band for ratio extraction. We extracted iron staining anomalies using the ratio of the red band to the blue band. We applied a threshold segmentation method to the mineral abundance maps to obtain the alteration intensity maps for each mineral. The calculation formulas are given in Equations (3)–(5).

For sericite (Equation (3)):

$$I_{\text{Sericite}} = \frac{R_{2000} + R_{2260}}{2 \times R_{2200}} \quad (3)$$

where R_{2000} is Band 137 (2014.404053 nm), R_{2200} is Band 148 (2199.482910 nm), and R_{2260} is Band 152 (2266.836426 nm).

For chlorite (Equation (4)):

$$I_{\text{Chlorite}} = \frac{R_{2250}}{R_{2330}} \quad (4)$$

where R_{2250} is Band 151 (2249.998047 nm) and R_{2330} is Band 156 (2334.189697 nm).

For iron staining (Equation (5)):

$$I_{\text{Iron}} = \frac{R_{660}}{R_{480}} \quad (5)$$

where R_{480} is Band 11 (481.829498 nm) and R_{660} is Band 32 (662.286926 nm).

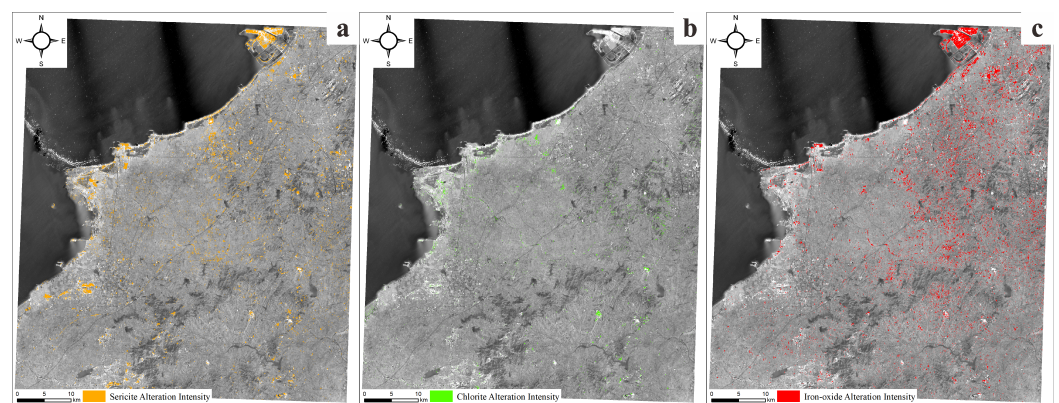


Figure 4. Point alteration intensity maps. (a) Sericite, (b) chlorite, (c) iron staining anomaly.

3.2.2. Alteration Influence Intensity

The alteration influence intensity indicator (Figure 5) was constructed using the Kernel Density Estimation (KDE) algorithm to quantify the spatial decay law of mineral potential with increasing distance from hydrothermal centers. First, anomalous patches were converted into point features, and kernel density was calculated with a search radius of 1200 m. By smoothing discrete anomaly signals, this indicator generates a continuous field reflecting the overall spatial influence range of the hydrothermal system. It effectively indicates the spatial proximity of a spatial unit to the hydrothermal core area or fluid channel leakage points.

3.2.3. Alteration Zoning Density

The alteration zoning density indicator (Figure 6) was used to characterize the spatial aggregation of alteration anomalies, indicating the frequency and scale of fluid activity. This study employed a circular sliding window (Focal Statistics) for spatial neighborhood statistics, selecting a local search radius of 600 m. By calculating the mean occurrence of anomalous pixels within the window, a quantitative description of alteration development frequency was achieved. Compared to the alteration influence intensity indicator,

alteration zoning density better captures the centripetal zoning characteristics of the hydrothermal system.

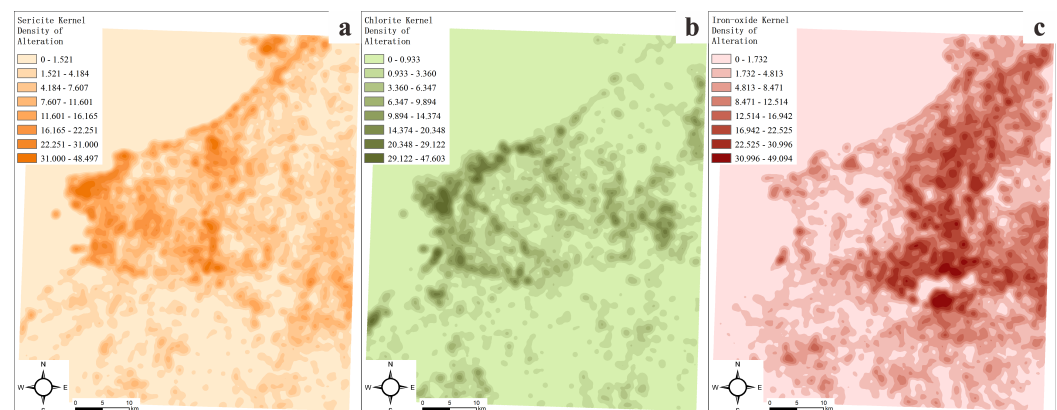


Figure 5. Alteration influence intensity maps. (a) Sericite, (b) chlorite, (c) iron staining anomaly.

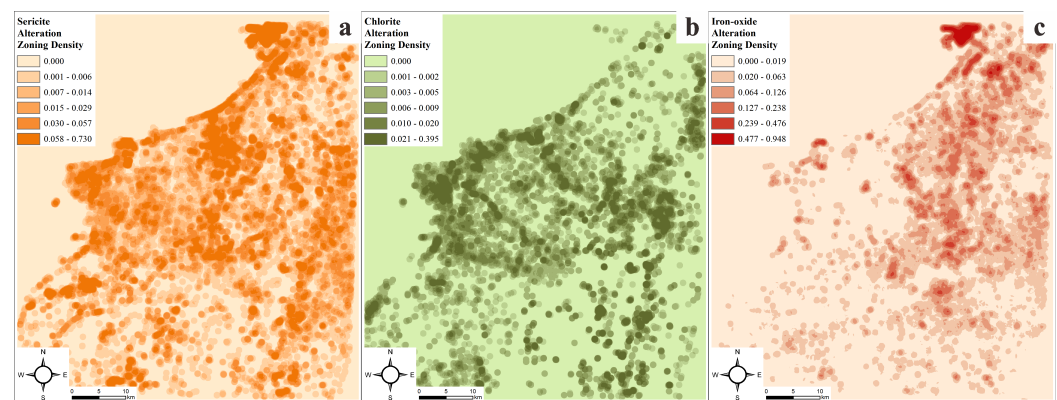


Figure 6. Alteration zoning density maps. (a) Sericite, (b) chlorite, (c) iron staining anomaly.

3.3. Deep Background Feature Analysis of Physical Fields

Potential-field background feature analysis used quality-checked 1:200,000 residual gravity anomaly and magnetic anomaly grids as inputs for subsequent potential-field transformations. This study constructed a feature system reflecting the deep ore-controlling environment through scale separation, gradient enhancement, and composite attribute extraction. All calculation processes were completed on the Geosoft Oasis montaj platform, employing frequency-domain filtering (FFT) and spatial operator transformations [44,45]. The indicator results are continuous grid data covering the entire domain, output in .grd format.

3.3.1. Gravity Field Features

Gravity features are primarily used to characterize density inhomogeneities and the deep structural framework of the study area. In this study, Fourier transform techniques were applied to the input residual gravity anomaly data for scale separation and derivative operations. A total of nine gravity indicators were constructed: the original gravity field (Figure 7a), 6 km upward continued gravity field (Figure 7b), vertical first-order derivative of the 6 km continued field (Figure 7c), 3 km upward continued gravity field (Figure 7d), 3 km residual gravity field (Figure 7e), vertical first-order derivative of the residual field (Figure 7f), X-direction first-order derivative of the residual field (Figure 7g), Y-direction first-order derivative of the residual field (Figure 7h), and horizontal gradient magnitude of the residual field (HGM_g) (Figure 7i). Upward continuation and residual-field separation were used to distinguish regional and local gravity-field components, while the derivative

and horizontal-gradient indicators enhanced density-contrast boundaries and lateral variations in the gravity field. The calculation formulas for the upward continued field and horizontal gradient magnitude are as follows (Equations (6) and (7)):

$$G_{\text{up}}(z) = \mathcal{F}^{-1}\{\mathcal{F}(G) \cdot e^{-kz}\} \quad (6)$$

where $\mathcal{F}(\cdot)$ is the forward Fourier transform, $\mathcal{F}^{-1}(\cdot)$ is the inverse Fourier transform, G is the input residual gravity anomaly value, and k is the wavenumber.

$$HGM_g = \sqrt{\left(\frac{\partial\phi}{\partial x}\right)^2 + \left(\frac{\partial\phi}{\partial y}\right)^2} \quad (7)$$

where ϕ is the residual gravity field, and $\partial\phi/\partial x$ and $\partial\phi/\partial y$ represent the rate of change of the gravity field in the horizontal x and y directions, respectively.

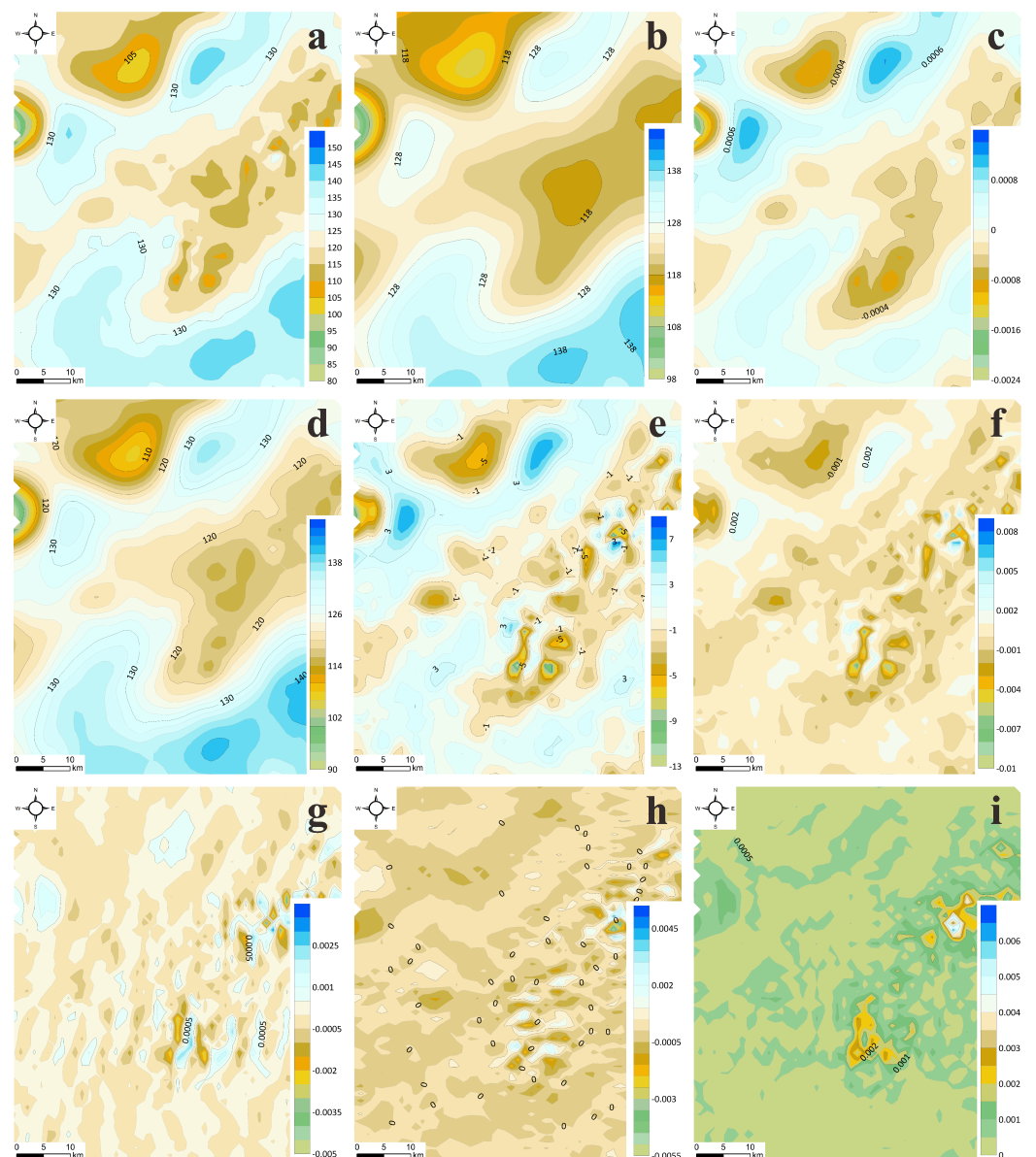


Figure 7. Gravity feature maps. (a) Original gravity field, (b) 6 km upward continued gravity field, (c) vertical first-order derivative of 6 km continued field, (d) 3 km upward continued gravity field, (e) 3 km residual gravity field, (f) vertical first-order derivative of residual field, (g) X-direction first-order derivative of residual field, (h) Y-direction first-order derivative of residual field, (i) horizontal gradient magnitude.

3.3.2. Magnetic Field Features

Magnetic features indicate the distribution of deep rock masses and concealed structures by reflecting variations in rock magnetism. In this study, reduction to the pole (RTP) was applied to the input magnetic anomaly data to reduce the influence of magnetization direction. Including the RTP magnetic anomaly, nine indicators were constructed: the RTP magnetic anomaly (Figure 8a), 6 km upward continued magnetic field (Figure 8b), 6 km residual magnetic anomaly (Figure 8c), X-direction first-order derivative of the residual field (Figure 8d), Y-direction first-order derivative of the residual field (Figure 8e), vertical first-order derivative of the residual field (Figure 8f), horizontal gradient magnitude of the residual field (HGM_t) (Figure 8g), analytical signal amplitude (AS) (Figure 8h), and Tilt derivative (Figure 8i). To enhance relatively deep magnetic responses and suppress shallow interference, the derivative indicators were calculated from the 6 km residual magnetic field. Among these indicators, the Tilt derivative provides a normalized representation of magnetic-field gradients and enhances the boundaries of both strong and weak anomalies. The calculation formulas for AS and the Tilt derivative are as follows (Equations (8) and (9)):

$$AS = \sqrt{\left(\frac{\partial T}{\partial x}\right)^2 + \left(\frac{\partial T}{\partial y}\right)^2 + \left(\frac{\partial T}{\partial z}\right)^2} \quad (8)$$

where T is the magnetic anomaly residual value after RTP, and $\partial T/\partial x$, $\partial T/\partial y$, and $\partial T/\partial z$ are the first-order derivatives of the magnetic anomaly residual value in the x , y , and z directions, respectively.

$$Tilt = \arctan\left(\frac{\partial T/\partial z}{HGM_t}\right) \quad (9)$$

where T is the magnetic anomaly residual value after RTP, $\partial T/\partial z$ is the vertical first-order derivative of the magnetic field, and HGM_t is the horizontal gradient magnitude of the magnetic field.

3.4. Geochemical Comprehensive Anomaly Analysis

Geochemical comprehensive anomaly analysis utilized 1:200,000 stream sediment geochemical survey data, obtaining contour maps for 38 elements. Sampling locations were selected in parts of the stream system favorable for the accumulation of fine-grained materials. To ensure data quality, pollution sources such as buildings and major transportation routes were avoided, with samples collected 20–30 m upstream from any potential contamination sources. Given the low sampling density of the 1:200,000-scale survey, the outer-zone assemblage of the orebody front halo was selected as the analysis target because it generally exhibits better continuity and statistical regularity in regional geochemical fields, thereby providing a more stable transition from background to anomaly. Accordingly, Au–Ag–Cu–Pb–Zn was used to construct the comprehensive geochemical anomaly factor (Chem_F). This assemblage effectively characterizes the elemental zoning patterns of the hydrothermal system in both horizontal and vertical directions.

To address the issues of low resolution and uneven spatial distribution of the geochemical data, the SOMML method developed in our previous study [31] was adopted as the methodological basis. SOMML consists of SOM-based topological mapping, metric-learning-based anomaly extension, and anomaly-boundary extraction. In the standard SOM framework, the network first maps high-dimensional data by identifying the Best Matching Unit (BMU) for a given input sample x , which is determined by the minimum distance to the weight vector w_i of the i -th neuron:

$$d_{BMU} = \min\|x - w_i\| \quad (10)$$

Subsequently, the network updates the weights of the BMU and its neighboring neurons over t training iterations using a learning rate $\alpha(t)$ and a Gaussian neighborhood function $h(i, \text{BMU})$:

$$w_i(t+1) = w_i(t) + \alpha(t) \cdot h(i, \text{BMU}) \cdot (x - w_i(t)) \quad (11)$$

$$h(i, \text{BMU}) = \exp\left(-\frac{\|r_i - r_{\text{BMU}}\|^2}{2\sigma(t)^2}\right) \quad (12)$$

where $\sigma(t)$ represents the neighborhood radius, and r_i and r_{BMU} are the position vectors of the i -th neuron and the BMU, respectively. Once the topological features are mapped, metric learning evaluates the similarity between anomalous neurons and neighboring neurons based on their distance $d(x, y)$ in the feature space, thereby supporting anomaly-boundary delineation:

$$\text{similarity}(x, y) = \frac{1}{d(x, y) + 1} \quad (13)$$

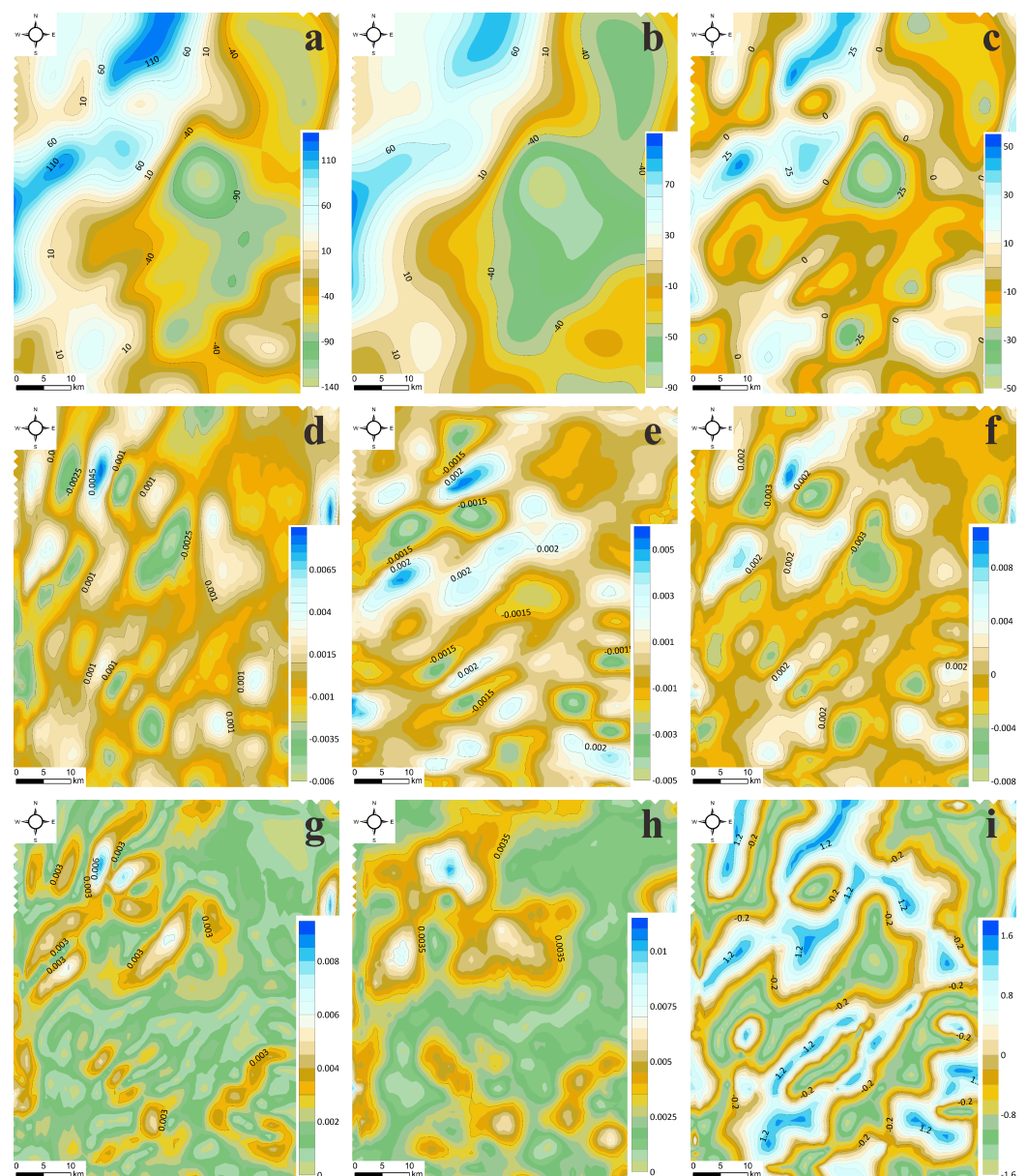


Figure 8. Magnetic feature maps. (a) RTP magnetic anomaly, (b) 6 km upward continued magnetic field, (c) 6 km residual magnetic anomaly, (d) X-direction first-order derivative of residual field, (e) Y-direction first-order derivative of residual field, (f) vertical first-order derivative of residual field, (g) horizontal gradient magnitude of residual field, (h) analytical signal amplitude, (i) tilt derivative.

Building on the original SOMML method, this study extended it into a geologically constrained SOMML model (G-SOMML) by introducing borehole-derived ore-bearing and non-ore-bearing information as geological prior constraints. The methodological extension operates at two stages. First, the borehole information is used to construct prior mineralization samples, thereby enhancing the aggregation of geochemical samples associated with similar metallogenic backgrounds in the feature space. Second, during anomaly extension, the prior information is used to refine the similarity relationships among neighboring neurons, making neurons with consistent metallogenic backgrounds more likely to be incorporated into the same anomaly system. In contrast, the original SOMML method determines anomaly identification and extension from the topological structure and inter-neuron similarities learned from the geochemical data without borehole-derived geological constraints. Thus, G-SOMML retains the nonlinear mapping and anomaly-extension architecture of SOMML while introducing geological constraints into feature-space representation and anomaly extension.

Accordingly, we prepared a common borehole dataset for the G-SOMML constraints and the subsequent supervised analyses. We vectorized a total of 2914 borehole records from the geological archives of the Sanshandao and Jiaojia areas and represented them as two-dimensional ore-bearing or non-ore-bearing points. To address the extremely uneven spatial distribution of the boreholes and local conflicts between positive and negative samples, we applied the k -nearest neighbors algorithm ($k = 5$) for neighborhood cleaning. This process removed 734 locally conflicting records and retained 2180 valid borehole records, including 1593 ore-bearing points and 587 non-ore-bearing points, corresponding to a positive-to-negative sample ratio of approximately 2.7:1. We consistently used this cleaned dataset for the G-SOMML geological constraints, Random Forest feature-importance evaluation, SHAP analysis, and comparative RF modeling. Prior to the corresponding model analyses, we standardized the continuous input variables using Z-score normalization to eliminate dimensional differences.

The known gold occurrences used for spatial evaluation constitute a separate point dataset from the borehole records used as G-SOMML constraints, although the two datasets are spatially related in the intensively explored Sanshandao and Jiaojia areas. To examine the potential influence of this spatial relationship, the evaluation results were reported separately for these two districts and for the remaining parts of the study area. The results of the comprehensive anomaly analysis (Figure 9) show that the anomalous zones and similar regions extracted by G-SOMML exhibit distinct NE–NNE-trending band-like patterns, consistent with the structural control of altered-rock-type gold mineralization in the study area. Anomalous responses were identified both within the principal borehole-concentrated districts and around known gold occurrences in other parts of the study area. Spatial quantitative evaluation using 45 known altered-rock-type gold occurrences showed that 21 of the 22 occurrences in the Sanshandao and Jiaojia areas were located within the extracted anomalous zones, corresponding to a deposit capture rate of 95.5%. In the other parts of the study area, 17 of the 23 known occurrences were captured, corresponding to a deposit capture rate of 73.9%. Overall, 38 of the 45 known occurrences were located within the anomalous zones, yielding an overall deposit capture rate of 84.4%. These results show that the anomalous zones extracted by G-SOMML also maintained spatial correspondence with known gold occurrences beyond the Sanshandao and Jiaojia areas, although the lower capture rate indicates reduced performance in the other parts of the study area. The spatial correspondence supports the use of Chem_F as a regional mineralization-related geochemical indicator and demonstrates the applicability of G-SOMML to low-density regional geochemical data.

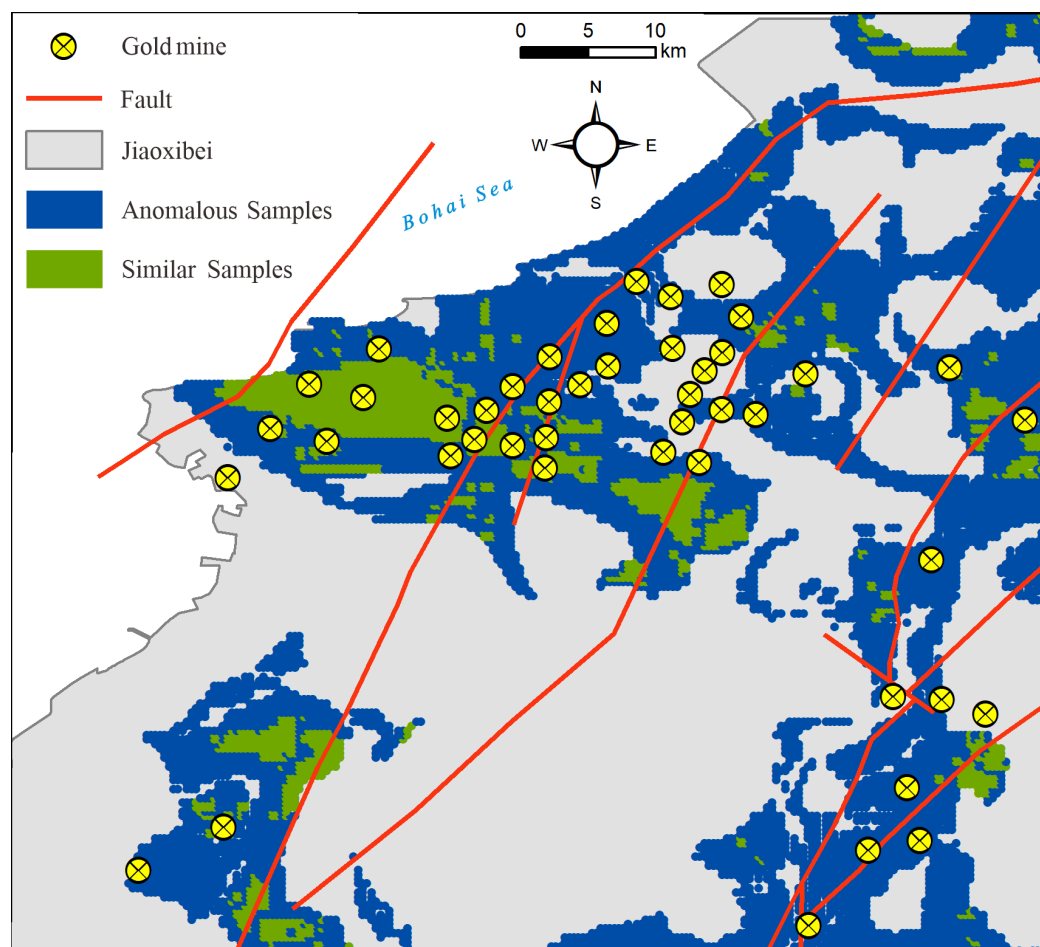


Figure 9. Results of geochemical comprehensive anomaly analysis.

4. Feature Engineering and Mineral Potential Evaluation

The methodological workflow links the indicator layers constructed in Section 3 to the final mineral potential map through a sequential data flow. The 32 initial indicators are first assembled into category-specific feature pools; within-category correlation analysis then reduces redundancy, after which RF importance, SHAP response patterns, and geological screening identify nine core factors. These factors are finally converted into fuzzy memberships according to their mineralization-response directions and integrated using category-balanced weights, producing a continuous potential surface from which the high-potential zones are statistically delineated.

4.1. Feature Pool Initialization

Based on the aforementioned mineralization indicator system, this study performed a parameterized integration of the extracted multi-source information from structures, remote sensing, gravity, magnetism, and geochemistry to construct an initial evaluation feature pool. This feature pool contains a total of 32 evaluation indicators (Table 1), aiming to characterize the metallogenic geological background and ore-controlling anomaly information in the study area from multi-scale and multi-dimensional perspectives.

Among these, there are 4 structural indicators, 9 remote sensing indicators, 18 geophysical indicators (9 each for gravity and magnetic data), and 1 geochemical comprehensive anomaly indicator. Because the geochemical indicator (Chem_F), representing the comprehensive anomaly of Au–Ag–Cu–Pb–Zn, is the sole indicator within the geochemical feature category, within-category redundancy analysis was not applicable to this category. Therefore, Chem_F was directly retained in the candidate feature pool and was subse-

quently included, together with the retained non-geochemical indicators, in the Random Forest importance ranking and SHAP analysis. Due to significant differences in physical dimensions and numerical magnitudes among the non-geochemical multi-source data, the 31 non-geochemical continuous variables were standardized using Z-score normalization to eliminate the influence of dimensional differences on machine learning model training. The standardized values of these variables, together with Chem_F, were then extracted at the 2180 retained borehole locations for subsequent feature evaluation.

Table 1. Initialization of the evaluation feature pool.

Category	Indicator ID	Meaning
Structure	Prox_fault	Structural proximity
	FD_2k	Structural fractal dimension (2 km)
	FD_5k	Structural fractal dimension (5 km)
	Junc_dens	Intersection density
Remote Sensing	Seri_Int	Sericite alteration intensity
	Chl_Int	Chlorite alteration intensity
	Iron_Int	Iron staining alteration intensity
	Seri_KDE	Sericite alteration influence intensity
	Chl_KDE	Chlorite alteration influence intensity
	Iron_KDE	Iron staining alteration influence intensity
	Seri_Dens	Sericite alteration zoning density
	Chl_Dens	Chlorite alteration zoning density
Geophysics-Gravity	Iron_Dens	Iron staining alteration zoning density
	Z_value	Original gravity field
	Z_vd1	Vertical first-order derivative (VD1)
	Z_res3k	3 km residual gravity field
	Z_up3k	3 km upward continued field
	Z_up6k	6 km upward continued field
	Z_up6_vd1	6 km continued vertical derivative
	Z_res3_xvd	X-direction first-order derivative
	Z_res3_yvd	Y-direction first-order derivative
Geophysics-Magnetism	Z_res3_hgm	Horizontal gradient magnitude (HGM)
	C_rtp	Reduction to the pole (RTP)
	Cr_res6k	6 km residual magnetic anomaly
	Cr_res6_vd	Vertical derivative of the residual field
	Cr_es6_xvd	X-derivative of the residual field
	Cr_es6_yvd	Y-derivative of the residual field
	Cr_es6_hgm	Horizontal gradient of the residual field (HGM)
	Cr_res6_as	Analytical signal amplitude (AS)
	Cr_e6_tilt	Tilt derivative (Tilt angle)
Geochemistry	Cr_up6k	6 km upward continued magnetic field
	Chem_F	Au–Ag–Cu–Pb–Zn comprehensive anomaly

4.2. Establishment of Feature Engineering

In mineral potential evaluation, feature engineering is not only a vital stage of data preprocessing but also a key process for the effective expression of geoscientific information. Departing from traditional processing methods primarily focused on statistical feature screening, this study views feature engineering as a reconstruction process of multi-source geoscientific information under the constraints of metallogenic system theory. By constructing representation indicators with explicit geological significance and incorporating redundancy control and collinearity reduction, the transformation of the feature space from high-dimensional coupling to low-dimensional effective expression is achieved. This process emphasizes the guiding role of geological knowledge in feature construction and

screening, thereby enhancing the interpretability and mineralization significance of the model input variables. On this basis, the feature space is optimized and reconstructed by further combining correlation analysis with model-based feature importance evaluation.

4.2.1. Feature Redundancy Analysis

Pearson correlation analysis was conducted separately within each feature category to characterize linear dependence among the continuous predictors, particularly among factors derived from common source fields or related transformations. The absolute correlation coefficients were interpreted hierarchically: values below 0.5 indicated weak correlation, values from 0.5 to 0.7 indicated moderate correlation, values above 0.7 indicated strong correlation, and values above 0.85 indicated a particularly high degree of redundancy. These ranges were used to identify potentially redundant feature pairs. Feature retention further considered the overlap in geological or physical meaning, scale representativeness, and complementarity with other indicators. Correlated features were retained when they characterized distinct spatial scales or geological processes, while the more representative feature was selected when strongly correlated variables contained substantially overlapping information. The retained candidate features were subsequently examined using RF importance and SHAP analysis to evaluate their predictive contributions and response patterns.

The analysis results of the structural features (Figure 10) show that the overall correlation among various fault features is low. Only Prox_fault and FD_2k exhibit a moderate negative correlation, quantitatively revealing the geological fact that local structural complexity increases in areas closer to fault zones. Notably, the correlation between FD_2k and FD_5k is extremely weak, indicating that fractal dimensions at different scales characterize local structural perturbations and regional structural backgrounds, respectively, demonstrating clear scale complementarity. Given that the various structural indicators have different physical orientations regarding spatial proximity, geometric complexity, and connectivity, all four indicators in the fault feature pool are retained.

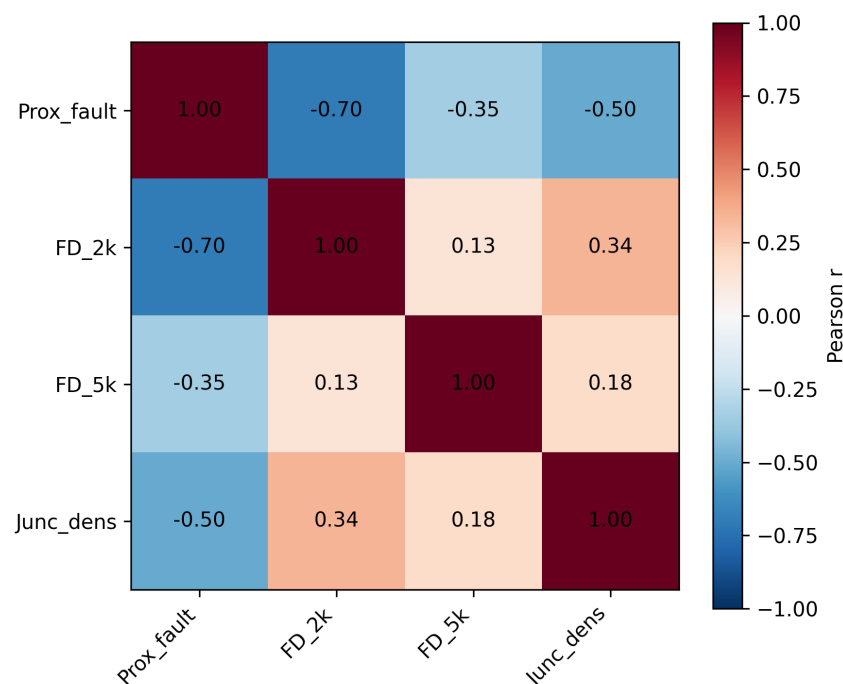


Figure 10. Pearson correlation matrix for the structural feature pool.

The analysis results of the remote sensing features (Figure 11) show that the spatial distributions of different mineral alterations exhibit good independence. However, within the same mineral type, the influence intensity of sericite alteration is strongly correlated with its zoning density, reflecting significant redundancy in characterizing the spatial aggregation features of sericite anomalies. Considering the large scale and continuity of sericitization in the study area, the sericite alteration zoning density indicator (Seri_Dens) is removed, while the remaining indicators are retained.

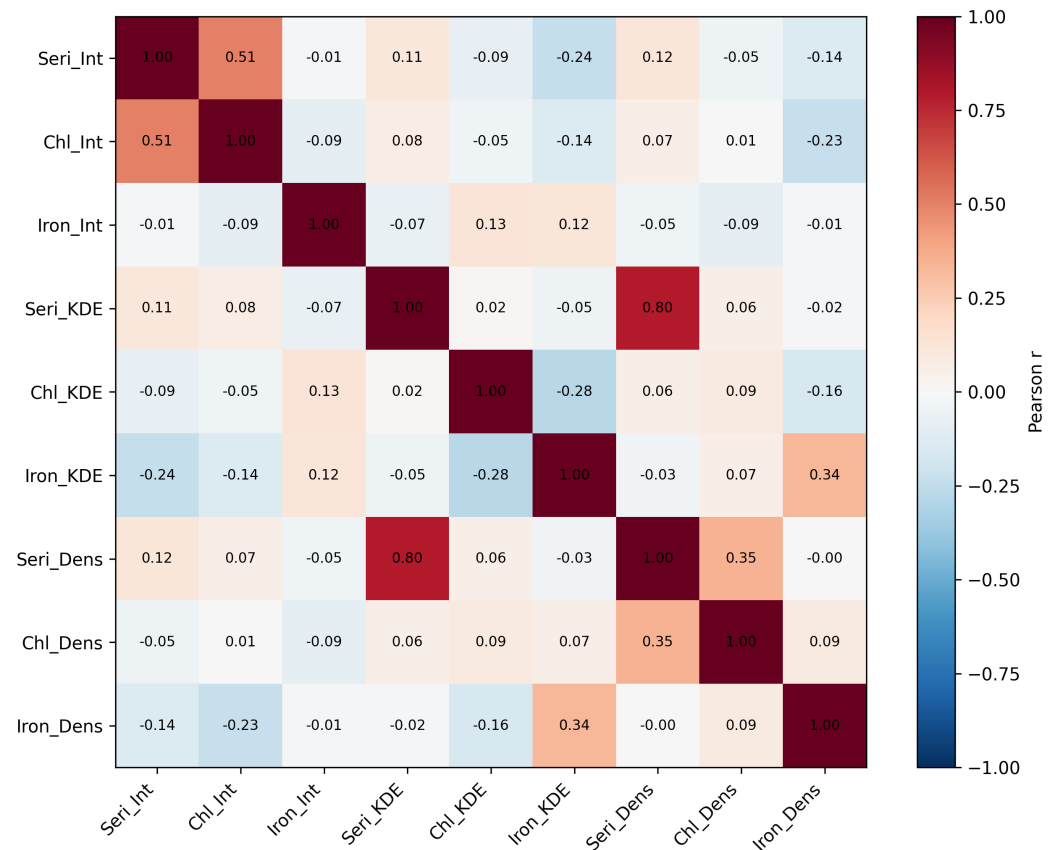


Figure 11. Pearson correlation matrix for the remote sensing feature pool.

Substantial multicollinearity exists within the gravity feature pool (Figure 12). The original gravity field (Z_{value}), the 3 km and 6 km upward continued fields (Z_{up3k} and Z_{up6k}), and the vertical derivative of the 6 km continued field (Z_{up6_vd1}) exhibit strong correlations, reflecting their common representation of the regional density background. To reduce redundancy while preserving the deep background information, Z_{up6k} was retained, whereas Z_{value} , Z_{up3k} , and Z_{up6_vd1} were removed. The 3 km residual gravity field (Z_{res3k}) is strongly correlated with the vertical first-order derivative (Z_{vd1}); therefore, Z_{vd1} was retained to represent local vertical variations, while Z_{res3k} was removed. Although moderate correlations occur among the directional derivatives and horizontal gradient magnitude of the residual field, Z_{res3_xvd} , Z_{res3_yvd} , and Z_{res3_hgm} were retained because they provide complementary representations of directional changes and density-contrast boundaries. Consequently, the gravity feature pool retained five indicators: Z_{up6k} , Z_{vd1} , Z_{res3_xvd} , Z_{res3_yvd} , and Z_{res3_hgm} .

The magnetic feature pool exhibits substantial multicollinearity (Figure 13). The correlation coefficient between the reduction-to-the-pole magnetic anomaly (C_{rtp}) and the 6 km upward continued field (Cr_{up6k}) reaches 0.989, indicating that both factors primarily represent the regional magnetic background. To reduce redundancy while retaining the relatively deep magnetic response, Cr_{up6k} was retained and C_{rtp} was removed.

The 6 km residual magnetic field (Cr_res6k) and its vertical, directional-gradient, and horizontal-gradient indicators contain partially overlapping information on local magnetic variations and source boundaries. Among these factors, Cr_e6_tilt was retained because its normalized gradient response enhances both strong and weak magnetic boundaries, while Cr_res6_as was retained to provide a complementary amplitude-based representation of magnetic-source responses. Accordingly, Cr_res6k, Cr_res6_vd, Cr_es6_xvd, Cr_es6_yvd, and Cr_es6_hgm were removed. The magnetic feature pool therefore retained three indicators: Cr_up6k, Cr_e6_tilt, and Cr_res6_as.

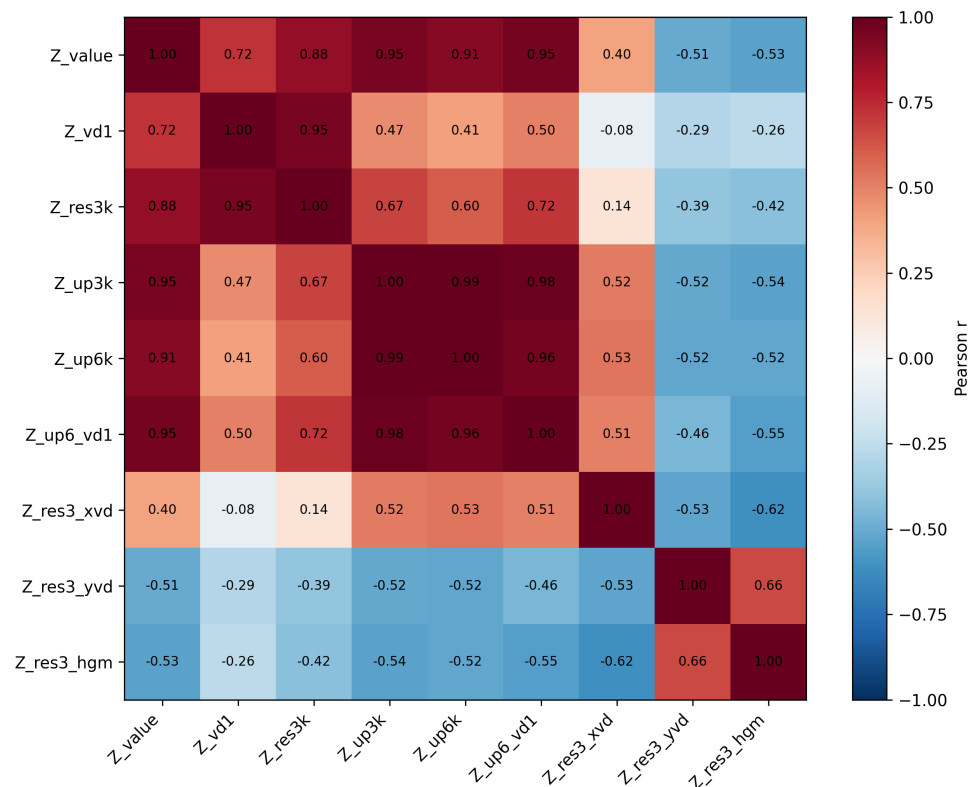


Figure 12. Pearson correlation matrix for the gravity feature pool.

In summary, redundancy analysis was conducted separately within the structural, remote-sensing, gravity, and magnetic feature categories, covering 31 non-geochemical indicators in total. A total of 11 highly redundant indicators were removed, leaving 20 retained non-geochemical features. Because Chem_F was the sole geochemical indicator, within-category redundancy analysis was not applicable to the geochemical category, and this indicator was directly retained. Consequently, the candidate feature set comprised 21 features for subsequent importance evaluation.

4.2.2. Feature Importance Evaluation

The Random Forest algorithm was employed to evaluate the importance of the 21 candidate features, comprising the 20 non-geochemical features retained after redundancy analysis and the directly retained geochemical indicator Chem_F (Figure 14). The cleaned borehole dataset was randomly divided into training and testing subsets at a ratio of 8:2 using `random_state = 42`. The RF model comprised 100 decision trees, and each tree was trained using bootstrap sampling from the training subset. No additional sample resampling or class-weight adjustment was applied, and the remaining hyperparameters followed the default settings of the scikit-learn implementation. Feature importance was quantified using the model's built-in impurity-based importance measure. The resulting

scores characterize the relative statistical contributions of the candidate factors to the discrimination of ore-bearing and non-ore-bearing borehole samples. The RF importance scores were subsequently interpreted together with feature redundancy, SHAP responses, geological meaning, and information complementarity during the multi-criteria screening.

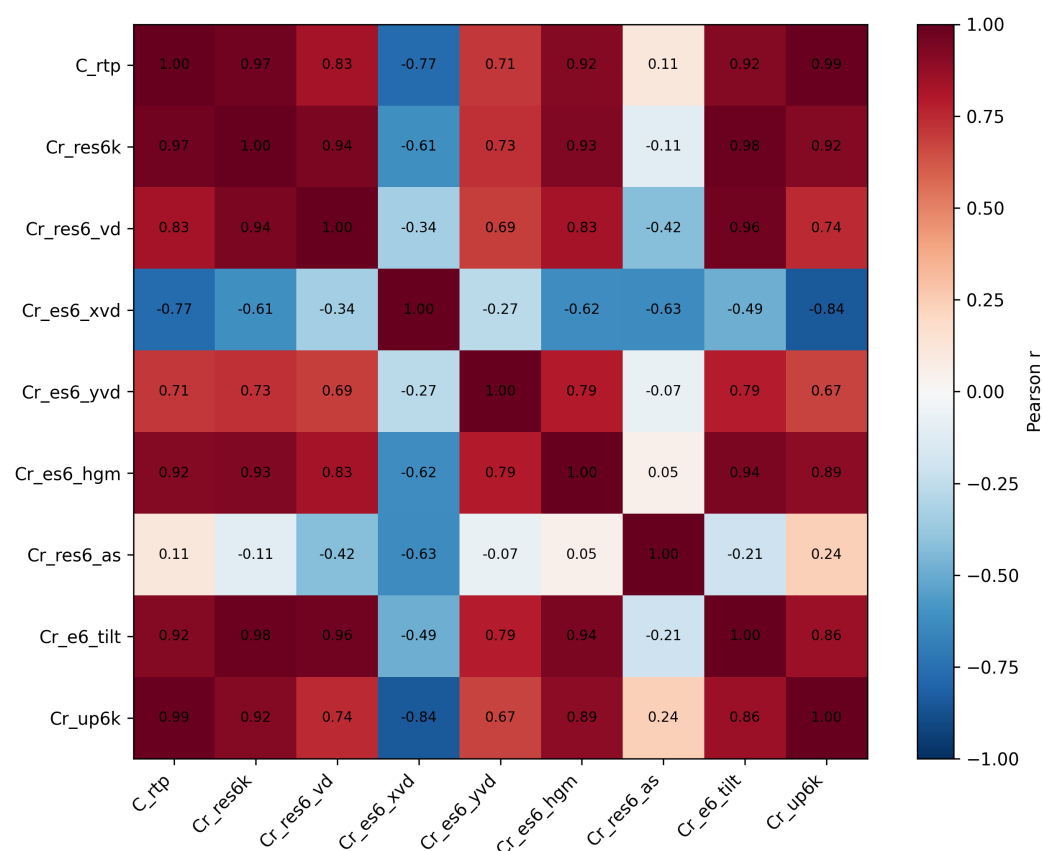


Figure 13. Pearson correlation matrix for the magnetic feature pool.

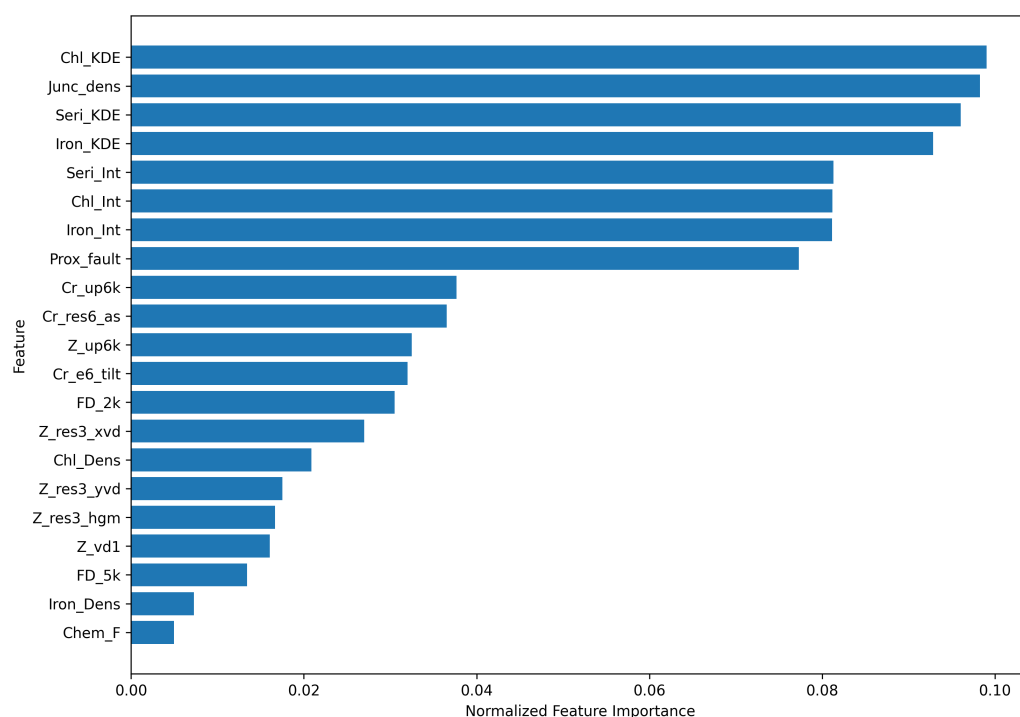


Figure 14. Feature importance ranking based on Random Forest.

First, the spatial aggregation of alteration anomalies and the features associated with structural intersections play dominant roles in the model. The RF feature importance ranking shows that the top three features are Chl_KDE, Junc_dens, and Seri_KDE. This result suggests that in the preliminary discrimination of mineralization probability, the spatial aggregation characteristics of alteration anomalies and the structurally complex areas formed by fault intersections are the most discriminative statistical indicators. In particular, the continuous anomaly fields constructed based on Kernel Density Estimation (KDE) exhibit significantly higher importance than point density indicators, supporting the effectiveness of incorporating distance-decay effects into the spatial representation of surficial alteration information.

Second, the feature importance results demonstrate the complementary value of sub-surface geophysical information. Among the geophysical indicators, the residual-gravity horizontal derivative (Z_{res3_xvd}) and magnetic tilt angle (Cr_e6_tilt) contributed additional discriminatory information to the fitted RF model. After potential-field separation and gradient enhancement, these variables characterize lateral variations and boundaries in sub-surface density and magnetic properties that may be associated with regional mineralization.

To further examine the magnitude and positive or negative response directions of the feature contributions and their geological implications, SHAP (SHapley Additive exPlanations) analysis was used to quantify the influence of each feature on the RF predictions. Impurity-based RF importance and SHAP values characterize different aspects of the fitted model and were therefore interpreted as complementary rather than interchangeable measures. The former summarizes the cumulative impurity reduction associated with feature-based splits across the ensemble, whereas the latter quantifies the contribution of each feature to individual model predictions. Both measures were considered together with feature redundancy, geological meaning, and information complementarity in the subsequent multi-criteria screening. The global importance summary plot (Figure 15) shows a distinct gradient in feature contributions. Among the geophysical indicators, the magnetic analytic signal (Cr_res6_as), residual-gravity horizontal derivative (Z_{res3_xvd}), and magnetic tilt angle (Cr_e6_tilt) showed comparatively high global contributions in the fitted model, with mean absolute SHAP values above 0.025. Together with the alteration KDE indicators and structural intersection density, these factors constitute the principal discriminatory feature group, while the geophysical derivative indicators provide complementary information on subsurface physical-property boundaries for mineral potential evaluation.

The quantitative analysis results (Table 2) show the directional associations between the feature values and their SHAP contributions to the RF predictions. Among the geophysical factors, higher values of the magnetic analytical signal (Cr_res6_as) were associated with positive contributions to the predicted mineralization class, whereas the residual gravity X-direction derivative (Z_{res3_xvd}) exhibited a predominantly negative association. As a signed directional derivative, Z_{res3_xvd} distinguishes the orientation and relative position of lateral density transitions. Its negative association therefore reflects the side of the density-gradient boundary preferentially represented by the ore-bearing boreholes. These patterns reflect the sensitivity of the analytical signal to magnetic-source boundaries and the occurrence of mineralization-related responses along transitions in the regional density field. The positive response of the residual gravity horizontal-gradient magnitude (Z_{res3_hgm}) indicates the contribution of density-contrast boundaries, while the negative response of the magnetic Tilt derivative (Cr_e6_tilt) characterizes magnetic-gradient transition zones associated with positive RF predictions. These associations summarize the response patterns represented by the fitted RF model and the present borehole dataset.

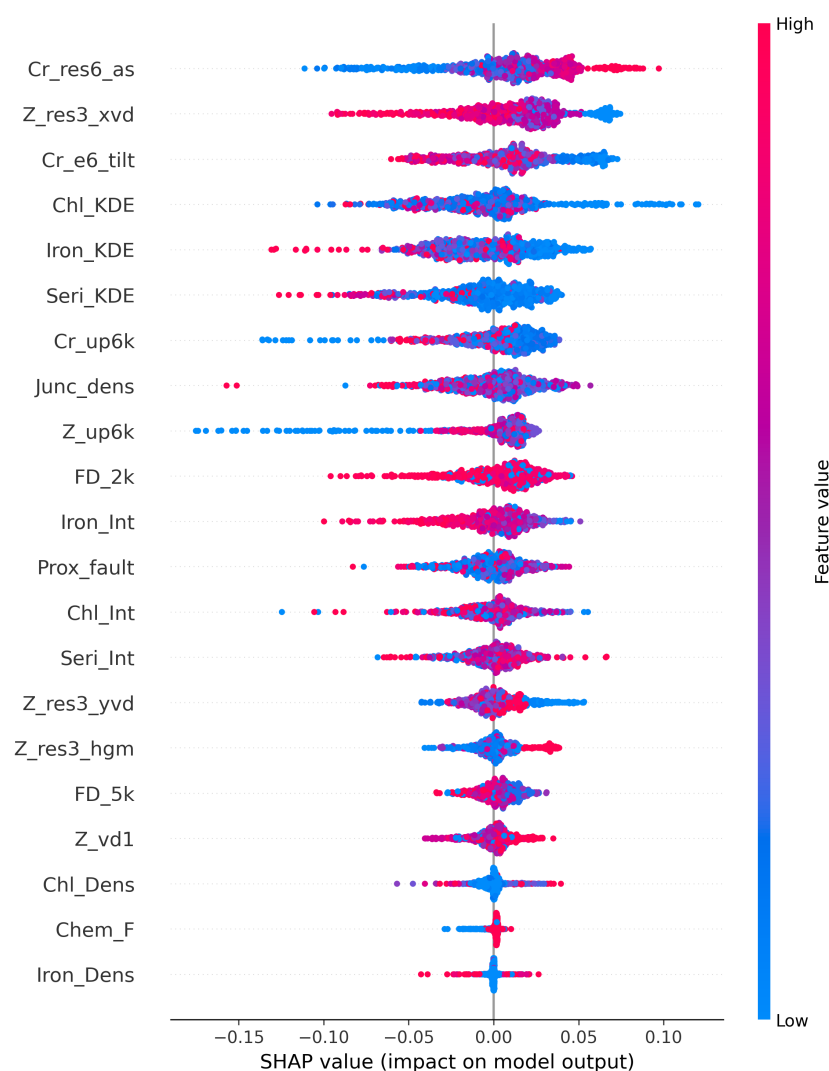


Figure 15. Global importance summary plot (SHAP).

Table 2. Quantitative analysis of SHAP values.

Factor Name	Category	Correlation	Directionality
Prox_fault	Structure	0.296	Positive
Junc_dens	Structure	−0.229	Negative
Z_res3_xvd	Geophysics-Gravity	−0.773	Negative
Z_res3_hgm	Geophysics-Gravity	0.336	Positive
Cr_res6_as	Geophysics-Magnetism	0.770	Positive
Cr_e6_tilt	Geophysics-Magnetism	−0.657	Negative
Chl_KDE	Remote Sensing	−0.274	Negative
Seri_KDE	Remote Sensing	−0.401	Negative
Chem_F	Geochemistry	0.456	Positive

The structural proximity factor (*Prox_fault*) showed a moderate positive association with the RF predictions. Because this factor represents the Euclidean distance to the nearest fault and the borehole samples were generally distributed within fault-proximal settings, the observed SHAP pattern primarily reflects variations among boreholes within these structural settings. Structural-intersection density (*Junc_dens*) exhibited a non-monotonic response within the borehole dataset, with an overall moderate negative association. Chloritization (*Chl_KDE*) and sericitization (*Seri_KDE*) also exhibited predominantly negative SHAP responses. The KDE layers represent spatially smoothed alteration influence, and

broad high-value zones may include extensive hydrothermal or weathering-related responses, while ore-bearing sites are concentrated within more localized, structurally focused parts of the alteration system. This spatial relationship provides a geological context for the observed SHAP responses in the borehole dataset. These statistical response patterns were considered together with the definitions and geological implications of the factors in the subsequent fuzzy membership transformation.

In summary, nine core factors were identified through a multi-criteria screening procedure integrating relative RF importance, mean absolute SHAP values, SHAP response patterns, and geological interpretation. Candidate factors were prioritized according to their statistical contributions, distinct geological meanings, and complementarity among the different categories of evidence. Factors providing substantially overlapping geological or physical information were excluded in favor of more representative variables. The resulting subset therefore balances statistical discrimination, feature-space parsimony, geological completeness, and complementarity among the different information sources.

Prox_fault and Junc_dens were retained to represent fault proximity and structural connectivity, respectively. Z_res3_xvd, Z_res3_hgm, Cr_res6_as, and Cr_e6_tilt were selected to characterize gradients and boundaries in the subsurface density and magnetic-property fields, thereby providing complementary information where surface geological observations are limited by Quaternary cover. Chl_KDE and Seri_KDE were retained because their continuous fields characterize the spatial influence of chloritization and sericitization, while their negative SHAP associations provide complementary information on the distribution of surface-alteration responses relative to the borehole samples. Chem_F provided the only geochemical evidence in the candidate feature set and showed a clear positive SHAP association, indicating that stronger comprehensive Au–Ag–Cu–Pb–Zn anomalies contributed positively to the RF predictions. Together, the nine factors represent structural configuration, surficial alteration, subsurface physical-property boundaries, and geochemical anomalies while reducing redundancy in the initial feature space.

4.3. Fuzzy Synthetic Evaluation Method

The fuzzy synthetic evaluation model adopted in this study, hereafter referred to as Fuzzy, follows the general aggregation procedure described in our previous work [32]. In the present framework, Fuzzy was applied as the final integration component for the nine core factors identified through correlation-based redundancy reduction, RF importance evaluation, SHAP interpretation, and geological screening. The membership directions were assigned by considering the factor definitions, geological implications, and SHAP response patterns, while category-balanced weights were used to coordinate the contributions of the different information sources. The methodological development therefore integrates geology-guided feature construction, borehole-constrained geochemical anomaly extraction, multi-criteria feature screening, membership transformation, and balanced fuzzy aggregation within a unified workflow. The evaluation procedure comprised five steps: construction of the factor domain, membership transformation, determination of the weight vector, fuzzy weighted overlay, and statistical delineation of the high-potential zones.

We transformed the nine selected factor layers into fuzzy membership values and integrated them directly through the category-balanced weighted overlay. We predefined the category weights to balance the five information sources. We used known gold occurrences in the subsequent spatial evaluation, specified the membership-function parameters according to the characteristics of the individual factors, and derived the statistical anomaly threshold from the distribution of the output potential scores. The preceding borehole-based RF and SHAP analyses informed feature selection and provided statistical evidence for the membership transformation, while the final directions additionally incorporated

the definitions and geological implications of the factors. We used linear membership functions to transform the factor values into the fuzzy interval $[0, 1]$. For the remote-sensing layers, we assigned a membership degree of 0 to cells coded as NoData (−9999) by the predefined spatial mask and retained them in the weighted overlay with their prescribed factor weights unchanged. We excluded NoData values caused by missing or invalid data in the other factor layers when determining the membership-function parameters and defining the valid evaluation domain.

For factors whose higher values were treated as more favorable for mineralization, the monotonically increasing membership function was defined as

$$\mu_i^+(x) = \begin{cases} 0, & x \leq a_i, \\ \frac{x - a_i}{b_i - a_i}, & a_i < x < b_i, \\ 1, & x \geq b_i, \end{cases} \quad (14)$$

whereas for factors whose lower values were treated as more favorable for mineralization, the monotonically decreasing membership function was defined as

$$\mu_i^-(x) = \begin{cases} 1, & x \leq a_i, \\ \frac{b_i - x}{b_i - a_i}, & a_i < x < b_i, \\ 0, & x \geq b_i. \end{cases} \quad (15)$$

Here, a_i and b_i denote the minimum and maximum valid values of factor i within the evaluation domain, respectively, and were used as the lower and upper breakpoints of the membership transformation. Based on the factor definitions, geological implications, and observed SHAP response patterns, monotonically increasing membership functions were assigned to Junc_dens, Z_res3_hgm, Cr_res6_as, and Chem_F, whereas monotonically decreasing membership functions were assigned to Prox_fault, Z_res3_xvd, Cr_e6_tilt, Chl_KDE, and Seri_KDE. Because Prox_fault represents the distance to the nearest fault, smaller values were assigned higher membership degrees. By contrast, higher Junc_dens values were assigned higher membership degrees to represent the regional favorability of structurally connected fault-intersection zones. The decreasing transformations of Chl_KDE and Seri_KDE downweighted the highest alteration-influence intensities, consistent with their predominantly negative SHAP associations in the borehole dataset.

The category-balanced weighting scheme was established through expert geological judgment based on the relative roles of the five information categories in the regional metallogenic system. Equal category-level weights of 0.2 were assigned to structure, remote sensing, gravity, magnetics, and geochemistry. The structural, remote-sensing, gravity, and magnetic categories each contained two selected factors; therefore, their category weights were divided equally, resulting in an individual factor weight of 0.1. Because geochemistry was represented by the single comprehensive anomaly factor Chem_F, this factor retained the full category weight of 0.2.

Accordingly, the weight vector was defined as

$$\mathbf{w} = [0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.2]^T, \quad (16)$$

where the factor order is Prox_fault, Junc_dens, Chl_KDE, Seri_KDE, Z_res3_xvd, Z_res3_hgm, Cr_res6_as, Cr_e6_tilt, and Chem_F. The membership-function parameters and factor weights are summarized in Table 3.

Table 3. Membership-function parameters and weights of the nine core evaluation factors.

Factor	Membership Function	a_i	b_i	Weight
Prox_fault	Linear decreasing	0	26,929.801	0.10
Junc_dens	Linear increasing	0	1	0.10
Chl_KDE	Linear decreasing	0	42	0.10
Seri_KDE	Linear decreasing	0	255	0.10
Z_res3_xvd	Linear decreasing	-4.61177×10^{-3}	2.64257×10^{-3}	0.10
Z_res3_hgm	Linear increasing	2.24457×10^{-5}	4.86659×10^{-3}	0.10
Cr_res6_as	Linear increasing	1.37320×10^{-4}	6.90525×10^{-3}	0.10
Cr_e6_tilt	Linear decreasing	−1.551383	1.518627	0.10
Chem_F	Linear increasing	0	1	0.20

For each grid cell p , the mineral potential score was calculated using the weighted-average fuzzy operator:

$$S(p) = \sum_{i=1}^9 w_i \mu_i(x_i(p)), \quad \sum_{i=1}^9 w_i = 1, \quad (17)$$

where $x_i(p)$ is the value of factor i at grid cell p , μ_i is its corresponding membership function, and w_i is its factor weight. The weighted fuzzy overlay was performed cell by cell on the co-registered raster layers, thereby producing a continuous mineral potential score surface for subsequent delineation and spatial evaluation.

A standardized statistical anomaly rule was used to delineate the high-potential zones:

$$T = \bar{S} + k\sigma_S, \quad (18)$$

where $\bar{S}$ and σ_S are the mean and standard deviation of the output potential scores within the valid evaluation domain, respectively, and k is the threshold multiplier. A value of $k = 1$ was adopted to delineate the upper-score domain exceeding the mean by one standard deviation. The same thresholding rule was subsequently applied to all model outputs in the comparative evaluation.

4.4. Mineral Potential Evaluation Results

Application of the statistical anomaly rule to the Fuzzy output yielded a mean potential score of $\bar{S} = 0.3827$ and a standard deviation of $\sigma_S = 0.0934$. Accordingly, the high-potential threshold was determined as $T = 0.476$. Grid cells with potential scores greater than this threshold were classified as high-potential zones, whereas the remaining cells were treated as background areas. The resulting mineral potential map is shown in Figure 16.

The mineral potential results exhibit a distinct NE–NNE-trending spatial pattern. The principal high-potential zones extend along the Sanshandao, Jiaojia, and Zhaoping fault systems and their associated secondary structures. Previous studies indicate that these major faults were reactivated during gold mineralization and that secondary fractures, fault bends, and intersections provided favorable pathways for hydrothermal-fluid migration and sites for gold deposition [37–39]. The predicted spatial pattern is therefore consistent with the established structural control on altered-rock-type gold mineralization in the Jiaoxibe area.

The high-potential zones reflect the integrated expression of multiple mineralization-related indicators. The structural factors characterize regional fluid pathways and structurally favorable sites, whereas Chl_KDE and Seri_KDE represent the spatial influence of hydrothermal alteration. Previous studies have shown that sericitization, chloritization, silicification, and pyritization are widely developed in the fractured alteration zones of

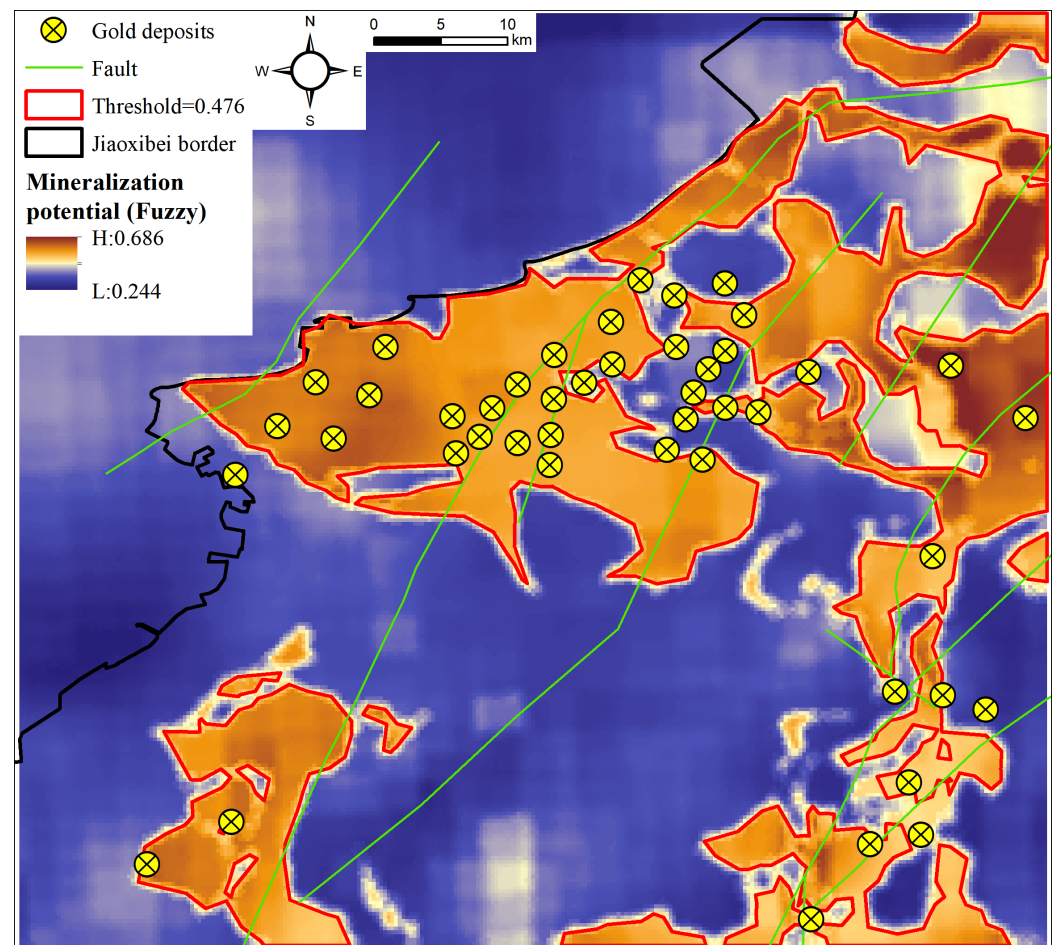


Figure 16. Mineral potential evaluation result obtained using the fuzzy synthetic evaluation method.

Jiaoxibei gold deposits and exhibit systematic spatial zoning around the principal mineralized structures [2,41,42]. However, strong or spatially extensive alteration does not necessarily correspond directly to economic mineralization, which is consistent with the decreasing membership transformations of the two alteration KDE factors. The gravity and magnetic derivatives provide complementary information on boundaries and lateral variations in subsurface density and magnetic properties, while Chem_F represents the regional hydrothermal geochemical response expressed by the Au–Ag–Cu–Pb–Zn assemblage. Therefore, the high-potential zones delineate areas where favorable structural, alteration, geophysical, and geochemical evidence occurs jointly rather than areas controlled by a single anomalous factor.

In structurally complex metallogenic settings, the mapped high-potential zones should therefore be interpreted as areas where complementary mineralization-related evidence converges spatially, rather than as direct expressions of any single anomaly. In particular, prospectivity is more strongly supported where major fault corridors and secondary structural complexity coincide spatially with alteration influence, potential-field boundaries, and multi-element geochemical anomalies than where any single high-amplitude response occurs in isolation. This evidence-convergence perspective reduces the risk of treating broad alteration or geophysical anomalies as direct proxies for mineralization.

In the northwestern Bohai Sea area, mineral potential scores are generally lower, reflecting the limited availability of geochemical and remote-sensing evidence in the off-shore region. Spatial comparison with the known gold occurrences showed that 21 of the 22 altered-rock-type gold occurrences in the Sanshandao and Jiaojia areas were located within the high-potential zones, corresponding to a deposit capture rate of 95.5%. In the

other parts of the study area, 17 of the 23 known occurrences were captured, corresponding to a deposit capture rate of 73.9%. Overall, 38 of the 45 known occurrences fell within the high-potential zones, yielding an overall deposit capture rate of 84.4%. The lower capture rate outside the Sanshandao and Jiaojia areas indicates some reduction in spatial performance, whereas the overall spatial correspondence supports the applicability of the integrated result for regional mineral potential evaluation and target delineation. The contribution of the RF- and SHAP-based feature-screening procedure to the prediction result is evaluated further through the comparison between the Fuzzy and AT-Fuzzy models in Section 5.

5. Discussion

Four mineral potential mapping models were evaluated within the common valid evaluation domain. The feature-engineering-based Fuzzy model integrated the nine selected core factors using the linear membership transformations, category-balanced weighting scheme, and weighted-average fuzzy operator described in Section 4.3. The all-feature fuzzy synthetic evaluation model (AT-Fuzzy) followed the same membership-transformation and integration procedures but retained all 32 initial factors. Equal category-level weights of 0.2 were assigned to structure, remote sensing, gravity, magnetics, and geochemistry, and the weight of each category was divided equally among the factors within that category. The comparison between Fuzzy and AT-Fuzzy therefore maintained a consistent fuzzy-integration procedure while differing in the input feature set. The feature-engineering-based Weights of Evidence model (WoE) and Random Forest model (RF) both used the same nine core factors as Fuzzy. WoE quantified the spatial associations between the evidence factors and mineralization and combined the resulting evidence weights at each grid cell under the conditional-independence assumption. RF learned nonlinear relationships between the nine factors and the ore-bearing or non-ore-bearing borehole labels and was subsequently applied to the co-registered raster layers to estimate the probability of the ore-bearing class for each grid cell.

The RF experiment used the same cleaned dataset of 2180 valid borehole records described above, comprising 1593 ore-bearing and 587 non-ore-bearing records. Model performance and potential overfitting were evaluated using a held-out test set and out-of-bag validation. The model achieved a training accuracy of 0.8378, a testing accuracy of 0.7238, and an out-of-bag (OOB) score of 0.7473. The four models generated spatially continuous mineral potential surfaces, but their output values have different meanings. Fuzzy and AT-Fuzzy produced aggregated fuzzy membership scores within the interval $[0, 1]$, representing the combined degree of mineral favorability derived from the input factors. WoE produced a combined evidential score reflecting the cumulative spatial association of the evidence factors with mineralization, whereas RF produced the predicted probability of the ore-bearing class within the interval $[0, 1]$. Although their absolute values and score distributions are model dependent, all four outputs provide continuous rankings from lower to higher mineral potential. Their regional mapping performance was consequently assessed within the same valid evaluation domain using the same statistical anomaly rule and the same map-level evaluation criteria. Borehole information also contributed to the overall feature-engineering framework through the geological constraints introduced into G-SOMML and the RF-SHAP analyses used for feature evaluation and membership transformation.

For the map-level comparison, the deposit capture rate was defined as the percentage of known gold occurrences located within the high-potential zones delineated by each model relative to the total number of known gold occurrences within the common valid evaluation domain. The high-potential zones of each model were determined using the corresponding statistical anomaly threshold described below. This metric was calculated

consistently for the Fuzzy, AT-Fuzzy, WoE, and RF models. It is distinct from the RF training and testing accuracies and the OOB score reported above, which measure borehole-level classification performance rather than the spatial capture performance of the mineral potential maps.

The Prediction–Area (P–A) plots were generated by progressively varying the prospectivity-score threshold from the maximum to the minimum value for each model. At each threshold, the prediction rate was calculated as the proportion of known gold occurrences located in grid cells with scores equal to or greater than the threshold, while the area rate was calculated as the proportion of the common valid evaluation domain occupied by those cells. The resulting plots show the joint variation in mineralization capture and prospective-area extent across the full score range. They complement the single-threshold deposit-capture and area-ratio statistics by displaying the threshold-dependent trade-off between occurrence capture and mapped area, while the success-rate curves and SR-AUC values provide an additional cumulative assessment of spatial-ranking efficiency.

To complement the single-threshold comparison, we constructed success-rate curves to evaluate the cumulative spatial-ranking performance of the four mineral potential maps. For each model, we progressively expanded the prospective zones from higher- to lower-ranked potential levels and plotted the cumulative proportion of known gold occurrences captured against the corresponding proportion of the valid evaluation area. We defined curve nodes at each observed increase in the cumulative number of captured occurrences and calculated the corresponding prospective-area proportions within the common valid evaluation domain. We applied the same occurrence–zone spatial correspondence criterion used in the threshold-based evaluation throughout this analysis. We calculated the area under each success-rate curve (SR-AUC) from the resulting right-continuous step function. A larger SR-AUC indicates that the model captures a greater proportion of known occurrences within a smaller proportion of the evaluation area, thereby providing a cumulative measure of map-level targeting efficiency that does not depend on any single prospective-area threshold.

To ensure comparability among the four model outputs, the same standardized statistical anomaly-thresholding rule was applied to each model. Because the models produce scores with different numerical ranges and statistical distributions, applying a common absolute threshold would not provide an equivalent statistical criterion. Therefore, the threshold for each model was calculated using $T_m = \bar{S}_m + k\sigma_{S,m}$, with $k = 1$, following Equation (18). Here, $\bar{S}_m$ and $\sigma_{S,m}$ denote the mean and standard deviation of the output scores of model m within the common valid evaluation domain, respectively.

As shown in Table 4, application of this consistent rule yielded thresholds of 0.476 for Fuzzy ($\bar{S} = 0.3827$, $\sigma_S = 0.0934$), 0.689 for AT-Fuzzy ($\bar{S} = 0.6473$, $\sigma_S = 0.0415$), 0.719 for WoE ($\bar{S} = 0.0217$, $\sigma_S = 0.6971$), and 0.376 for RF ($\bar{S} = 0.3058$, $\sigma_S = 0.0705$). The different numerical thresholds reflect differences in the score distributions of the four models rather than differences in the thresholding criterion. Accordingly, the prospective-area ratios and deposit capture rates were compared under the same statistical thresholding criterion.

Table 4. Quantitative and cumulative success-rate comparison of the mineral potential mapping results.

Method	Mean ($\bar{S}$)	Standard Deviation (σ_S)	Threshold (T)	Deposit Capture Rate	Area Ratio	SR-AUC
Fuzzy	0.3827	0.0934	0.476	84.4%	25.356%	0.762
AT-Fuzzy	0.6473	0.0415	0.689	51.1%	50.951%	0.386
WoE	0.0217	0.6971	0.719	40.0%	16.457%	0.540
RF	0.3058	0.0705	0.376	71.1%	50.875%	0.590

Based on the quantitative evaluation results (Table 4) and the mineral potential maps (Figure 17), the Fuzzy model achieved the most favorable overall balance between deposit

capture and prospective-area extent. Its high-potential zones were spatially consistent with the principal metallogenic fault zones and exhibited coherent spatial continuity. The Prediction–Area plots (Figure 18) further demonstrate the joint variation in occurrence capture and mapped area with changing prospectivity thresholds.

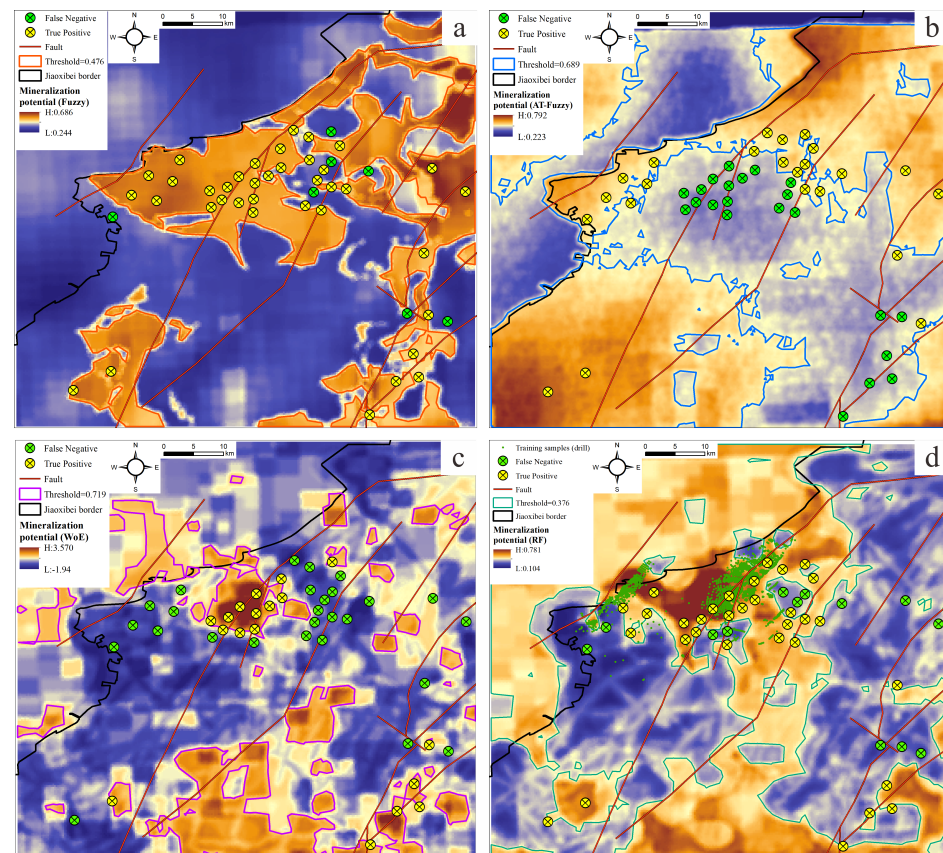


Figure 17. Mineral potential maps of the comparative experimental groups ((a) Fuzzy, (b) AT-Fuzzy, (c) WoE, (d) RF).

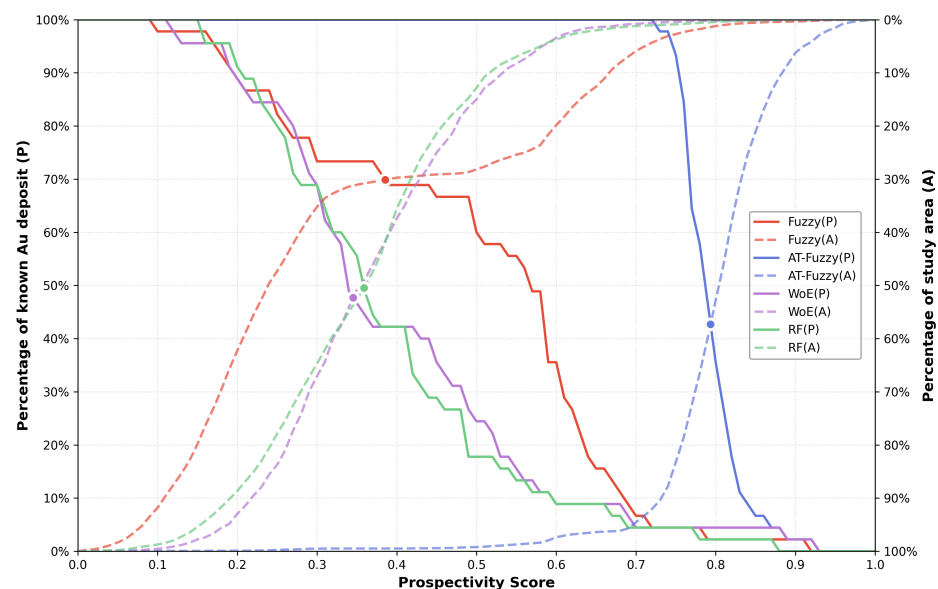


Figure 18. Prediction–Area (P–A) plots of the comparative experimental groups.

The success-rate analysis provides an additional cumulative comparison of spatial-ranking performance over the full range of mapped prospective areas (Figure 19). The

Fuzzy model achieved the highest SR-AUC of 0.762, followed by RF (0.590), WoE (0.540), and AT-Fuzzy (0.386). Although RF and WoE captured relatively more known occurrences within some of the smallest top-ranked area fractions, the Fuzzy curve increased rapidly and remained above the other models over most of the evaluation-area range. This result indicates that the feature-engineering-based Fuzzy model provided the strongest overall concentration of known mineralization within the higher-ranked prospective areas. The SR-AUC results are consistent with the single-threshold comparison, in which the Fuzzy model captured 84.4% of the known occurrences within 25.356% of the valid evaluation area.

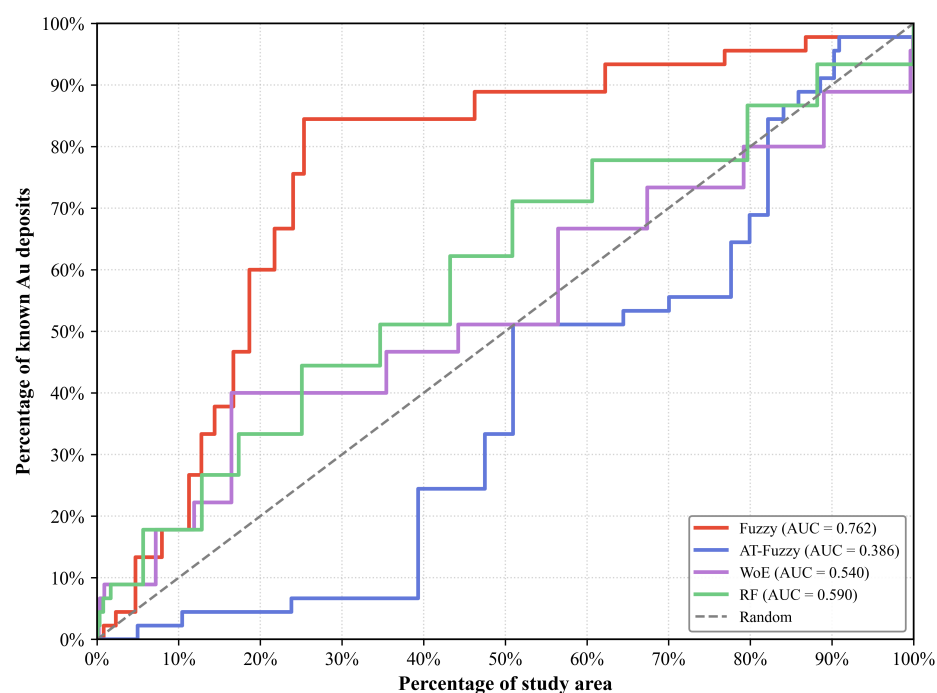


Figure 19. Success-rate curves and corresponding SR-AUC values for the four mineral potential mapping models. The diagonal line represents random spatial ranking.

In contrast, the AT-Fuzzy model delineated an overly extensive high-potential area, covering more than half of the valid evaluation domain (Figure 17b). Its SR-AUC of 0.386, which is below the 0.5 random-ranking reference, further indicates weak cumulative spatial-ranking efficiency when the full initial factor set was retained. These results highlight the role of feature engineering in suppressing redundant geological-background information and emphasizing the principal ore-controlling features, thereby improving prospecting efficiency while retaining a relatively high deposit capture rate.

The resulting nine-factor combination reflects the available data, the relative contributions identified by RF and SHAP, and the metallogenic characteristics of the Jiaoxibei area. Under different data coverage, sample composition, geological settings, or deposit types, the same workflow may yield a different retained subset. Application to other deposit types would require reconstruction of the initial indicator system, membership transformations, and category weights according to the corresponding metallogenic model and available exploration data. Accordingly, the integrated procedure of correlation analysis, RF importance evaluation, SHAP interpretation, geological screening, and category-balanced integration represents the transferable component of the framework, whereas the specific factor composition and parameterization should be re-evaluated for each application.

Notably, the G-SOMML method, by introducing geological prior constraints, partially addresses the limitations of traditional unsupervised geochemical anomaly-detection methods in identifying mineralization-related information. These findings indicate that

knowledge-guided feature extraction is particularly valuable for low-density regional geochemical data.

Among the three comparative experiments, WoE yielded the lowest deposit capture rate (Table 4). The Weights of Evidence method assumes conditional independence among the evidence factors, whereas substantial correlations exist among the structural, geophysical, remote-sensing, and geochemical information in the Jiaoxibei metallogenic system. These correlations may have reduced the reliability of the combined WoE estimates. The Fuzzy model integrated the correlated evidence through membership transformation and category-balanced weighted overlay and achieved a higher deposit capture rate under the same statistical thresholding criterion.

Although the RF model achieved a relatively high deposit capture rate (Table 4), its predicted area ratio was approximately twice that of the Fuzzy model (Figure 17d). RF can learn nonlinear relationships and feature interactions directly from labeled borehole data, while its held-out testing accuracy and OOB score provide explicit diagnostics of borehole-level classification performance. Its regional predictions, however, are influenced by the representativeness, class balance, and spatial distribution of the training samples. In the present dataset, the positive-to-negative sample ratio was approximately 2.7:1, and the boreholes were concentrated in and around known mining areas, which may have contributed to the relatively extensive high-potential area delineated by RF. The Fuzzy model provides a transparent mechanism for integrating heterogeneous geoscientific evidence through explicit membership transformations and category-balanced weights. Its implementation depends partly on expert geological knowledge during indicator-system construction, geological screening, membership-direction assignment, and category-weight determination. Linear membership functions simplify potentially nonlinear or spatially varying factor responses, and weighted-average aggregation has limited capacity to represent higher-order feature interactions. In addition, feature screening cannot completely eliminate biases associated with the spatial distribution and geological representativeness of the available borehole samples. The spatial evaluation against known gold occurrences conducted in this study therefore primarily reflects spatial correspondence within the Jiaoxibei area, whereas application to geologically different or poorly explored regions requires recalibration using region-specific information and validation against independent exploration data. In addition, the preprocessing and co-registration of multi-source geoscientific data, together with the RF and SHAP analyses, increase the overall data-processing requirements of the complete framework. These methodological characteristics help explain the differences in mapped extent and targeting efficiency observed between RF and Fuzzy under the present data conditions.

Known gold occurrences were used for the spatial evaluation of the mineral potential maps. Because feature selection drew on the borehole-based RF and SHAP analyses and the membership transformations incorporated both statistical responses and geological implications, deposit capture rates were reported separately for the Sanshandao and Jiaojia areas and for the other parts of the study area. The single-threshold statistics, cumulative success-rate analysis, and regional deposit-capture statistics jointly characterize spatial correspondence and relative targeting efficiency under different prospective-area constraints.

The threshold-based statistics and success-rate curves quantify spatial correspondence with known gold occurrences and relative targeting efficiency. Independent prospective validation will require newly discovered mineralization, independent drilling results, or spatially independent validation samples. Most of the available occurrences represent deposits or mineralized sites identified through predominantly shallow exploration; therefore, assessment of the framework for deep or concealed mineralization requires additional drilling evidence and three-dimensional geological constraints. The gravity and mag-

netic derivative features provide indirect information on subsurface physical-property boundaries and structural settings, but their predictive significance for deep mineralization requires further geological verification.

Overall, the feature-engineering-based fuzzy synthetic evaluation framework delineated a relatively compact high-potential area while retaining a comparatively high deposit capture rate. The resulting spatial pattern and cumulative targeting performance support its application to regional mineral potential evaluation and target prioritization in the Jiaoxibei area.

6. Conclusions

This study developed a geologically constrained feature-engineering framework for regional mineral potential evaluation in the Jiaoxibei area. Multi-source structural, remote-sensing, gravity, magnetic, geochemical, and borehole information was reconstructed and screened to establish a geologically interpretable evaluation factor set. The primary conclusions are as follows:

- (1) A multi-source mineralization indicator system comprising 32 initial factors was established to represent structural geometry, surface alteration, subsurface physical-property boundaries, and geochemical anomalies. Correlation-based redundancy analysis retained 20 non-geochemical factors, which, together with the directly retained geochemical factor Chem_F, formed 21 candidate features. Random Forest importance evaluation, SHAP interpretation, and geological screening subsequently identified nine core factors for mineral potential evaluation.
- (2) Feature engineering substantially improved the spatial focusing of the evaluation results. Compared with the AT-Fuzzy model using all available factors, the feature-engineering-based Fuzzy model reduced the high-potential area ratio from 50.951% to 25.356%, while increasing the deposit capture rate from 51.1% to 84.4%. These results indicate that redundancy reduction and feature screening improved the balance between known-occurrence capture and prospective-area extent.
- (3) The nine core factors were integrated through linear membership transformation and category-balanced fuzzy weighted overlay. Under the same statistical anomaly-thresholding criterion, the Fuzzy model achieved a deposit capture rate of 84.4% within 25.356% of the valid evaluation area, compared with 71.1% within 50.875% for RF and 40.0% within 16.457% for WoE. The cumulative success-rate analysis yielded an SR-AUC of 0.762 for Fuzzy, compared with 0.590 for RF, 0.540 for WoE, and 0.386 for AT-Fuzzy. The Fuzzy result therefore achieved the highest cumulative spatial-ranking efficiency among the four mineral potential maps under the present evaluation conditions.
- (4) Spatial evaluation showed that 21 of the 22 known gold occurrences in the Sanshandao and Jiaojia areas were located within the high-potential zones, corresponding to a deposit capture rate of 95.5%. In the other parts of the study area, 17 of the 23 known occurrences were captured, corresponding to a deposit capture rate of 73.9%. Overall, 38 of the 45 known occurrences were captured, yielding an overall deposit capture rate of 84.4%. The regional difference in deposit capture indicates that the model achieved stronger spatial correspondence in the borehole-concentrated Sanshandao and Jiaojia areas. Further evaluation of deep or concealed mineralization requires independent drilling results and three-dimensional geological constraints.

Author Contributions: Conceptualization, Y.F.; methodology, Y.F.; validation, Y.C.; formal analysis, Y.F. and Y.C.; investigation, Y.C.; resources, Y.W. and Z.J.; data curation, S.W. and Z.A.; writing—original draft preparation, Y.F.; writing—review and editing, Y.F.; visualization, Y.F. and J.Z.; supervision, Y.W. and Z.J.; project administration, Y.W.; funding acquisition, Y.W. and Z.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key R & D Program of China (2023YFC2906903, 2023YFC2907105), the Science and Technology Research Project of Shandong Provincial Bureau of Geology and Mineral Resources (KY202502), the Key Project of the National Natural Science Foundation of China (42230810) and the Key Science & Technology Support Project of Ministry of Natural Resources of China (ZKKJ202419).

Data Availability Statement: The datasets for the current study are available from the corresponding author on reasonable request.

Acknowledgments: The authors would like to thank the anonymous reviewers for their constructive comments and suggestions that improved the quality of this manuscript.

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

References

- Deng, J.; Wang, Q.; Zhang, L.; Xue, S.; Liu, X.; Yang, L.; Yang, L.; Qiu, K.; Liang, Y. Metallogenic model of Jiaodong-type gold deposits, eastern China. *Sci. China Earth Sci.* **2023**, *66*, 2287–2310. [\[CrossRef\]](#)
- Song, M.C.; Li, S.Z.; Santosh, M.; Zhao, S.; Yu, S.; Yi, P.H.; Cui, S.X.; Lv, G.X.; Xu, J.X.; Song, Y.X.; et al. Types, characteristics and metallogenesis of gold deposits in the Jiaodong Peninsula, Eastern North China Craton. *Ore Geol. Rev.* **2015**, *65*, 612–625. [\[CrossRef\]](#)
- Tian, Z.; Ding, Z.; MacKenzie, D.; Chen, G.; Zhang, Q.; Liu, F.; Xiao, W. Structural Control of Gold Mineralization in the Jiaodong Peninsula, Eastern China. *J. Struct. Geol.* **2026**, *206*, 105654. [\[CrossRef\]](#)
- Deng, J.; Yang, L.Q.; Li, R.H.; Groves, D.I.; Santosh, M.; Wang, Z.L.; Sai, S.X.; Wang, S.R. Regional structural control on the distribution of world-class gold deposits: An overview from the Giant Jiaodong Gold Province, China. *Geol. J.* **2019**, *54*, 378–391.
- Li, X.H.; Fan, H.R.; Luo, J.; Zhang, Y.W.; Zeng, Q.D.; Bai, R.; Wu, J.J.; Zhang, Z.M.; Hu, F.F. Characteristics of deep ore-forming fluids in the Jiaojia gold belt, Jiaodong gold province (China). *Ore Geol. Rev.* **2025**, *186*, 106832. [\[CrossRef\]](#)
- Song, M.; Wang, B.; Song, Y.; Li, J.; Zheng, J.; Li, S.; Fan, J.; Yang, Z.; He, C.; Gao, M.; et al. Spatial coupling relationship between faults and gold deposits in the Jiaodong ore concentration area and the effect of thermal doming-extension on mineralisation. *Ore Geol. Rev.* **2023**, *153*, 105277. [\[CrossRef\]](#)
- Chen, G.; Liu, Z.; Chen, G.; Xie, S.; Yang, X.; Li, X.; Chen, Y.; Hao, Z.; Zhong, H.; Jiang, L. Deep gold prospectivity modeling in the Jiaojia gold belt, Jiaodong Peninsula, eastern China using machine learning of geometric and geodynamic variables. *Front. Earth Sci.* **2024**, *12*, 1308426. [\[CrossRef\]](#)
- Li, B.; Peng, Y.; Zhao, X.; Liu, X.; Wang, G.; Jiang, H.; Wang, H.; Yang, Z. Combining 3D geological modeling and 3D spectral modeling for deep mineral exploration in the Zhaoxian gold Deposit, Shandong Province, China. *Minerals* **2022**, *12*, 1272. [\[CrossRef\]](#)
- Kuang, H.; Yan, J.; Zhang, K.; Tang, W.; Fu, C.; Liang, J.; Zhang, G.; You, Y. Integrated geophysical signatures of the Jiaodong region in China and their implications for deep architecture and gold metallogenic systems. *Minerals* **2025**, *15*, 417. [\[CrossRef\]](#)
- Liu, X.; Wang, G.; Lv, J.; Peng, Y.; Zhao, X. Three-dimensional modeling and geostatistical structural analysis for ore deposit prospecting trend and wisdom mining: A case study in the Jiaoxibei gold field, China. *Front. Earth Sci.* **2024**, *12*, 1217016. [\[CrossRef\]](#)
- Na, Y.; Yang, Y. Study on Genesis of Gold Deposit and Mineralization Enrichment Regularity Based on Mapping Analysis. In *Proceedings of the 2017 9th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*; IEEE: New York, NY, USA, 2017; pp. 236–239.
- Song, M.; Li, S.; Zheng, J.; Wang, B.; Fan, J.; Yang, Z.; Wen, G.; Liu, H.; He, C.; Zhang, L.; et al. A 3D Predictive Method for Deep-Seated Gold Deposits in the Northwest Jiaodong Peninsula and Predicted Results of Main Metallogenic Belts. *Minerals* **2022**, *12*, 935. [\[CrossRef\]](#)
- Xu, J.; Li, Y.; Liu, W.; Han, S.; Tan, K.; Xie, Y.; Zhao, Y. Research on Multi-Source Information-Based Mineral Prospecting Prediction Using Machine Learning. *Minerals* **2025**, *15*, 1046. [\[CrossRef\]](#)
- Abedini, M.; Ziaii, M.; Timkin, T.; Pour, A.B. Machine learning (ML)-based copper mineralization prospectivity mapping (MPM) using mining geochemistry method and remote sensing satellite data. *Remote Sens.* **2023**, *15*, 3708. [\[CrossRef\]](#)

15. Ghane, B.; Lentz, D.R.; Thorne, K.G.; Ugalde, H.A. Calibration of Airborne Geophysical Data with In Situ Petrophysical Measurements for Mineral Prospectivity Mapping Using XGBoost Random Forest. *Nat. Resour. Res.* **2026**, *35*, 245–277.
16. Liu, L.; Lu, J.; Tao, C.; Liao, S.; Su, C.; Huang, N.; Xu, X. Fuzzy forest machine learning predictive model for mineral prospectivity: A case study on Southwest Indian Ridge 48.7°E–50.5°E. *Nat. Resour. Res.* **2022**, *31*, 99–116. [[CrossRef](#)]
17. Fu, Z.; Zheng, X.; Yan, Y.; Xu, X.; Zhou, F.; Li, X.; Zhou, Q.; Mai, W. The Evolution of Machine Learning in Large-Scale Mineral Prospectivity Prediction: A Decade of Innovation (2016–2025). *Minerals* **2025**, *15*, 1042. [[CrossRef](#)]
18. Liu, Y.; Sun, T.; Wu, K.; Xiang, W.; Zhang, J.; Zhang, H.; Feng, M. Interpretability analysis of data augmented convolutional neural network in mineral prospectivity mapping using black-box visualization tools. *Nat. Resour. Res.* **2025**, *34*, 759–783. [[CrossRef](#)]
19. Mou, N.; Carranza, E.J.M.; Wang, G.; Sun, X. A Framework for Data-Driven Mineral Prospectivity Mapping with Interpretable Machine Learning and Modulated Predictive Modeling. *Nat. Resour. Res.* **2023**, *32*, 2439–2462. [[CrossRef](#)]
20. Liu, H.; Harris, J.; Sherlock, R.; Behnia, P.; Grunsky, E.; Naghizadeh, M.; Rubingh, K.; Tuba, G.; Roots, E.; Hill, G. Mineral prospectivity mapping using machine learning techniques for gold exploration in the Larder Lake area, Ontario, Canada. *J. Geochem. Explor.* **2023**, *253*, 107279. [[CrossRef](#)]
21. Wang, L.; Yang, J.; Wu, S.; Hu, L.; Ge, Y.; Du, Z. Enhancing mineral prospectivity mapping with geospatial artificial intelligence: A geographically neural network-weighted logistic regression approach. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *128*, 103746. [[CrossRef](#)]
22. Li, C.; Yan, J. Tectono-magmatic control on the intensity of decratonic gold mineralization in the southeastern margin of the North China Craton: A perspective from geochemical comparison. *Int. Geol. Rev.* **2024**, *66*, 582–606.
23. Liu, Z.J.; Yang, L.Q.; Xie, D.; Yang, W.; Li, D.P.; Feng, T.; Deng, J. Hydrothermal Alteration Processes of Xincheng Gold Deposit Jiaodong Peninsula, China: Constraints from Composition of Hydrothermal Rutile. *Minerals* **2024**, *14*, 417. [[CrossRef](#)]
24. Ni, J.L.; Liu, J.L.; Wang, Y.R.; Xia, Z.M.; Ji, L.; Li, G.Y. Control of the shift in subduction direction on upper plate tectonic extension and gold mineralization: A case study from the early cretaceous Jiaodong Peninsula, eastern China. *Tectonics* **2025**, *44*, e2024TC008585. [[CrossRef](#)]
25. Glowacki, T.; Bortnowski, P. Long-term prediction of post-mining land deformation with XGBoost, Bayesian optimization, and time series feature engineering. *Measurement* **2025**, *256*, 119630.
26. Zhang, H.; Xie, M.; Dan, S.; Li, M.; Li, Y.; Yang, D.; Wang, Y. Optimization of feature selection in mineral prospectivity using ensemble learning. *Minerals* **2024**, *14*, 970. [[CrossRef](#)]
27. Zuo, R.; Cheng, Q.; Xu, Y.; Yang, F.; Xiong, Y.; Wang, Z.; Kreuzer, O.P. Explainable artificial intelligence models for mineral prospectivity mapping. *Sci. China Earth Sci.* **2024**, *67*, 2864–2875. [[CrossRef](#)]
28. Zuo, R.; Yang, F.; Cheng, Q.; Kreuzer, O.P. A novel data-knowledge dual-driven model coupling artificial intelligence with a mineral systems approach for mineral prospectivity mapping. *Geology* **2025**, *53*, 284–288. [[CrossRef](#)]
29. Li, Q.; Chen, G.; Luo, L. Mineral prospectivity mapping using attention-based convolutional neural network. *Ore Geol. Rev.* **2023**, *156*, 105381. [[CrossRef](#)]
30. Mahboob, M.A.; Celik, T.; Genc, B. Predictive modelling of mineral prospectivity using satellite remote sensing and machine learning algorithms. *Remote Sens. Appl. Soc. Environ.* **2024**, *36*, 101316. [[CrossRef](#)]
31. Feng, Y.; Wang, Y.; Wang, C.; Tian, J.; Zeng, Z.; Cheng, Y.; Wen, S. Automated Geochemical Anomaly Detection Using Self-Organizing Map and Metric Learning. *Nat. Resour. Res.* **2025**, *34*, 2915–2932. [[CrossRef](#)]
32. Feng, Y.; Wang, Y.; Wang, C.; Tian, J.; Wen, S.; Zhou, Y.; Cheng, Y. A fuzzy gold mineral prediction model integrating with knowledge-driven and data-driven: A case study of the Hatu region in Xinjiang, China. *Minerals* **2024**, *14*, 1209. [[CrossRef](#)]
33. Wang, B.; Zhou, J.; Ding, Z.; Zhao, T.; Song, M.; Bao, Z.; Lyu, J.; Xu, S.; Yan, C.; Liu, X.; et al. Gold mineralization in Jiaodong Peninsula and destruction of North China Craton: Insights from Mesozoic granite. *Glob. Geol.* **2023**, *26*, 98–113.
34. Wang, B.; Zhou, J.B.; Ding, Z.J.; Wilde, S.A.; Song, M.C.; Zhao, T.Q.; Bao, Z.Y. Late Mesozoic magmatism and gold metallogeny of the Jiaodong Peninsula, China: A response to the destruction of the North China Craton. *Bulletin* **2024**, *136*, 1395–1412. [[CrossRef](#)]
35. Liu, H.; Yin, Y.; Xiao, P.; Peng, K. Modeling of in situ stress field distribution near metallogenic faults and analysis of the interaction mechanism between faults and stress field: A case from the northwestern Jiaodong Peninsula, China. *Min. Metall. Explor.* **2025**, *42*, 235–253. [[CrossRef](#)]
36. Wang, J.; Zhu, W.; Liu, X.; Guo, J.; Yan, J. Multi-scale geological modeling and in-situ stress inversion of Xincheng Gold Mine at the Jiaodong Peninsula, China. *Bull. Eng. Geol. Environ.* **2024**, *83*, 492. [[CrossRef](#)]
37. Qiu, Y.; Groves, D.I.; McNaughton, N.J.; Wang, L.G.; Zhou, T. Nature, age, and tectonic setting of granitoid-hosted, orogenic gold deposits of the Jiaodong Peninsula, eastern North China craton, China. *Miner. Depos.* **2002**, *37*, 283–305. [[CrossRef](#)]
38. Deng, J.; Wang, C.; Bagas, L.; Carranza, E.J.M.; Lu, Y. Cretaceous–Cenozoic tectonic history of the Jiaojia Fault and gold mineralization in the Jiaodong Peninsula, China: Constraints from zircon U–Pb, illite K–Ar, and apatite fission track thermochronometry. *Miner. Depos.* **2015**, *50*, 987–1006. [[CrossRef](#)]
39. Goldfarb, R.; Qiu, K. The Jiaodong gold deposits of the North China block—A possible Sulu terrane source. *Aust. J. Earth Sci.* **2025**, *72*, 795–811. [[CrossRef](#)]

40. Wang, H.; Liang, Y.; Liu, X.; Wang, Q.; Ma, Y.; Xue, W.; Fang, Y.; Xue, S.; Deng, J. Heterogeneity of the late Mesozoic lithospheric mantle beneath the Jiaodong Peninsula, eastern North China Craton: A synthesis from geochemistry and Sr-Nd-Pb isotopes of mafic dykes. *Int. Geol. Rev.* **2024**, *66*, 1765–1793. [[CrossRef](#)]
41. Deng, J.; Wang, Q.; Liu, X.; Zhang, L.; Yang, L.; Yang, L.; Qiu, K.; Guo, L.; Liang, Y.; Ma, Y. The formation of the Jiaodong gold province. *Acta Geol. Sin.-Engl. Ed.* **2022**, *96*, 1801–1820. [[CrossRef](#)]
42. Feng, Y.; Wang, Y.; Ding, Z.; Wang, B.; He, Y.; An, Z.; Liu, D. The ore-forming model and evolution of prospecting techniques for gold deposits in Jiaoxibei. *Earth Sci. Front.* **2025**, *32*, 165–181.
43. McCuaig, T.C.; Beresford, S.; Hronsky, J. Translating the mineral systems approach into an effective exploration targeting system. *Ore Geol. Rev.* **2010**, *38*, 128–138. [[CrossRef](#)]
44. Karpfinger, C. Fourier series: Calculation of fourier coefficients. In *Calculus and Linear Algebra in Recipes: Terms, Phrases and Numerous Examples in Short Learning Units*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 813–827.
45. Oliveira, S.P.; Porsani, M.J.; Fries, M.; Zago, M.M. Phase Transformation of the Analytic Signal for Enhancing the Resolution of Potential Field Data. *Minerals* **2025**, *15*, 1266. [[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.