https://doi.org/10.5194/egusphere-2026-2616
Preprint. Discussion started: 16 June 2026

Considering mineralization local-global geological features: An interpretable DCN-Transformer hybrid model with attribution for mineral prospectivity mapping

Yunfei Hao, Yihui Xiong

State Key Laboratory of Geological Processes and Mineral Resources, China University of Geosciences, Wuhan 430074, China

Corresponding author. Email: xiongyh426@cug.edu.cn

Abstract. Mineral prospectivity mapping is a critical task in mineral exploration, effectively integrating which demands models capable of capturing both local and global geological features. While deep learning models excel in this domain, their "black-box" nature often limits the trust and insights geologists can derive from their predictions. This paper introduces a novel and interpretable hybrid model to explicitly address this challenge. Our architecture synergistically combines a deformable convolutional network for adapting spatially varying local mineralization features, such as geochemical anomalies, a Transformer module for modelling long-range global-scale spatial features governing mineral deposition. A pivotal innovation is the incorporation of an attribution branching network that generates significance scores for each input predictive factor to the final prospectivity probability. These scores not only provide a direct interpretation of factor relevance but are also fed back to dynamically modulate the key values in the Transformer's attention mechanism, effectively injecting prior geological knowledge into the local and global feature learning process. This design fosters a more geologically informed integration of local and global representations. The model's performance is evaluated against benchmark models including standalone deformable convolutional network, Transformer, and a hybrid model with deformable convolutional network and Transformer. Results demonstrate a superior predictive accuracy and more geologically plausible prospectivity maps. Furthermore, we provide a multi-faceted interpretation framework: the attribution branching network quantifies the contribution of each evidence layer, while gradient-weighted class activation mapping visualizes the discriminative local regions highlighted by the convolutional components and attention maps reveal the long-range feature relationships prioritized by the Transformer to elucidate how the model hierarchically integrates local and global features to arrive at its decisions. This dual-path interpretation strategy demystifies the model's decision-making process, offering geologists tangible insights into both "where" and "why" the model identifies high-potential zones, thereby bridging the gap between high-performing deep learning and actionable c intelligence.

https://doi.org/10.5194/egusphere-2026-2616
Preprint. Discussion started: 16 June 2026

1. Introduction

Mineral Prospectivity Mapping (MPM) is a pivotal geoscientific task that integrates diverse geological, geochemical, geophysical, and remote sensing to delineate areas with high potential for hosting mineral deposits of a specific type (Bonham-Carter, 1994). The goal is to reduce exploration risk and cost by prioritizing favourable ore-controlling factors. The evolution of MPM methodologies has progressed from knowledge-driven approaches, such as fuzzy logic and expert systems, to data-driven techniques, primarily employing machine learning (ML) models (Porwal et al., 2003; Carranza, 2008). Among these, deep learning (DL), with their superior capacity for automatically learning complex, high-order, non-linear relationships from multi-source geoscience data, have demonstrated remarkable predictive performance, often surpassing traditional machine learning methods (Xiong et al., 2018; Zuo et al., 2022; Zuo and Xu, 2023; Yang et al., 2024).

The conceptual foundation of MPM rests on the "mineral system" model, which posits that ore formation is the result of synergistic, multi-scale geological processes that create both localized ore bodies (proximal features) and extensive, indicative alteration halos (distal features) (Wyborn et al., 1994; McCuaig et al., 2010). Effectively capturing and integrating these multi-scale spatial patterns, from the local features of alteration and structure to the global geological features, is the central challenge in modern data-driven MPM. Local features are often related to the immediate ore-forming environment, such as specific lithological contacts or local geochemical anomalies. These features are typically confined and have a high spatial frequency. In contrast, global features pertain to the broader geological context and tectonic framework that create a permissive environment for mineralization, such as large-scale fault systems. These features are extensive and exhibit low spatial frequency. A robust MPM model must, therefore, be capable of effectively integrating both local and global spatial contexts to accurately model the multi-scale nature of mineralization controls (Li et al., 2024; Yu et al., 2024).

Traditional machine learning (ML) methods, such as random forests and support vector machines, have been widely adopted for MPM due to their proficiency in handling high-dimensional data. However, their capacity to model spatial relationships is inherently limited. These models typically treat each pixel or grid cell independently, failing to explicitly capture the contextual spatial patterns-both local and global-that are the very essence of mineral deposit genesis (Zuo and Xu, 2023). While traditional methods often struggled to explicitly model this hierarchy, the advent of DL has provided powerful tools, yet each family of architectures has historically excelled in only one domain, creating a compelling rationale for the hybrid models that synergize their strengths. This necessity to capture multi-scale features has driven the adoption of specific DL architectures (Zuo and Xu, 2023).

https://doi.org/10.5194/egusphere-2026-2616
Preprint. Discussion started: 16 June 2026

Convolutional Neural Networks (CNN) have emerged as the dominant tool for modeling local features (Li et al., 2021). Their intrinsic design, utilizing localized filters and hierarchical stacking, is exceptionally well-suited for extracting spatially local patterns. Several studies have successfully applied CNN to MPM, demonstrating their prowess in identifying patterns from geological and geophysical data (Li et al., 2021; McMillan et al., 2021; Zhang et al., 2021; Yang et al., 2022; Li et al., 2023; Yang and Zuo, 2024). However, traditional CNN suffer from a critical limitation stemming from fixed regular convolution kernels, which fail to adequately model the anisotropic patterns of geochemical distributions (Dai et al., 2017). In contrast, GNN can directly represent non-Euclidean spatial relationships through nodes of geochemical samples and edges reflecting spatial proximity or geological associations, facilitating multi-source geoscience learning via complex neighborhood interactions (Zuo and Xu, 2023; Sihombing et al., 2024; Shi et al., 2024). Even so, GNN relies heavily on high-quality data and professional domain expertise to construct valid graph structures; unreasonable edge definition may trigger noise interference, hide true geochemical anomalies, and reduce model accuracy (Gong and Cheng, 2019; Zhou et al., 2020). As an effective alternative, Deformable Convolutional Networks (DCN) break the fixed geometric restriction of conventional CNN kernels (Dai et al., 2017). With learnable spatial offsets, DCN adaptively adjusts sampling positions and warps receptive fields to match irregular, complex spatial structures. It delivers superior robustness to changes in orientation, scale and deformation, achieving precise feature extraction for irregular local geospatial characteristics (Dai et al., 2017; Zhu et al., 2018). By capturing subtle spatial deformations from inherently irregular geoscientific data, DCN is therefore adopted in the present study (Zhang et al., 2026).

However, the above deep learning architecture which proficiency with local features also imposes a fundamental limitation: their receptive field is inherently local. They struggle to capture long-range dependencies and the global geological structures that are equally critical for mineral prospectivity (Yu et al., 2024). A mineral system operates at a basin or crustal scale, where the spatial relationship between fluid source, large-scale fault acting as a fluid pathway, and local trap site is paramount (McCuaig et al., 2010). The above deep learning architecture (CNN, GNN, and DCN) analyzing a small patch of data may identify a favourable local structure but fail to recognize that this structure is situated within a much larger, regionally significant global tectonic zone. Therefore, an MPM model that excels only at local feature extraction provides an incomplete picture of mineralizing systems.

To address the challenge of capturing global context, the Transformer architecture, originally developed for natural language processing, has revolutionized sequence modeling with its self-attention mechanism (Vaswani et al., 2017). When adapted for computer vision as Vision Transformers (ViTs), this mechanism allows each patch of an image to interact directly with every other patch, irrespective of distance, thereby enabling the direct modeling of global contextual information (Dosovitskiy et al., 2020). In the context of MPM, this translates to a model that can, for example, directly

https://doi.org/10.5194/egusphere-2026-2616
Preprint. Discussion started: 16 June 2026

learn the association between a specific geochemical signature in one part of a region and a structural feature hundreds of kilometers away, provided they are part of the same mineralizing system (Li et al., 2024; Yu et al., 2024). While the application of pure Transformers in MPM is still nascent, their theoretical potential for capturing the global scale of mineral systems is undeniable.

The complementary strengths and weaknesses of DCN and Transformers present a compelling case for a hybrid architecture. The local method can serve as a powerful feature engine, extracting rich, multi-level local features—from edges and textures to more complex motifs—from the input geo-data. The Transformer can then act as a global reasoning engine, analyzing the relationships between these DCN-derived features across the entire map. This DCN-Transformer hybrid paradigm can synergistically leverage the DCN's excellence in extracting hierarchical local features and the Transformer's supremacy in modeling global dependencies. This hybrid idea is gaining traction in high-level feature extraction and is beginning to permeate geoscientific applications, demonstrating superior performance over standalone models (Li et al., 2024; Zhao et al., 2024; Hou et al., 2025; Song et al., 2025). We posit that this architecture is ideally suited for MPM, where the DCN component can act as a local geologist identifying detailed ore-controlling features, while the Transformer component serves as a regional tectonic expert understanding the broader geological framework that makes a district prospective. This hybrid design directly addresses the multi-scale nature of mineralization processes. Despite their superior predictive performance, DL models, including hybrids, are often criticized as black box, making their decision-making processes opaque and difficult to trust (Samek et al., 2017; Li et al., 2022). The inability to understand why a model flags a specific area as prospective hinders its trustworthiness and practical utility for decision-making (Zuo et al., 2023, 2024). Since the application of DL in MPM is not merely a quest for higher prediction accuracy; it is increasingly recognized that the geological intelligibility and interpretability of these black box models are equally, if not more, critical for gaining the trust of geologists and providing actionable insights for exploration targeting (Xu and Zuo, 2024; Zuo et al., 2024; Zuo et al., 2025).

The field of Explainable AI (XAI) offers various techniques to illuminate the decision-making processes of DL models. These can be broadly categorized into post-hoc methods (applied after training) and ante-hoc (intrinsically interpretable models) (Samek et al., 2021). The ante-hoc model interpretability represents the inherent interpretability within the models, which can be implemented by the theoretical rules and constraints used in building and training models (Zuo et al., 2024). The core premise of informed modeling synthesizes observational data and domain priors (physical/chemical laws) to learn feature representations. By embedding this knowledge to direct training and inference processes, DL models gain improved accuracy and interpretability (Von Rueden et al., 2021). A notable example is the geologically-constrained neural network, which integrates geological constraints into the training of neural networks to approximate certain

https://doi.org/10.5194/egusphere-2026-2616
Preprint. Discussion started: 16 June 2026

geological regulations (Zuo et al., 2022). These constraints are often implemented as different loss functions to guide the training of neural networks, where the decision-making process of the trained neural networks is potentially limited by the corresponding geological regulations (Zuo et al., 2022; Xu and Zuo, 2024; Yang et al., 2024; Sabbaghi et al., 2024). Post-hoc model interpretability primarily involves techniques applied after model training is complete (Carvalho et al., 2019; Luo et al., 2023; Zuo et al., 2024). Attribution methods (i.e., SHapley Additive exPlanations (SHAP) framework, gradient-weighted Class Activation Mapping (Grad-CAM), backpropagation) that assign an importance score to each input feature concerning the model's final prediction, are widely adopted post-hoc interpretation methods to DL models. SHAP leverages Shapley values to quantify the contribution of individual ore-controlling factors to model prediction (Pradhan et al., 2022; Luo et al., 2023; Chen et al., 2024). However, SHAP probe the model by perturbing input features and observing output changes. Though model-agnostic, they can be computationally expensive and may produce unstable explanations. Conversely, CAM and grad-CAM assign contribution weights to features within the network and generates a heatmap via weighted linear combination, thereby explaining the influence of each factor on the mapping of mineral prospectivity (Yang et al., 2023). In contrast to previous two types of methods, DeepLIFT (Deep Learning Important FeaTures) offers a powerful and computationally efficient alternative (Shrikumar et al., 2017). DeepLIFT decomposes the output prediction by comparing the activation of each neuron to a 'reference' activation, effectively tracing the contribution of each input feature through a series of backpropagation-like rules. This allows it to assign contribution scores without relying on gradients, which can vanish or explode (Shrikumar et al., 2017). DeepLIFT provides a clear, pixel-wise attribution map that indicates which parts of the input geospatial data were most critical for the model's "prospective" or "non-prospective" classification.

The rationale for integrating DeepLIFT with DCN-Transformer hybrid for MPM is a novel and critical step by discovering the key contributing ore-controlling factors of specific known mineral deposits. This study proposes a hybrid interpretable model with DCN, DeepLIFT, and Transformer. The hybrid model first attempt to extract local feature of ore-controlling factors based on DCN branching network, then an interpretation branch network driven by DeepLIFT computes the significance scores for each contributing factor in the feedforward fully connected network, finally a classification network with the Transformer component, illuminate the global reasoning by showing which long-range interactions between different parts of the map the model deemed important. The DeepLIFT method can be effectively applied to calculate significance scores for key ore-controlling factors, thereby enhancing both post-hoc and ante-hoc interpretability of the model. On one hand, these scores allow for the ranking and visualization of the relative significance of each geological factor in controlling mineralization. This provides clear, post-hoc insights into which features most influence the model’s predictions, helping geoscientists validate and interpret the output. On the other hand, the importance scores

https://doi.org/10.5194/egusphere-2026-2616
Preprint. Discussion started: 16 June 2026

can be utilized to update the key values in a transformer architecture, effectively embedding domain knowledge directly into the model. By adjusting attention mechanisms based on geological prior knowledge, the model's internal reasoning process becomes more aligned with expert understanding, thereby improving its ante-hoc interpretability. Together, these approaches not only make the model's decisions more transparent after prediction but also guide its learning process toward geologically meaningful patterns, ultimately increasing both trust and performance in mineral prospectivity mapping.

In light of the above, this study proposes an interpretable DCN-Transformer hybrid model integrated with DeepLIFT attribution (called DCN-attribution-Transformer) for MPM. The primary objectives are:

1. To design and implement a hybrid DL architecture that effectively leverages DCN for local spatial feature extraction and Transformers for modeling long-range, global dependencies within multi-source geoscience data.
2. To rigorously evaluate the predictive performance of the proposed hybrid model against standalone DCN and Transformer benchmarks in a case study of lithium polymetallic deposits in Nanling region.
3. To employ the DeepLIFT attribution method to bridges post-hoc and ante-hoc interpretability, where the former generates spatially interpretable maps, providing a transparent explanation of the model's prediction and linking them directly to local and global geological features; the latter dynamically modulates the key values within the Transformer's attention mechanism, thereby injecting prior geological knowledge into the global feature learning process.

By achieving these objectives, this research aims to advance the field of MPM not only by improving predictive accuracy through a more complete representation of multi-scale features but also by bridging the critical interpretability gap, fostering greater confidence and deeper geological insight among exploration geoscientists.

2. Geological setting and Datasets

2.1. Geological setting

This study focuses on the polymetallic and critical metal mineralization belts in the Nanling region of South China (Fig. 1). The Nanling area represents one of the major concentrations of lithium, beryllium, niobium, and tantalum deposits in China and is also among the few regions worldwide characterized by a high density of granite–pegmatite-type rare metal deposits (Hu et al., 2005; Qiao et al., 2015). The region has undergone multiple phases of tectonic evolution and magmatic activity, resulting in a complex geological structural framework and diverse mineralization systems (Chen and Jahn, 1998). The tectonic evolution of this region has been influenced by multiple orogenic events, notably including the Jinning,

https://doi.org/10.5194/egusphere-2026-2616
Preprint. Discussion started: 16 June 2026

Caledonian, Indosinian-Hercynian, Yanshanian, and Himalayan orogenies. During the Mesozoic era, the tectonic regime underwent a significant transition from the Tethyan tectonic system to the Paleo-Pacific tectonic system, a shift that profoundly impacted magmatic activity and mineralization processes. In the Yanshanian period, a northeast-trending fault system developed, superimposed upon the northwest- and east-west-trending faults inherited from the Indosinian phase (Wang et al., 2011). This northeast-oriented fault system served as the principal conduit for magmatic intrusions and controlled the spatial distribution of granitic bodies as well as polymetallic mineral deposits, including tungsten, tin, and other rare metals, within the Nanling and South China regions. During the Indosinian period, the Nanling region was predominantly influenced by the Paleotethys tectonic domain. With the progressive closure of the Paleotethys Ocean, the area underwent intense continental collision, resulting in a series of tectonic deformations and magmatic activities (Hua et al., 2005). Within this tectonic framework, multiple metamorphic domes and associated fold structures developed in the Nanling region. Rare metal mineralization is frequently spatially associated with structural sites such as the margins of metamorphic domes or the cores of anticlines within fold systems.

Regarding the timing of magmatic events and their relationship to rare metal mineralization, two principal mineralized granite-forming episodes can be identified in the Nanling area. Indosinian granites typically formed in collisional and post-collisional tectonic settings (Chen and Jahn, 1998) and serve as the primary protoliths for several characteristic lithium-beryllium-niobium-tantalum deposits in the region, such as the Touping lithium-beryllium-niobium-tantalum deposit in Guangchang County, Jiangxi Province (~246 Ma; Che et al., 2019), and the lithium-beryllium-niobium-tantalum deposits in Heyuan, Guangdong Province (~200 Ma; Hu et al., 2005). In contrast, magmatic activity during the Yanshanian period was more temporally concentrated and extensive, predominantly occurring between 165 and 155 Ma. This phase produced widespread large granite bodies (Li et al., 2007) that constitute significant ore-bearing lithologies for lithium polymetallic deposits in the Nanling region, with the Jianfengling lithium-rubidium-niobium-tantalum deposit in Hunan Province (~158 Ma; Wen et al., 2021) being the most representative example.

https://doi.org/10.5194/egusphere-2026-2616
Preprint. Discussion started: 16 June 2026

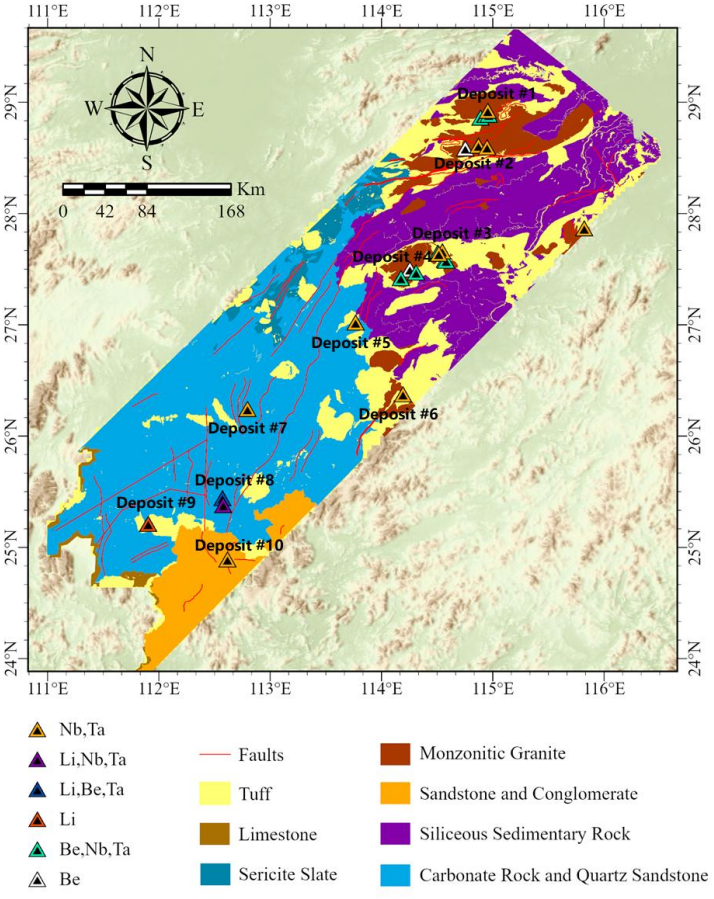


Figure 1. Simplified geological setting of the Nanling region in Southern China (modified from Zhou et al., 2006)

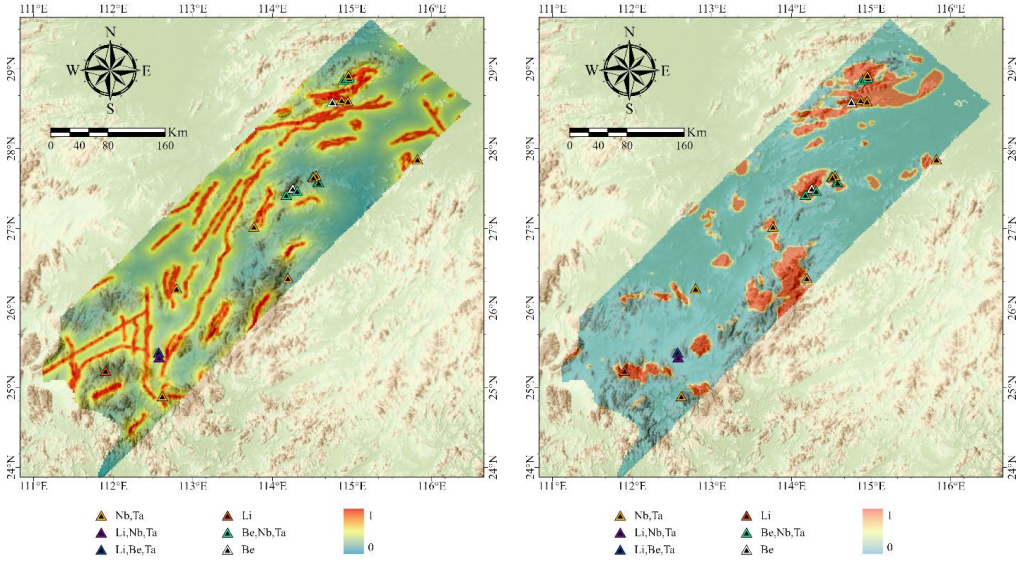
2.2. Datasets

Previous geological data (Wang et al., 2020) indicate that: (1) the mineralization of granite-type lithium-beryllium-niobium-tantalum deposits in the Nanling region predominantly occurred during the peak phase of magmatic evolution in the Yanshanian period; (2) the ore bodies are strictly controlled by NE-NNE trending faults and metamorphic dome structures, typically occurring at stress-relieved zones such as the edges of domes or the cores of anticlines; and (3) highly evolved S-type granites intruded into the Sinian-Cambrian basement strata, providing essential source materials and dynamic conditions for the enrichment of rare metals. Based on these approaches, mineralization-related granites, faults, and ore-controlling strata were selected as input data for the construction of the mineral prospectivity map. A weight map was generated, incorporating buffer zone analysis and Euclidean distance analysis. The power-law function employed

https://doi.org/10.5194/egusphere-2026-2616
Preprint. Discussion started: 16 June 2026

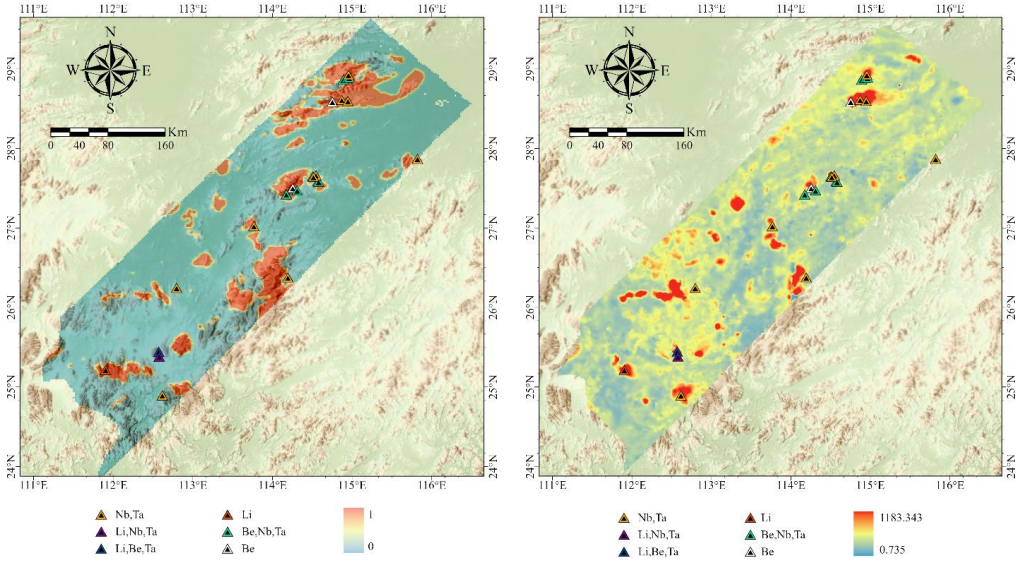
corresponds to that proposed by Zuo (2016).

Besides, geochemical samples at a 1:200,000 scale were obtained from the National Geochemical Mapping Project of China (Xie et al., 1997). A standardized dataset covering the study area was acquired and preprocessed, comprising a total of 32 elements and 7 oxides: Bi, Cu, P, La, Li, Ag, Sn, Au, Mo, Th, U, Y, W, Sb, Hg, Mn, Cr, Sr, Nb, Pb, Ni, Ti, Cd, Co, Ba, Be, V, Zn, B, As, Zr, F, as well as Fe_2O_3, K_2O, CaO, MgO, Na_2O, Al_2O_3, and SiO_2. The concentration data of 39 geochemical elements collected were visualized to generate raster datasets. Subsequently, each of the 39 elements was interpolated onto a 1×1km grid using the inverse distance weighting method. This procedure aimed to transform discrete measurement point data into continuous raster data. All the evidence layers are illustrated in Figure 2.



(a) faults **(b) Yanshanian rocks**

https://doi.org/10.5194/egusphere-2026-2616
Preprint. Discussion started: 16 June 2026



(c) ore-controlling strata (d) Li element

Figure 2. Geological evidence layers including distances to (a) faults, (b) Yanshanian rocks, (c) ore-controlling strata, and (d) Geochemical patterns (Only Li element is shown here)

3. Methods

The proposed framework consists of four principal components: an attribution network employing reference baseline differencing, a deformable convolutional feature extraction network, an attribution-guided attention fusion module, and a Transformer-based classification head. Initially, the model separately extracts attribution information relative to the background distribution and local spatial structural features (Fig. 3). These two types of information are then integrated through an attribution-guided attention mechanism, culminating in the prediction of the mineralization probability for the central pixel.

3.1. Attribution network

The objective of the attribution branch is to quantify the extent to which the input sample deviates from the background distribution of the training dataset, and to convert this quantification into informative guidance for subsequent attention modeling. Given an input sample x, a convolutional encoder is initially employed to extract its latent representation:$f(x) \in R^{Ch\times H\times W}$, Here, $H = W = 9$. To establish a reference state that holds geological relevance, the model employs the average of the encoded features derived from the training set samples as the baseline reference.

https://doi.org/10.5194/egusphere-2026-2616
Preprint. Discussion started: 16 June 2026

$f_{ref} = \sum_{i=1}^{N} f(x_i)$, (1)

Based on this framework, the differential representation of the input sample with respect to the reference distribution is defined as follows:

$\Delta f = f(x) - f_{ref}$, (2)

This definition aligns with the conceptual framework of DeepLIFT, which quantifies feature contributions by comparing the input sample to a reference baseline (Shrikumar et al., 2017; Lundberg and Lee, 2017). Specifically, feature attributions are determined by the variations of the sample relative to this baseline. Following this, the model employs a trainable parameter matrix to execute a linear projection of Δf, and derives the spatial attribution intensity by computing the L2 norm of the resulting projection:

$s(u, v) = \sigma(\| W_a \Delta f(u, v) \| 2)$, (3)

In this context, W_a denotes the trainable weight matrix, and $\sigma(\cdot)$ corresponds to the Sigmoid activation function. Consequently, the attribution score at the coordinate (u, v), denoted as $s(u, v) \in (0,1)$, signifies that higher values reflect a greater degree of anomaly at that location in comparison to the background feature distribution. Ultimately, the feature representation weighted by the attribution scores is expressed as follows:

$f_{attr}(u, v) = \Delta f(u, v) \cdot s(u, v)$, (4)

Subsequently, the data is projected into a unified feature space via a 1×1 convolution operation to produce the Key features necessary for the ensuing attention mechanism.

3.2. Deformable convolutional networks

In order to accurately characterize the irregular spatial structures associated with mineralization, this study utilizes the deformable convolutional network DCNv2 (Zhu et al., 2019) as the local feature extraction component. Unlike conventional convolutional operations with fixed sampling positions, deformable convolution dynamically adjusts both the sampling locations and weights in response to the input data (Dai et al., 2017). This adaptive capability renders it particularly well-suited for representing complex spatial patterns influenced by the interplay of faults, lithological boundaries, and elemental anomalies.

DCNv2 incorporates learnable spatial offsets and modulation scalars, allowing the convolutional kernel to dynamically adjust its sampling configuration and weight allocation in response to the specific spatial characteristics of the input data (Zhu et al., 2019). Denoting the standard grid sampling location as p, the output of DCNv2 at position p can be formulated as follows:

$y(p) = \sum_{k=1}^{K} w_k \cdot x(p + p_k + \Delta p_k) \cdot \Delta m_k$, (5)

https://doi.org/10.5194/egusphere-2026-2616
Preprint. Discussion started: 16 June 2026

In this context, Δp_k denotes the offset parameter learned by the network, which facilitates the adjustment of sampling points toward areas exhibiting more pronounced mineral control signals. The variable $\Delta m_k \in [0,1]$ functions as a modulation mechanism designed to attenuate the influence of irrelevant background noise. Together, these components enable the model to accurately capture the authentic non-rigid distribution patterns inherent in the 42-dimensional multi-source geoscientific spatial dataset.

3.3. Attribution-guided attention fusion

In order to facilitate the integrated modeling of attribution information alongside local structural features, this study proposes an attribution-guided attention fusion module. Denote the output of the deformable convolution branch as F_{dcn}, the guided features derived from the attribution branch as F_{attr}, and the attribution score map as S. The model initially formulates the query, key, and value, in this framework, Q and V are derived from deformable convolution features, while K is obtained from the output of the attribution network. This design enables the attention mechanism to preserve the capacity for local structural representation while incorporating attribution priors characterized by background deviation (Chen and Fan, 2023).

In the computation of multi-head attention, the model incorporates attribution scores as additional bias terms into the attention weights:

$$A = Softmax\left(\frac{Q^{\top}K}{d_h} + \lambda S\right), \quad (6)$$

Here, d_h denotes the dimensionality of a single-head feature, and λ represents a learnable fusion coefficient. Unlike conventional self-attention mechanisms, this approach allocates weights not only based on feature similarity but also enhances the prominence of anomalous regions within the attention distribution by leveraging attribution intensity. The fusion outcome can be expressed as follows:

$$F_{fuse} = Norm(W_o(AV) + F_{dcn}), \quad (7)$$

Residual connections and normalization operations are employed to enhance training stability and preserve the original local feature representations.

3.4. Visualization and interpretability analysis methods

To analyze the decision-making basis of the model from multiple perspectives, this study employs three types of visualization techniques: Grad-CAM (Selvaraju et al., 2016), Integrated Gradients (Sundararajan et al., 2017), and Transformer attention mapping.

Firstly, at the convolutional feature level, Grad-CAM is utilized to visualize the intermediate features of the deformable

https://doi.org/10.5194/egusphere-2026-2616
Preprint. Discussion started: 16 June 2026

convolutional module. Specifically, the gradients of the target class score with respect to the feature maps are subjected to global average pooling to obtain channel weights α_k, which are then used to construct the class activation map:

$$L^{C}_{Grad-CAM} = ReLU(\sum_k \alpha_k A^k), \quad (8)$$

This method can identify the spatial regions within a local window that are emphasized by deformable convolutional branches, facilitating the analysis of the response locations of dynamic sampling convolutions in mineralization discrimination.

Secondly, at the level of input variables, Integrated Gradients was employed to perform attribution analysis on all grid points of the entire map. Using the standardized baseline corresponding to the mean of the training set as the reference input, the contribution of each channel to the prediction outcome was estimated by accumulating the gradients along the path from the baseline to the input.

$$IGi(x) = (x_i - x_i') \int_0^1 \frac{F\left(x' + \alpha(x - x')\right)}{\partial x_i} d\alpha, \quad (9)$$

This method enables the quantification of the positive and negative contributions of various geochemical elements and ore-controlling factors at the pixel scale. Furthermore, it facilitates the identification of key mineralization indicator factors and their spatial distribution characteristics through spatial mapping.

Finally, at the global decision-making level, the multi-head self-attention matrices within the Transformer encoder are extracted, with a particular focus on the attention weights of the CLS token towards the central patch. Since the CLS token directly contributes to the final classification (Dosovitskiy et al., 2020), its attention distribution over spatial tokens can reflect the model's allocation of importance across different spatial locations during the discrimination process (Caron et al., 2021). By projecting this attention metric back onto the study area, it becomes possible to reveal the model's focal points at the regional scale and to examine their correspondence with the distribution of known mineral deposits (Reed et al., 2023).

https://doi.org/10.5194/egusphere-2026-2616
Preprint. Discussion started: 16 June 2026

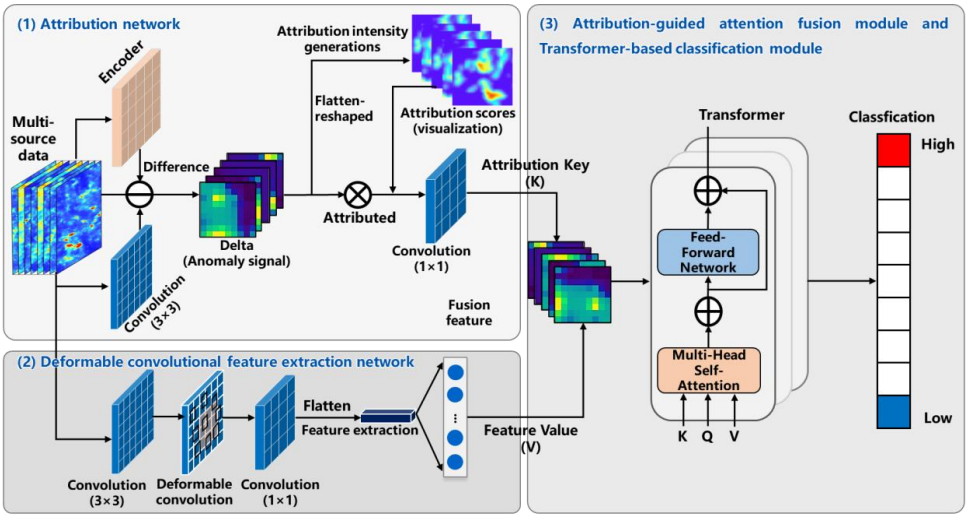


Figure 3. The whole structure of our proposed model: (1) an attribution network employing reference baseline differencing; (2) a deformable convolutional feature extraction network; (3) an attribution-guided attention fusion module, and a Transformer-based classification module.

4. Results

The model processes input data in the form of a 9×9×42 multi-channel neighborhood window centered on a target pixel, where the 42 channels comprise 39 geochemical elements alongside three categories of ore-controlling geological factors. In the supervised learning task for MPM, samples must be labeled as positive or negative to facilitate model training. Specifically, known mineral deposit locations are designated as positive samples, while non-mineralized areas serve as negative samples. The positive samples comprise 23 polymetallic Li-Be-Nb-Ta deposits within the study region. To maintain a balance between positive and negative samples, an equal number of 23 negative samples were selected from specific locations. The selection of negative samples adhered to the principle of random sampling, ensuring they were situated at a considerable distance from known deposits (within non-mineralized zones) and possessed corresponding values across all evidence layers to guarantee their validity (Nykänen et al., 2015). Prior to model training, data augmentation, which was adopted in Yang and Zuo (2024), was performed, resulting in 196 positive and negative samples, respectively. Subsequently, the dataset was partitioned into training and validation sets at an 8:2 ratio, comprising 314 and 78 samples, respectively.

Using the fully balanced augmented training dataset, the primary classifier is optimized via the cross-entropy loss function to model the binary classification probability distribution. Additionally, to facilitate spatial contrastive learning within the

https://doi.org/10.5194/egusphere-2026-2616
Preprint. Discussion started: 16 June 2026

DeepLIFT attribution network, the loss function incorporates an attribution spatial variance maximization regularization term weighted by a coefficient $\lambda = 0.001$. This regularizer guides the model to adaptively enlarge the score disparity between local anomalous points and background regions.

To prevent gradient explosion during backpropagation—particularly due to the complexity of high-dimensional geoscientific feature mappings—gradient clipping with a maximum norm of 1.0 is applied throughout training. Model parameter updates are performed using the AdamW optimizer, with a base learning rate of 5×10^{-5} and a weight decay factor of 0.01. The training process integrates a cosine annealing learning rate scheduler with CosineAnnealingWarmRestarts (Loshchilov and Hutter, 2016). The initial 5 epochs serve as a linear warm-up phase to stabilize early gradients, followed by periodic cosine decay cycles. This strategy balances stable convergence in the early training stages with the ability to escape local minima later on, making it well-suited for non-convex optimization problems under limited sample conditions.

4.1. Mineral prospectivity mapping

The application of the proposed DCN-Attribution-Transformer hybrid model to the study area resulted in a distinct and geologically insightful mineral prospectivity map (Fig. 4(a)). The model successfully delineated high-prospectivity zones that exhibit a strong spatial correlation with known mineral occurrences and key geological structures. For comparative analysis, the MPM generated by the standalone DCN, standalone Transformer, and the DCN-Transformer (without attribution) models were evaluated, revealing significant differences in the spatial coherence, and geological plausibility of the predicted targets (Fig. 4(b)-4(d)).

The prospectivity map generated by our proposed DCN-Attribution-Transformer model is characterized by highly focused and spatially coherent high-prospectivity zones. These zones are not merely isolated pixels but form contiguous, well-defined polygons that align remarkably well with major fault zones and the contacts between specific lithological units. The model appears to have effectively captured the multi-scale controls on mineralization; it identifies both the broader structural corridors that host fluid flow and the localized, intricate features within these corridors where ore deposition is most likely. The boundaries of these high-prospectivity areas are sharp and geologically reasonable, avoiding the overly dispersed patterns. This refined output suggests a successful integration where the DCN component has precisely delineated local alteration patterns and subtle geological boundaries, while the Transformer component has contextualized these local features within the larger structural framework, with the attribution mechanism ensuring this synthesis is driven by the most relevant evidence.

In contrast, the MPM produced by the standalone DCN model shows a strong emphasis on localized features. The high-

https://doi.org/10.5194/egusphere-2026-2616
Preprint. Discussion started: 16 June 2026

prospectivity zones are numerous and capture fine-grained geological details with high precision. However, they lack the interconnectedness that would suggest a comprehensive understanding of the larger mineralizing system. The model excels at identifying the immediate host rocks and alteration zones, but struggles to consistently delineate the broader structural pathways that control district-scale fertility. Conversely, the prospectivity map from the standalone Transformer model identifies regional-scale zones of elevated prospectivity that broadly correspond to major tectonic boundaries and intrusive complexes. The model successfully captures the global features by understanding long-range dependencies between geological elements. These zones are spatially coherent but lack definition at a local scale. The DCN-Transformer hybrid model (without the attribution component) represents a significant step forward, combining aspects of both previous models. Its output shows improved spatial coherence over the standalone DCN and better local detail than the standalone Transformer. The high-prospectivity zones are larger and more connected than those from the DCN model and contain more internal structure and variation than those from the Transformer model. However, a critical observation is that the boundaries of these zones can sometimes be less sharp and slightly more conservative than those of our proposed model. There is a sense that the integration, while effective, may not be optimally weighted, potentially allowing less relevant features to influence the final spatial distribution. This sometimes results in high-prospectivity areas that are spatially accurate but slightly over- or under-extended compared to the most stringent geological constraints.

In summary, the spatial comparison demonstrates the superior interpretability of the DCN-Attribution-Transformer model's output. It does not merely average the characteristics of its components but synthesizes them to produce a prospectivity map where the high-confidence zones are both precisely localized and contextually grounded within the regional geology. The result is a spatially optimized map that pinpoints specific, coherent targets, offering a more reliable and actionable guide for mineral exploration by effectively bridging the local-global gap in mineralization features.

https://doi.org/10.5194/egusphere-2026-2616
Preprint. Discussion started: 16 June 2026

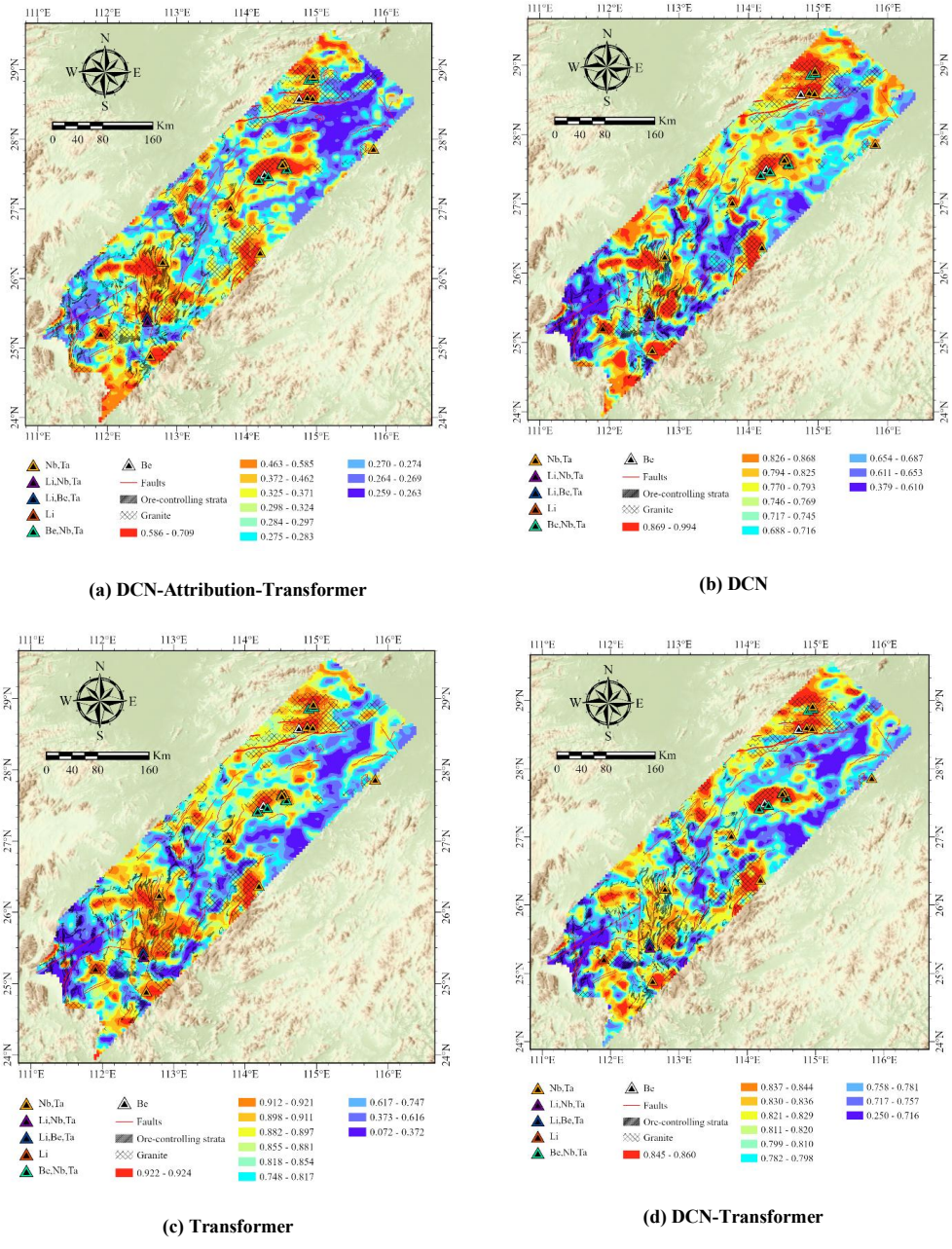


Figure 4. Geochemical anomalies associated with mineralization obtained by (a) DCN-Attribution-Transformer; (b) DCN; (c) Transformer; (d) DCN-Transformer

https://doi.org/10.5194/egusphere-2026-2616
Preprint. Discussion started: 16 June 2026

4.2 Model performance

To quantitatively evaluate the efficacy of the proposed DCN-Attribution-Transformer hybrid model for MPM, we conducted a comprehensive comparative analysis against three benchmark architectures: a standalone DCN, a standalone Transformer, and a baseline DCN-Transformer (without the attribution). The evaluation was performed on a held-out test set, leveraging multiple performance metrics—Accuracy (ACC), Area Under the Curve (AUC), Kappa, Matthews Correlation Coefficient (MCC), Precision, Recall, and F1-Score—to provide a holistic view of the models' predictive capabilities and generalization performance. The above indices can be expressed as:

$$ACC = \frac{TP+TN}{TP+FP+TN+FN}, \quad (10)$$

$$Kappa = \frac{ACC-P_e}{1-P_e};\ P_e = \frac{\frac{n}{2}(TP+FP)+\frac{n}{2}(FN+FTN)}{n^2}, \quad (11)$$

$$MCC = \frac{TP\times TN-FP\times FN}{\sqrt{(TP+FP)\times(TP+FN)\times(TN+FP)\times(TN+FN)}}, \quad (12)$$

$$Precision = \frac{TP}{TP+FP}, \quad (13)$$

$$Recall = \frac{TP}{TP+FN}, \quad (14)$$

$$F1 = \frac{2\times TP}{2\times TP+FP+FN}, \quad (15)$$

Here, TP (True Positive), TN (True Negative), FP (False Positive), and FN (False Negative) represent the correspondence between the actual binary labels and the classification predictions, with *n* standing for the total number of samples. AUC, short for Area Under the Curve, quantifies the area enclosed by the ROC curve relative to the area of a unit square. The ROC curve is constructed by plotting the true positive rate (TPR, also known as sensitivity) against the false positive rate (FPR, equivalent to 1−specificity) (Fawcett, 2006).

The quantitative results, summarized in Fig. 5, unequivocally demonstrate the superior performance of the proposed DCN-Attribution-Transformer model across all evaluation metrics. Most notably, our model achieved the highest AUC value of 0.960, significantly outperforming the standalone DCN (0.874), the standalone Transformer (0.915), and the baseline DCN-Transformer hybrid (0.951). The AUC metric, which reflects the model's ability to distinguish between prospective and non-prospective areas across all classification thresholds, underscores the hybrid model's enhanced robustness and its superior capacity for ranking mineralization potential. This improvement in AUC is statistically significant, indicating that the integration of the attribution mechanism provides a substantial leap in predictive power beyond what is achievable by simply combining DCN and Transformer modules.

https://doi.org/10.5194/egusphere-2026-2616
Preprint. Discussion started: 16 June 2026

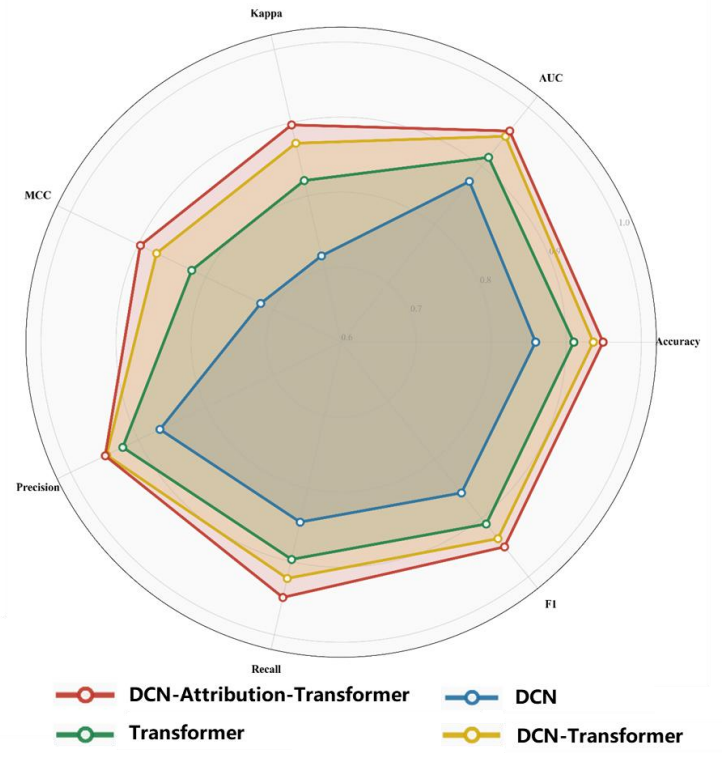


Figure 5. Evaluation of model performance between DCN-Attribution-Transformer, DCN, Transformer and DCN-Transformer in Accuracy, Precision, Recall, F1-score, AUC, Kappa, and MCC.

A detailed analysis of the threshold-based metrics at the optimal probability cutoff further validates the advantages of our proposed architecture. The model attained the highest accuracy (94.9%), confirming its overall correctness in classification. More critically, for MPM where the cost of missing a mineralized zone (false negative) is high, the Recall rate is of paramount importance. Our model achieved a Recall of 0.949, meaning it successfully identified 94.9% of all known mineralized locations, a marked improvement over the other models. This suggests that the model is effective at minimizing false negatives, a crucial characteristic for a practical exploration tool. Furthermore, the high Precision indicates that when the model predicts a location as prospective, it is highly likely to be correct, thereby reducing the cost and effort associated with false positives. The F1-Score, as the harmonic mean of Precision and Recall, provides a single metric to evaluate the balance between these two competing concerns. Our model's top F1-Score signifies that it achieves the most robust and balanced trade-off, effectively minimizing both false alarms and missed targets.

The performance gap between the baseline DCN-Transformer and our proposed model highlights the critical contribution

https://doi.org/10.5194/egusphere-2026-2616
Preprint. Discussion started: 16 June 2026

of the attribution mechanism. While the baseline hybrid already showed improvement over the standalone models by leveraging both local (via DCN) and global (via Transformer) features, its performance was limited by treating all features equally. The attribution mechanism acts as a dynamic feature modulator, guiding the Transformer's attention to the most salient multi-scale features identified by the DCN. This process effectively filters out irrelevant feature activations, allowing the model to focus its computational resources on the most diagnostic patterns for mineralization. The result is a more refined and interpretable feature representation that directly translates into higher predictive accuracy.

In contrast, the standalone models exhibited inherent limitations. The DCN model, while powerful for capturing local geometric variations and subtle mineralization-related features, struggled to contextualize these features within the broader geological framework, leading to a higher rate of false positives in structurally complex but barren areas. The standalone Transformer model, excelling at modeling long-range dependencies, overlooked local features that are critical indicators of mineralization, resulting in a lower Recall. In conclusion, the results presented in this section support the hypothesis that a hybrid model explicitly designed to leverage both local and global features, augmented with an attribution-driven feature refinement process, constitutes a state-of-the-art approach for mineral prospectivity mapping. The proposed DCN-Attribution-Transformer model sets a new benchmark for predictive performance, as evidenced by its superior scores across all key evaluation metrics.

5 Interpretations Discussions

5.1 Interpretations of the factor importance by attribution branching network

A pivotal advancement presented in this study is the integration of an attribution branching network, powered by the DeepLIFT algorithm, which provides an interpretation of the learned relationships within DCN-Transformer hybrid model. Moving beyond the black box paradigm, this approach allows us to deconstruct the model's complex, non-linear decision-making process for MPM and directly attribute predictive importance to the input geological features. The interpretations derived from this analysis not only validate the model's adherence to established geological principles but also offer profound insights into the localized and hierarchical control of mineralization, thereby substantiating DeepLIFT's critical role in enhancing both the performance trustworthiness and the interpretative power of our framework.

At the global level, the attribution analysis yields an overarching ranking of the importance of all input ore-controlling factors. By aggregating and averaging the attribution scores across all prospective and non-prospective cells in the study area, we can determine which evidence layers the hybrid DCN-Transformer model consistently relies upon most heavily for distinguishing high-potential zones (Fig. 6). For instance, our analysis reveals that proximity to major crustal faults

https://doi.org/10.5194/egusphere-2026-2616
Preprint. Discussion started: 16 June 2026

and granites, and specific geochemical anomaly complexes (e.g., W-Sn-Li-Be-Nb-Ba-P association) received high global attribution scores, aligning robustly with the documented metallogenic model in this region. This global feature importance is not a simple statistical correlation but a reflection of the model's learned, complex non-linear dependencies. It validates the relevance of the selected criteria and provides a quantitative, model-informed confirmation of established geological knowledge. Conversely, factors with persistently low global attribution scores may indicate redundant information or factors less critical at the regional predictive scale, guiding future data collection and feature engineering efforts.

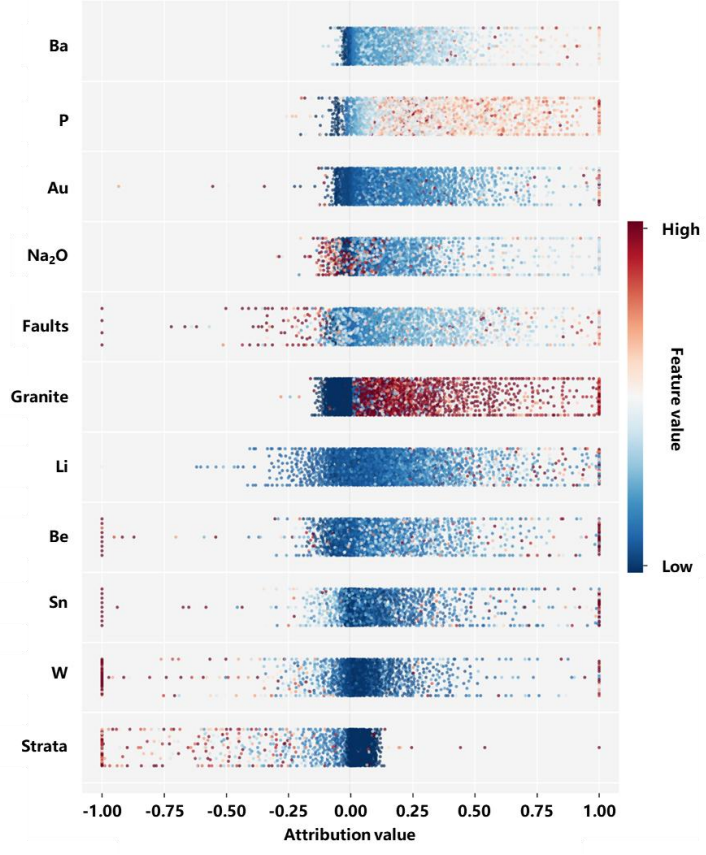
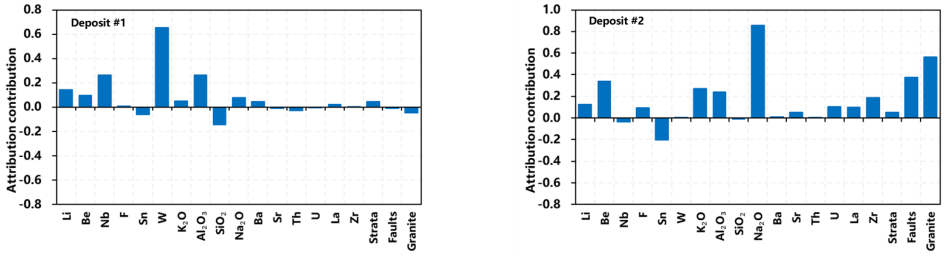


Figure 6. Summary scatter plots showing feature contribution attribution value obtained from DeepLIFT

The true power of the DeepLIFT-based attribution, however, is unlocked at the local level. For each single grid cell in the study area, the algorithm calculates a distinct set of attribution scores, decomposing the specific prediction for that location into the precise contribution of each input feature. This means that the importance of a given geological factor is not static but spatially variable, dependent on the unique multi-factor context at each point. DeepLIFT combined with Integrated

https://doi.org/10.5194/egusphere-2026-2616
Preprint. Discussion started: 16 June 2026

Gradients can determine the significant contributing factors of a particular mineral deposit. Figure 7 showed the significance values of the contributing factors predicted by the attribution branch network of our model checking the contribution of ore-controlling factors to particular mineral deposits (taking 10 deposits as example). This local explicability allows us to move beyond answering "which factors are important?" to answering "where and why is this factor critically important?". More importantly, these significance score rankings reveal areas where high model prospectivity is driven by different dominant factor combinations, such as some deposits may be primarily explained by structural factors, while other deposits may be predominantly due to geochemical signatures. For instance, the algorithm consistently assigned high positive attribution scores to proximity to major fault zones. This is not a simple linear correlation; DeepLIFT illuminates how the model uses this information. Thus, the interpretability provided by DeepLIFT is not ancillary; it is direct evidence of the model's sophisticated and geologically sound reasoning, justifying its high performance. The contribution of DeepLIFT within this architecture is therefore twofold. First, it enhances model performance transparency and trustworthiness. By providing a post-hoc rationale for each prediction, it allows geologists to critically evaluate whether the model's reasoning is geologically plausible, fostering greater confidence in its predictions. Second, it transforms the model from a pure prediction tool into a knowledge-discovery paradigm. The contribution of ore-controlling factors to the particular mineral deposits obtained by DeepLIFT and Integrated Gradients reveal previously unrecognized relationships or contextual dependencies between controls (Fig. 7). These interpretable outputs bridge the gap between data-driven deep learning and knowledge-driven conceptual modeling, creating a synergistic loop where the model's learned patterns can inform and refine the very geological understanding used to construct it.



https://doi.org/10.5194/egusphere-2026-2616
Preprint. Discussion started: 16 June 2026

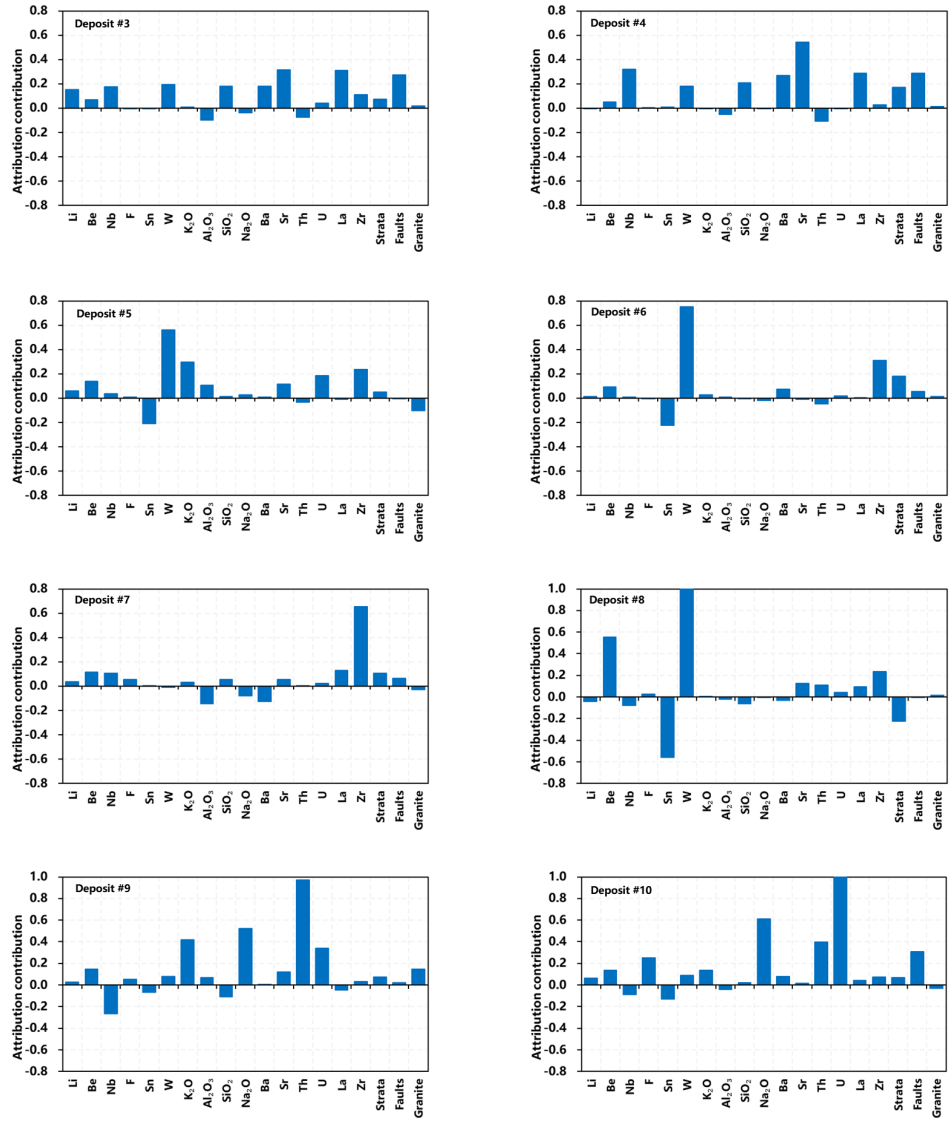


Figure 7. Significance values of the contributing factors corresponding to ten mineral deposits marked in Fig. 2

5.2 Interpretations of local and global feature by visualization

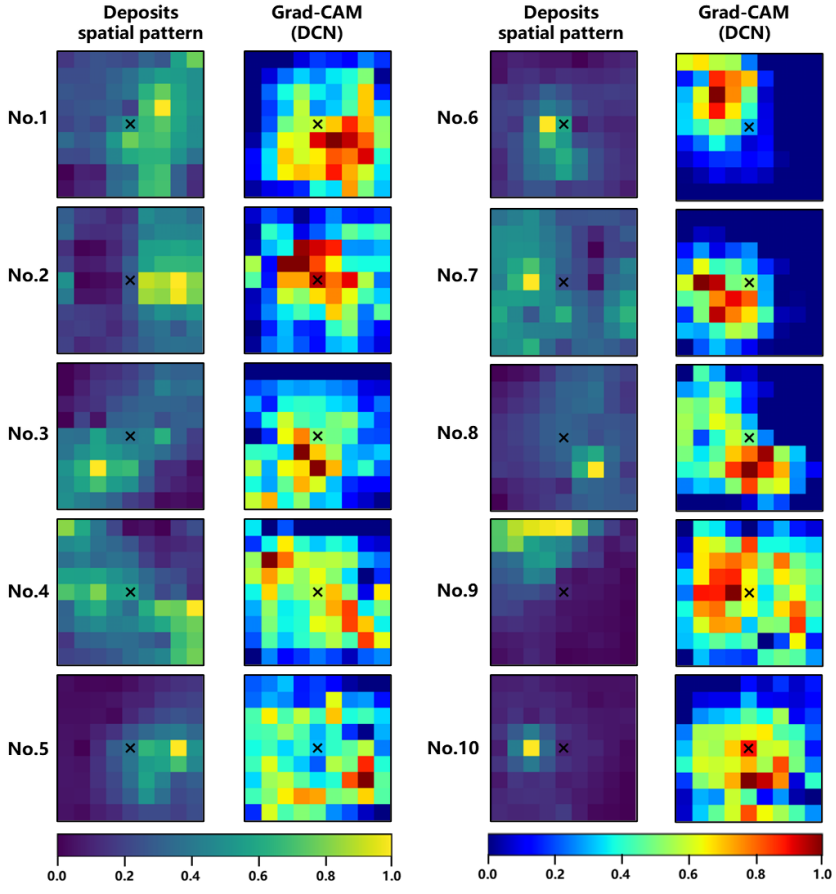
The synergistic integration of DCN and Transformer architecture is fundamentally validated through the visualization of their respective feature extraction mechanisms. The distinct yet complementary nature of these components, DCN in capturing localized anomalies and structures and the Transformer in modeling long-range, global dependencies, is clearly elucidated using Grad-CAM and attention map visualizations, respectively. These visual tools not only demystify the

https://doi.org/10.5194/egusphere-2026-2616
Preprint. Discussion started: 16 June 2026

model's decision-making process but also provide tangible, geological evidence for its superior performance.

The visualization of local features via Grad-CAM offers a compelling narrative for the superiority of DCN over standard convolutional layers. In our MPM task, critical ore-controlling factors such as specific subtle geochemical patterns often manifest as spatially constrained and geometrically irregular patterns. Standard CNN, with their fixed geometric structures, struggle to adapt to these non-rigid shapes, potentially leading to a loss of critical information. The Grad-CAM heatmaps generated from our DCN demonstrably overcome this limitation. They reveal highly focused and spatially adaptive activation regions that precisely align with known mineralized locations, even where the mineralizing signals are weak or have complex geometries (Fig. 8). For instance, the heatmaps outline the spatially anisotropic distribution of geochemical fields, exhibiting higher spatial coupling with actual geochemical spatial patterns, and enhance the interpretability of model decisions (Fig. 8). This demonstrates that DCN's ability to dynamically adjust its receptive field allows it to "zoom in" on these decisive local geochemical anomalies related to mineralization, directly contributing to enhanced model discriminative power (Fig. 9). From an interpretability standpoint, this provides mineral explorers with a spatially explicit map highlighting where the most diagnostic local evidence for mineralization is located within the broader prospective area, making the model's local reasoning transparent and geologically verifiable.

https://doi.org/10.5194/egusphere-2026-2616
Preprint. Discussion started: 16 June 2026



500 **Figure 8. Grad-CAM maps obtained by DCN with the geochemical patterns of ten mineral deposits marked in Fig. 2. The black crosses represent the known deposits.**

https://doi.org/10.5194/egusphere-2026-2616
Preprint. Discussion started: 16 June 2026

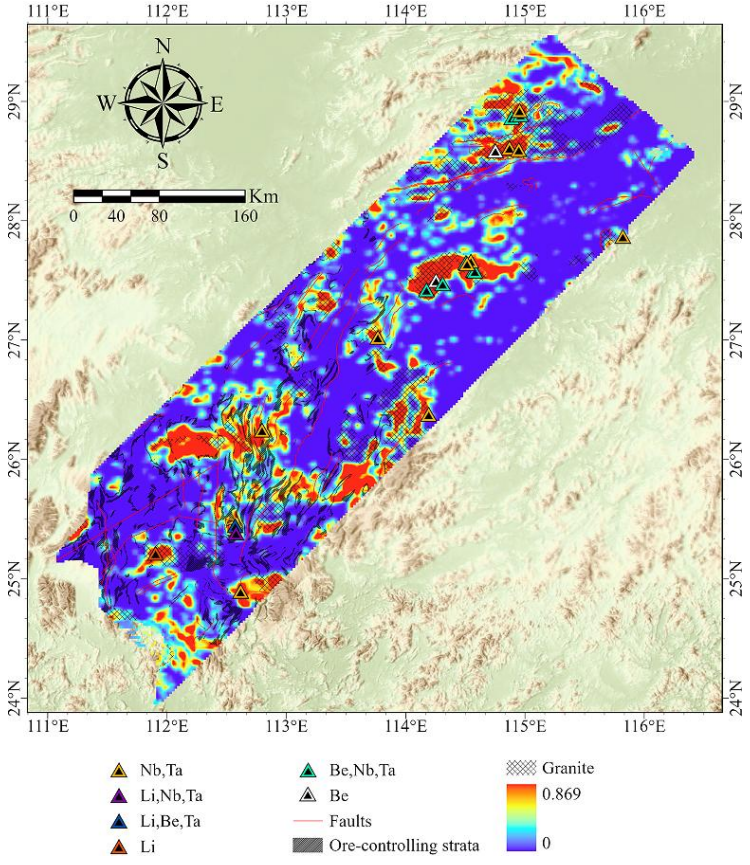


Figure 9. Grad-CAM map covered the region in geochemical pattern obtained by DCN

Conversely, the visualization of attention maps from the Transformer module unveils the model's capacity for global reasoning, a facet notoriously difficult to achieve with purely convolutional architectures. Ore formation is seldom the result of isolated factors; it is a consequence of the synergistic interplay of multiple geological processes acting over a large area. The self-attention mechanism inherently quantifies these pairwise relationships between all locations in the input feature space. The resulting attention maps vividly illustrate how the model connects disparate geological elements to form a coherent prospecting hypothesis. For example, a high attention score between a distal geochemical anomaly and a major regional fault, bypassing the intervening barren areas, suggests the model has learned the pathway of mineralizing fluids. Similarly, strong attention linking a specific intrusive rock body to a surrounding alteration zone underscores the model's grasp of genetic relationships. This global feature is crucial for distinguishing true mineral potential from areas that may exhibit strong local features but lack the necessary large-scale geological setting. By making these long-range dependencies visually explicit, the Transformer module elevates the model's interpretability from a local

https://doi.org/10.5194/egusphere-2026-2616
Preprint. Discussion started: 16 June 2026

"feature-spotter" to a global "systems-thinking" tool. It answers not just where the indicators are, but why they are significant in a broader geological context, thereby building greater trust and providing deeper insights for geological inference.

The dual visualization strategy solidifies the core thesis of our hybrid model. The Grad-CAM visualizations empirically confirm that the DCN component excels in extracting and highlighting the intricate, deformation-sensitive local features that are paramount for precise targeting. Simultaneously, the Transformer's attention maps provide an unprecedented window into the model's global reasoning, capturing the essential geological correlations that define a fertile mineral system. This multi-scale interpretability is not merely an ancillary benefit but is intrinsically linked to the model's robust performance. It ensures that the predictions are not a black-box output but are grounded in a spatially and contextually aware analysis that resonates with geological principles, thereby offering a more reliable and insightful tool for mineral prospectivity mapping. For instance, the model did not treat all faults equally; it discerned higher importance for faults with a specific orientation (e.g., NE-SW), which is known to have controlled fluid flow during the mineralizing event. This level of interpretation—quantifying not just "which" fault but "where" along it and under "what spatial context"-is a significant leap over traditional methods, which can provide a global, spatially-agnostic rank of feature importance (Fig. 10).

https://doi.org/10.5194/egusphere-2026-2616
Preprint. Discussion started: 16 June 2026

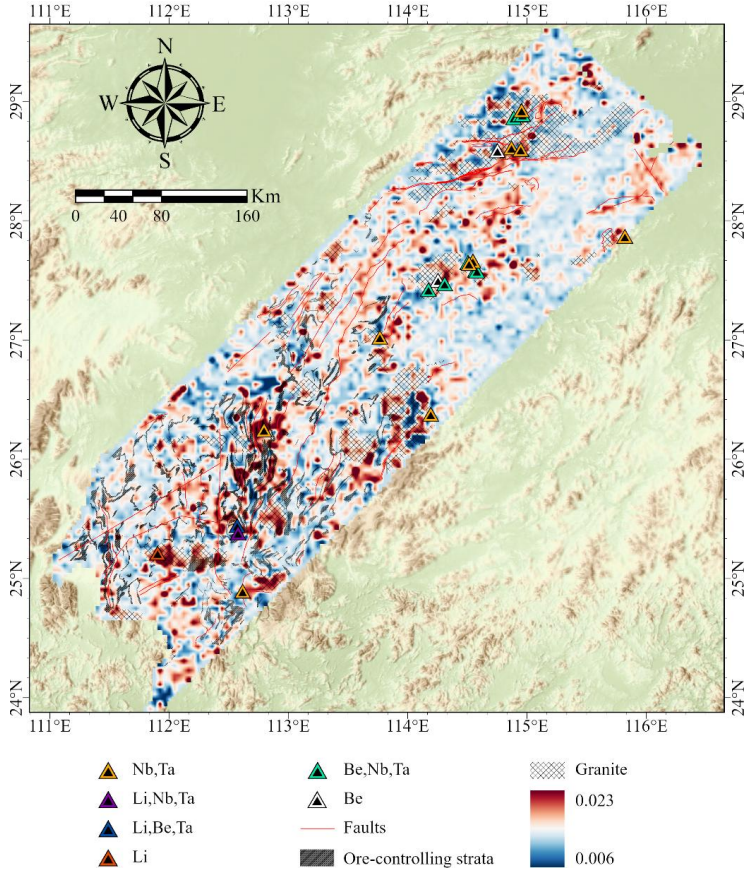


Figure 10. The attention map covered the region in geochemical pattern obtained by Transformer

6 Conclusions

This study successfully developed an interpretable DCN-Attribution-Transformer hybrid model for MPM, demonstrating its superior predictive performance over the standalone DCN, Transformer, and the baseline DCN-Transformer models, as evidenced by higher evaluation indices and more geologically plausible prospectivity maps with greater spatial precision and reliability, effectively balancing local structural features with global contextual relationships. The incorporation of the attribution branching network provided a critical, model-intrinsic, and pixel-level interpretation of factor importance, quantitatively revealing the specific contribution of each evidence layer and its spatial correlation with known mineral deposits, thereby moving beyond "black-box" predictions to offer a geological knowledge-aware MPM framework. Furthermore, the complementary visual interpretations using Grad-CAM and attention maps delivered profound insights into the model's decision-making process: Grad-CAM effectively highlighted the localized, spatially

https://doi.org/10.5194/egusphere-2026-2616
Preprint. Discussion started: 16 June 2026

constrained geological features (e.g., geochemical anomalies) that the model prioritized, while the attention maps unveiled how the model globally contextualizes these local features by identifying long-range dependencies between geological units. This dual-scale interpretability not only builds greater trust in the model's predictions but also provides a more comprehensive and geologically meaningful understanding of the controlling factors of mineralization, establishing a new, interpretable paradigm for data-driven mineral exploration.

Code and Data Availability

The code and data used for mineral prospectivity mapping based on an interpretable DCN-Transformer hybrid model with attribution are archived on Zenodo (https://doi.org/10.5281/zenodo.20039700; Hao and Xiong, 2026).

CRediT Authorship Contribution Statement

YH: Conceptualization, Methodology, Writing – original draft. **YX:** Conceptualization, Resources, Methodology, Writing – original draft.

Declaration of Competing Interest

The contact author has declared that none of the authors has any competing interests.

Acknowledgement

This research was supported by the National Key Research and Development Program of China (Nos. 22023YFC2906400, 22023YFC2906404), the Deep Earth Probe and Mineral Resources Exploration - National Science and Technology Major Project (Nos. 2025ZD1006700 and 2025ZD1006705), and the Key Research, Development Program of Xinjiang Uygur Autonomous Region, China (No. 2024B03010-3) , and the State Key Laboratory of Geological Processes and Mineral Resources, China University of Geosciences (No. GPMR-2025-QT01).

References

Bonham-Carter, G.: Geographic information systems for geoscientists: modelling with GIS (No. 13). Elsevier, 1994.

Carranza, E.J.M.: Geochemical anomaly and mineral prospectivity mapping in GIS (Vol. 11). Elsevier, 2008.

Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., and Joulin, A.: Emerging properties in self-supervised vision transformers. In Proceedings of the IEEE/CVF international conference on computer vision, pp.

https://doi.org/10.5194/egusphere-2026-2616
Preprint. Discussion started: 16 June 2026

9650-9660, 2021.

Carvalho, D.V., Pereira, E.M., and Cardoso, J.S.: Machine learning interpretability: A survey on methods and metrics. Electronics, 8(8), 832, https://doi.org/10.3390/electronics8080832, 2019.

Che, X. D., Wang, R. C., Wu, F. Y., Zhu, Z. Y., Zhang, W. L., Hu, H., Xie, L., Lu, J. J., and Zhang, D.: Episodic Nb–Ta mineralisation in South China: Constraints from in situ LA–ICP–MS columbite-tantalite U–Pb dating. Ore Geology Reviews, 105, 71-85, https://doi.org/10.1016/j.oregeorev.2018.11.023, 2019.

Chen, C., and Fan, L.: An attribution deep learning interpretation model for landslide susceptibility mapping in the three gorges reservoir area. IEEE Transactions on Geoscience and Remote Sensing, 61, 1-15, https://doi.org/10.1109/TGRS.2023.3323668, 2023.

Chen, J., and Jahn, B. M.: Crustal evolution of southeastern China: Nd and Sr isotopic evidence. Tectonophysics, 284(1-2), 101-133, https://doi.org/10.1016/S0040-1951(97)00186-8, 1998.

Chen, Y., Chen, B., Shayilan, A.: Combining categorical boosting and Shapley additive explanations for building an interpretable ensemble classifier for identifying mineralization-related geochemical anomalies. Ore Geology Reviews, 173, 106263, https://doi.org/10.1016/j.oregeorev.2024.106263, 2024.

Chen, Z., and Zuo, R.: Geological-knowledge-guided graph self-supervised pretraining framework for identifying mineralization-related geochemical anomalies. Computers & Geosciences, 199, 105913, https://doi.org/10.1016/j.cageo.2025.105913, 2025.

Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., and Wei, Y.: Deformable convolutional networks, In Proceedings of the IEEE International Conference on Computer Vision, pp. 764-773, 2017.

Dosovitskiy, A.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929, 2020.

Fawcett, T.: An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861-874, https://doi.org/10.1016/j.patrec.2005.10.010, 2006.

Gong, L., and Cheng, Q.: Exploiting edge features for graph neural networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 9211-9219, 2019.

Hao, Y., and Xiong, Y.: Considering mineralization local-global geological features: An interpretable DCN-Transformer hybrid model with attribution for mineral prospectivity mapping, GitHub [code], https://github.com/Ccat-chuan/YFF, 2026.

Hou, Q., Gao, Z., Lu, M., and Yu, Y.: A Hybrid Transformer-CNN Model for Interpolating Meteorological Data on the Tibetan Plateau. Atmosphere, 16(4), 431, https://doi.org/10.3390/atmos16040431, 2025.

https://doi.org/10.5194/egusphere-2026-2616
Preprint. Discussion started: 16 June 2026

Hu, W. Z., Huang, J. P., and Huang, X. W.: Characteristics and genesis of pegmatite type Nb-Ta spodumene deposit in Xigang, South Jiangxi. Resources Survey & Environment, 26(4), 258-266, 2005.

Hua, R., Chen, P., Zhang, W., Yao, J., Lin, J., and Zhang, Z.: Metallogeneses and their geodynamic settings related to mesozoic granitoids in the Nanling Range. Geological Journal of China Universities, 11(3), 291-304, 2005.

Lundberg, S. M., and Lee, S. I.: A unified approach to interpreting model predictions. Advances in neural information processing systems, 30, 2017.

Li, T., Zuo, R., Xiong, Y., and Peng, Y.: Random-drop data augmentation of deep convolutional neural network for mineral prospectivity mapping. Natural Resources Research, 30(1), 27-38, https://doi.org/10.1007/s11053-020-09742-z, 2021.

Li, X., Xiong, H., Li, X., Wu, X., Zhang, X., Liu, J., Bian, J., and Dou, D.: Interpretable deep learning: Interpretation, interpretability, trustworthiness, and beyond. Knowledge and Information Systems, 64(12), 3197-3234, https://doi.org/10.1007/s10115-022-01756-8, 2022.

Li, Q., Chen, G., and Luo, L.: Mineral prospectivity mapping using attention-based convolutional neural network. Ore Geology Reviews, 156, 105381, https://doi.org/10.1016/j.oregeorev.2023.105381, 2023.

Li, C., Xiao, K., Sun, L., Tang, R., Dong, X., Qiao, B., and Xu, D.: CNN-Transformers for mineral prospectivity mapping in the Maodeng–Baiyinchagan area, Southern Great Xing'an Range. Ore Geology Reviews, 167, 106007, https://doi.org/10.1016/j.oregeorev.2024.106007, 2024.

Li, X., Li, W., and Li, Z.: On the genetic classification and tectonic implications of the Early Yanshanian granitoids in the Nanling Range, South China. Chinese Science Bulletin, 52(14), 1873-1885, https://doi.org/10.1007/s11434-007-0259-0, 2007.

Loshchilov, I., and Hutter, F.: Sgdr: Stochastic gradient descent with warm restarts. arXiv preprint arXiv:1608.03983, 2016.

Luo, Z., Zuo, R., Xiong, Y., and Zhou, B.: Metallogenic-factor variational autoencoder for geochemical anomaly detection by ad-hoc and post-hoc interpretability algorithms. Natural Resources Research, 32(3), 835-853, https://doi.org/10.1007/s11053-023-10200-9, 2023.

McCuaig, T.C., Beresford, S., and Hronsky, J.: Translating the mineral systems approach into an effective exploration targeting system. Ore Geology Reviews 38(3), 128-138, https://doi.org/10.1016/j.oregeorev.2010.05.008, 2010.

McMillan, M., Haber, E., Peters, B., and Fohring, J.: Mineral prospectivity mapping using a VNet convolutional neural network. The Leading Edge 40(2), 99-105, https://doi.org/10.1190/tle40020099.1, 2021.

Nykänen, V., Lahti, I., Niiranen, T., and Korhonen, K.: Receiver operating characteristics (ROC) as validation tool for

https://doi.org/10.5194/egusphere-2026-2616
Preprint. Discussion started: 16 June 2026

prospectivity models—A magmatic Ni–Cu case study from the Central Lapland Greenstone Belt, Northern Finland. Ore Geology Reviews, 71, 853-860, https://doi.org/10.1016/j.oregeorev.2014.09.007, 2015.

Porwal, A., Carranza, E.J.M., and Hale, M.: Knowledge-driven and data-driven fuzzy models for predictive mineral potential mapping. Natural Resources Research 12(1), 1-25, https://doi.org/10.1023/A:1022693220894, 2003.

Pradhan, B., Jena, R., Talukdar, D., Mohanty, M., Sahu, B.K., Raul, A. K., and Abdul Maulud, K.N.: A new method to evaluate gold mineralisation-potential mapping using deep learning and an explainable artificial intelligence (XAI) model. Remote Sensing 14(18), 4486, https://doi.org/10.3390/rs14184486, 2022.

Qiao, G. B., Zhang, H. D., Wu, Y. Z., Jin, M. S., Du, W., Zhao, X. J., and Chen, D. H.: Petrogenesis of the Dahongliutan monzogranite in western Kunlun: Constraints from SHRIMP zircon U-Pb geochronology and geochemical characteristics. Acta Geologica Sinica, 89(7), 1180-1194, 2015.

Reed, C. J., Gupta, R., Li, S., Brockman, S., Funk, C., Clipp, B., Keutzer, K., Candido, S., Uyttendaele, M., and Darrell, T.: Scale-mae: A scale-aware masked autoencoder for multiscale geospatial representation learning. In Proceedings of the IEEE/CVF International Conference on Computer Vision,pp. 4088-4099, 2023.

Sabbaghi, H., Tabatabaei, S.H., Fathianpour, N.: Geologically-constrained GANomaly network for mineral prospectivity mapping through frequency domain training data. Scientific Reports, 14(1), 6236, https://doi.org/10.1038/s41598-024-56644-8, 2024.

Samek, W., Wiegand, T., and Müller, K.R.: Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. arXiv preprint arXiv:1708.08296, 2017.

Samek, W., Montavon, G., Lapuschkin, S., Anders, C.J., and Müller, K.R.: Explaining deep neural networks and beyond: A review of methods and applications. Proceedings of the IEEE 109(3), 247-278, https://doi.org/10.1109/JPROC.2021.3060483, 2021.

Selvaraju, R. R., Das, A., Vedantam, R., Cogswell, M., Parikh, D., and Batra, D.: Grad-CAM: Why did you say that?, arXiv preprint arXiv:1611.07450, 2016.

Shi, L., Xu, Y., and Zuo, R.: A heterogeneous graph construction method for mineral prospectivity mapping. Natural Resources Research, 33(4), 1365-1376, https://doi.org/10.1007/s11053-024-10344-2, 2024.

Shrikumar, A., Greenside, P., and Kundaje, A.: Learning important features through propagating activation differences. In International conference on machine learning, pp. 3145-3153, 2017.

Sihombing, F. M., Palin, R. M., Hughes, H. S., and Robb, L. J.: Improved mineral prospectivity mapping using graph neural networks. Ore Geology Reviews, 172, 106215, https://doi.org/10.1016/j.oregeorev.2024.106215, 2024.

Song, D., Dai, F., Liu, Y., Wei, M., and Tan, H.: A multi-scale CNN-transformer hybrid network for microseismic signal

https://doi.org/10.5194/egusphere-2026-2616
Preprint. Discussion started: 16 June 2026

arrival picking: Model analysis and engineering application. Engineering Geology, 353, 108109, https://doi.org/10.1016/j.enggeo.2025.108109, 2025.

Sundararajan, M., Taly, A., and Yan, Q.: Axiomatic attribution for deep networks. In International conference on machine learning, pp. 3319-3328, 2017.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., and Polosukhin, I.: Attention is all you need. Advances in neural information processing systems, 30, 2017.

Von Rueden, L., Mayer, S., Beckh, K., Georgiev, B., Giesselbach, S., Heese, R., Kirsch, B., Pfrommer, J., Pick, A., Ramamurthy, R., Walczak, M., Garcke, J., Bauckhage, C., and Schuecker, J.: Informed machine learning–a taxonomy and survey of integrating prior knowledge into learning systems. IEEE Transactions on Knowledge and Data Engineering 35(1), 614-633, https://doi.org/10.1109/TKDE.2021.3079836, 2021.

Wang, F. Y., Ling, M. X., Ding, X., Hu, Y. H., Zhou, J. B., Yang, X. Y., Liang, H. Y., Fan, W. M., and Sun, W.: Mesozoic large magmatic events and mineralization in SE China: oblique subduction of the Pacific plate. International Geology Review, 53(5-6), 704-726, https://doi.org/10.1080/00206814.2010.503736, 2011.

Wang, D., Huang, F., Wang, Y., He, H., Li, X., Liu, X., Sheng, J., and Liang, T.: Regional metallogeny of Tungsten-tin-polymetallic deposits in Nanling region, South China. Ore Geology Reviews, 120, 103305, https://doi.org/10.1016/j.oregeorev.2019.103305, 2020.

Wen, C., Shao, Y., Xiong, Y., Li, J., and Jiang, S.: Ore genesis of the Baishawo Be-Li-Nb-Ta deposit in the northeast Hunan Province, south China: Evidence from geological, geochemical, and U-Pb and Re-Os geochronologic data. Ore Geology Reviews, 129, 103895, https://doi.org/10.1016/j.oregeorev.2020.103895, 2021.

Wyborn, L.A.I., Heinrich, C.A., and Jaques, A.L.: Australian Proterozoic mineral systems: essential ingredients and mappable criteria. In 1994 AusIMM Annual Conference: Australian Mining Looks North-the Challenges and Choices, pp. 109-115, 1994.

Xie, X., Mu, X., and Ren, T.: Geochemical mapping in China. Journal of Geochemical Exploration, 60, 99–113, https://doi.org/10.1016/S0375-6742(97)00029-0, 1997.

Xiong, Y., Zuo, R., and Carranza, E.J.M.: Mapping mineral prospectivity through big data analytics and a deep learning algorithm. Ore Geology Reviews, 102, 811-817, https://doi.org/10.1016/j.oregeorev.2018.10.006, 2018.

Xu, Y., Zuo, R., and Zhang, G.: The graph attention network and its post-hoc explanation for recognizing mineralization-related geochemical anomalies. Applied Geochemistry, 155, 105722, https://doi.org/10.1016/j.apgeochem.2023.105722, 2023.

Xu, Y., and Zuo, R.: An interpretable graph attention network for mineral prospectivity mapping. Mathematical

https://doi.org/10.5194/egusphere-2026-2616
Preprint. Discussion started: 16 June 2026

Geosciences, 56(2), 169-190, https://doi.org/10.1007/s11004-023-10076-8, 2024.

Xu, Y., Shi, L., and Zuo, R.: Geologically constrained unsupervised dual-branch deep learning algorithm for geochemical anomalies identification. Applied Geochemistry, 174, 106137, https://doi.org/10.1016/j.apgeochem.2024.106137, 2024.

Xu, Y., Zuo, R., Chen, Z., Shi, Z., and Kreuzer, O.P.: Recent advances and future research directions in deep learning as applied to geochemical mapping. Earth-Science Reviews, 105209, https://doi.org/10.1016/j.earscirev.2025.105209, 2025.

Yang, N., Zhang, Z., Yang, J., and Hong, Z.: Applications of data augmentation in mineral prospectivity prediction based on convolutional neural networks. Computers & Geosciences, 161, 105075, https://doi.org/10.1016/j.cageo.2022.105075, 2022.

Yang, F., Zuo, R., Xiong, Y., Wang, J., and Zhang, G.: An interpretable attention branch convolutional neural network for identifying geochemical anomalies related to mineralization. Journal of Geochemical Exploration, 252, 107274, https://doi.org/10.1016/j.gexplo.2023.107274, 2023.

Yang, F., and Zuo, R.: Geologically constrained convolutional neural network for mineral prospectivity mapping. Mathematical Geosciences, 56(8), 1605-1628, https://doi.org/10.1007/s11004-024-10141-w, 2024.

Yang, F., Zuo, R., and Kreuzer, O.P.: Artificial intelligence for mineral exploration: A review and perspectives on future directions from data science. Earth-Science Reviews, 258, 104941, https://doi.org/10.1016/j.earscirev.2024.104941, 2024.

Yu, S., Deng, H., Liu, Z., Chen, J., Xiao, K., and Mao, X.: Identification of geochemical anomalies using an end-to-end transformer. Natural Resources Research, 33(3), 973-994, https://doi.org/10.1007/s11053-024-10334-4, 2024.

Zhang, S., Carranza, E.J.M., Wei, H., Xiao, K., Yang, F., Xiang, J., Zhang, S., and Xu, Y.: Data-driven mineral prospectivity mapping by joint application of unsupervised convolutional auto-encoder network and supervised convolutional neural network. Natural Resources Research, 30(2), 1011-1031, https://doi.org/10.1007/s11053-020-09789-y, 2021.

Zhang, X., Xiong, Y., and Chen, Z.: Recognizing spatial geochemical anomaly patterns using deformable convolutional networks guided with geological knowledge. Geoscientific Model Development, 19(5), 2219-2238, https://doi.org/10.5194/gmd-19-2219-2026, 2026.

Zhao, Z., Chen, T., Dou, J., Liu, G., and Plaza, A.: Landslide susceptibility mapping considering landslide local-global features based on CNN and transformer. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 17, 7475-7489, https://doi.org/10.1109/JSTARS.2024.3379350, 2024.

https://doi.org/10.5194/egusphere-2026-2616
Preprint. Discussion started: 16 June 2026

Zhou, X., Sun, T., and Shen, W.: Petrogenesis of Mesozoic granitoids and volcanic rocks in South China: a response to tectonic evolution. Episodes, 29:26–33, https://doi.org/10.18814/epiiugs/2006/v29i1/004, 2006.

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., and Sun, M.: Graph neural networks: A review of methods and applications. AI open ,1, 57-81, https://doi.org/10.1016/j.aiopen.2021.01.001, 2020.

Zhu, J., Fang, L., and Ghamisi, P.: Deformable convolutional neural networks for hyperspectral image classification. IEEE Geoscience and Remote Sensing Letters, 15(8), 1254-1258, https://doi.org/10.1109/LGRS.2018.2830403, 2018.

Zhu, X., Hu, H., Lin, S., and Dai, J.: Deformable convnets v2: More deformable, better results, In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9308-9316, 2019.

Zuo, R.: A nonlinear controlling function of geological features on magmatic–hydrothermal mineralization. Scientific Reports, 6(1): 27127, https://doi.org/10.1038/srep27127, 2016.

Zuo, R., Luo, Z., Xiong, Y., and Yin, B.: A geologically constrained variational autoencoder for mineral prospectivity mapping. Natural Resources Research, 31(3), 1121-1133, https://doi.org/10.1007/s11053-022-10050-x, 2022.

Zuo, R., and Xu, Y.: Graph deep learning model for mapping mineral prospectivity. Mathematical Geosciences, 55(1), 1-21, https://doi.org/10.1007/s11004-022-10015-z, 2023.

Zuo, R., Xiong, Y., Wang, Z., Wang, J., and Kreuzer, O.P.: A new generation of artificial intelligence algorithms for mineral prospectivity mapping. Natural Resources Research, 32(5), 1859-1869, https://doi.org/10.1007/s11053-023-10237-w, 2023.

Zuo, R., Cheng, Q., Xu, Y., Yang, F., Xiong, Y., Wang, Z., and Kreuzer, O.P.: Explainable artificial intelligence models for mineral prospectivity mapping. Science China Earth Sciences, 67(9), 2864-2875, https://doi.org/10.1007/s11430-024-1309-9, 2024.

Zuo, R., Yang, F., Cheng, Q., and Kreuzer, O.P.: A novel data-knowledge dual-driven model coupling artificial intelligence with a mineral systems approach for mineral prospectivity mapping. Geology, 53(3), 284-288, https://doi.org/10.1130/G52970.1, 2025.