minerals

MDPI

Article

Unsupervised Autoencoder-Based Feature Ranking and Anomaly Detection for Porphyry Copper Prospectivity Mapping from Multi-Source Geospatial Datasets

Mobin Saremi [1], Zohre Hoseinzade [2], Adel Shirazy [1,3], Aref Shirazi [1] and Amin Beiranvand Pour [4,5,*]

[1] Department of Mining Engineering, Amirkabir University of Technology, Tehran 1591634311, Iran; mobinsaremi@gmail.com (M.S.); adel.shirazy@gmail.com (A.S.); aref.shirazi@gmail.com (A.S.)
[2] Department of Mining Engineering, Hamedan University of Technology, Hamedan 65155-579, Iran; zhoseinzade9@gmail.com
[3] Karlsruhe Institute of Technology (KIT), Kaiserstraße 1276131, Karlsruhe Campus Nord, Her-mann-von Helmholtz-Platz, 176344 Eggenstein-Leopoldshafen, Germany
[4] Institute of Oceanography and Environment (INOS), Higher Institution Center of Excellence (HICoE) in Marine Science, University Malaysia Terengganu (UMT), Kuala Nerus 21030, Malaysia
[5] School of Earth Sciences, Damghan University, Damghan 3671641167, Iran
* Correspondence: beiranvand.pour@umt.edu.my; Tel.: +609-6683824; Fax: +609-6692166

Abstract

The mineral system model formalizes the critical geological processes and mappable parameters that control ore formation, which can then be translated into spatial predictors used as input features in machine learning (ML)-based mineral prospectivity mapping (MPM). In most MPM studies, exploration evidence features are indeed derived from the mineral system model of the targeted deposit type. However, not all features produced in this way are necessarily informative or favorable for prospectivity analysis. This challenge can be addressed by using feature selection frameworks to identify the most relevant features before applying ML and deep learning (DL) algorithms for mathematical integration. To address this need, this study employs an unsupervised variational autoencoder (VAE) framework to evaluate and rank exploration evidence layers. The VAE quantifies feature importance through a systematic strategy that measures the sensitivity of reconstruction-error components, mean squared error (MSE), mean absolute error (MAE), and Kullback–Leibler (KL) divergence, to individual feature variations. In this way, the VAE ranks the exploration features and helps to identify those that are the most useful for prospectivity mapping. The proposed approach was applied to a real geo-dataset from a porphyry copper district in Iran. Based on the conceptual model of porphyry copper mineralization, 15 evidence layers were generated, including proximity to phyllic, argillic, propylitic, iron oxide, and silicification alteration zones; proximity to intrusive rocks, faults, and fault intersections; and geochemical maps of Cu, Mo, Sb, Pb, Zn, As, and W. The VAE-based ranking indicated that evidence layers related to hydrothermal alterations, intrusive rocks, and faults were the most influential exploration features, whereas geochemical evidence layers showed lower relative importance. Based on this evaluation, two modeling scenarios were considered: in the first, all available features were used, and in the second, only the features selected by the VAE framework were included. In both cases, the final prospectivity model was produced by an autoencoder (AE). For comparison, the prediction-area (P–A) plots of the two prospectivity models were generated using 14 known mineral occurrences as positive ground-truth labels, indicating that the model based on the selected features achieved a higher prediction rate (80%) than the model based on all features (72%). These results demonstrate that the evidence layers

Academic Editors: Xavier Emery and Yongzhang Zhou

Received: 6 May 2026
Revised: 18 June 2026
Accepted: 18 June 2026
Published: 22 June 2026

derived from the mineral system approach can benefit from unsupervised VAE-based evaluation, leading to improved performance of the prospectivity modeling.

Keywords: porphyry copper prospectivity mapping; machine learning; variational autoencoder (VAE); unsupervised feature selection; deep learning; reconstruction error

1. Introduction

In recent years, a wide range of approaches, from classical mathematical methods to advanced data-driven mineral prospectivity techniques such as machine learning (ML) and deep learning (DL), have been applied in mineral exploration, targeting to identify promising areas and discover new mineral deposit sites [1–7]. All prospectivity mapping methods require a set of geo-features, commonly known as evidence layers, which represent different ore-forming subsystems (e.g., pre-, syn-, and post-mineralization processes) [8–10]. These methods combine and analyze the evidence layers to delineate favorable zones that may host undiscovered mineral deposits [11]. Because of this, the suitability and favorability of input features play a critical role in the performance of mineral prospectivity mapping (MPM). In practice, the most widely used and generally accepted approach among exploration geologists is to analyze the mineral system and translate it into a conceptual model [12–14]. This procedure (as the first approach) is commonly regarded as the foundation of MPM, as it provides the primary set of exploration features [8]. A complementary approach involves evaluating all layers derived from the conceptual model using various performance-assessment techniques to identify the most effective evidence layers for prospectivity analysis and to further evaluate the first approach.

As mentioned above, the first approach represents a knowledge-driven feature selection strategy in which exploration geologists determine which evidence layers should be constructed and used in MPM. This approach relies on the expertise and interpretive skills of specialists (e.g., exploration geologists, geochemists) and is based on their understanding of the mineral system, conceptual models, and ore-forming processes associated with the targeted deposit type [11,13–15]. For example, Mirzabozorg, Saremi [2] produced eight evidence layers for Fe prospectivity mapping in the Esfordi area by considering the geological setting and the conceptual model of Fe mineralization in that region. Similarly, Yousefi, Lindsay [16] generated five evidence layers for regional-scale mineral exploration targeting in the Nagisan area of Iran using the mineral-system model of porphyry copper mineralization. Although this approach forms the basis of MPM and remains essential, the process of deriving exploration features is inherently subjective. It depends on the experts' understanding of the mineral system, which may result in an incomplete or poorly defined conceptual model, overlooking important features, or retaining redundant and low-value geo-features that introduce unnecessary repetition into the geospatial dataset. Individually, an exploration feature may appear relevant; however, when combined with other correlated features, it may provide little or no additional information. Therefore, while the knowledge-driven procedure provides the fundamental feature set for MPM, the evidence layers produced through this approach can benefit from further evaluation through different approaches aimed at selecting a subset of discriminative features and improving the overall performance of prospectivity modeling.

In response to these requirements, exploration geologists adopt several procedures to assess the features used in prospectivity modeling. One common approach is based on the principle that a suitable exploration feature (i.e., evidence layer) should

demonstrate a strong ability to predict known mineral occurrences associated with the targeted deposit type [17,18]. Following this principle, various quantitative assessment methods, such as ROC curves [19], success-rate curve [20], and prediction-area (P-A) plot [21], can be used to evaluate the predictive performance of each feature. These techniques rely on ground truth positive points (i.e., known mineral occurrences) [8] to rank the evidence layers generated in the baseline knowledge-driven approach (e.g., conceptual model procedures). Such analyses help exploration geologists examine the predictive contribution of each feature and identify the most informative geo-features for subsequent prospectivity modeling [21,22]. For example, Daviran, Maghsoudi [22] used a success-rate curve to evaluate twelve selected elements for generating a multi-element geochemical evidence layer and found that Au, Cu, Mo, Pb, Sb, and Zn were the most effective indicators. While these methods are effective for identifying mineralization-related features, they may not fully capture the complex interrelationships among the variables. Additionally, methods like ROC and P-A plot may be limited in regions with few or no known mineral deposits, such as greenfield environments. In addition, each method has inherent limitations and thus cannot comprehensively assess all aspects of feature behavior.

Alongside these performance evaluation techniques, data-driven approaches have also been used to examine the relative importance of exploration features. A common group of such approaches leverages the capabilities of supervised learning algorithms to evaluate and rank features based on their relationship with labeled samples. Measures such as SHAP values, Gini index, or information gain are commonly used in several algorithms, like random forest and so on [23–26].

These approaches are useful in mineral exploration targeting because they can identify important features that experts might overlook. However, they are less effective in regions where reliable ground truth labels are scarce or unavailable [17]. In such cases, the results of feature selection or feature ranking can be associated with uncertainty. It is therefore better to use an approach that relies on the intrinsic structure of the data without requiring labeled samples. Such methods can be effectively applied in both greenfield and brownfield settings. In this context, unsupervised anomaly detection approaches are a suitable alternative, as they rank features based on the internal structure and statistical patterns within the geospatial data. In this regard, several studies have examined the use of AEs for unsupervised feature selection and evaluation. For example, Feng and Duarte [27] used a graph AE-based framework for unsupervised feature selection in which a single-layer AE was used to achieve nonlinear dimensionality reduction, demonstrating improved performance over linear approaches. In another study, Han, Wang [28] introduced an AE-based feature selector (AEFS) that integrates AE regression with group lasso regularization to capture linear and nonlinear feature relationships. Sharifipour, Fayyazi [29] proposed an unsupervised approach for ranking and selecting features using AEs, in which the average reconstruction error obtained by omitting each feature in turn, together with the feature's contribution to the latent bottleneck, is used for feature importance. Inspired by previous AE-based unsupervised frameworks, the present study adopts an unsupervised scheme based on a variational autoencoder (VAE) [30] to improve the evaluation and utilization of features derived from the conceptual model approach. The VAE compresses the input data into a constrained latent space and then reconstructs it. The reconstruction error of the input is then used for feature evaluation and selection, representing a practical method in this domain.

The proposed approach was applied to a dataset of porphyry copper deposits in the Shahr-e-Babak region of Iran (Figure 1) to identify the most relevant exploration features

for mineral exploration targeting. To assess the effectiveness of the proposed method, two modeling strategies were implemented. In the first strategy, all available features were integrated using an AE method. In the second strategy, only the features identified by the unsupervised VAE-based evaluation were integrated using the same AE. The resulting prospectivity models were then evaluated using the prediction-area (P-A) plot to compare their predictive performance.

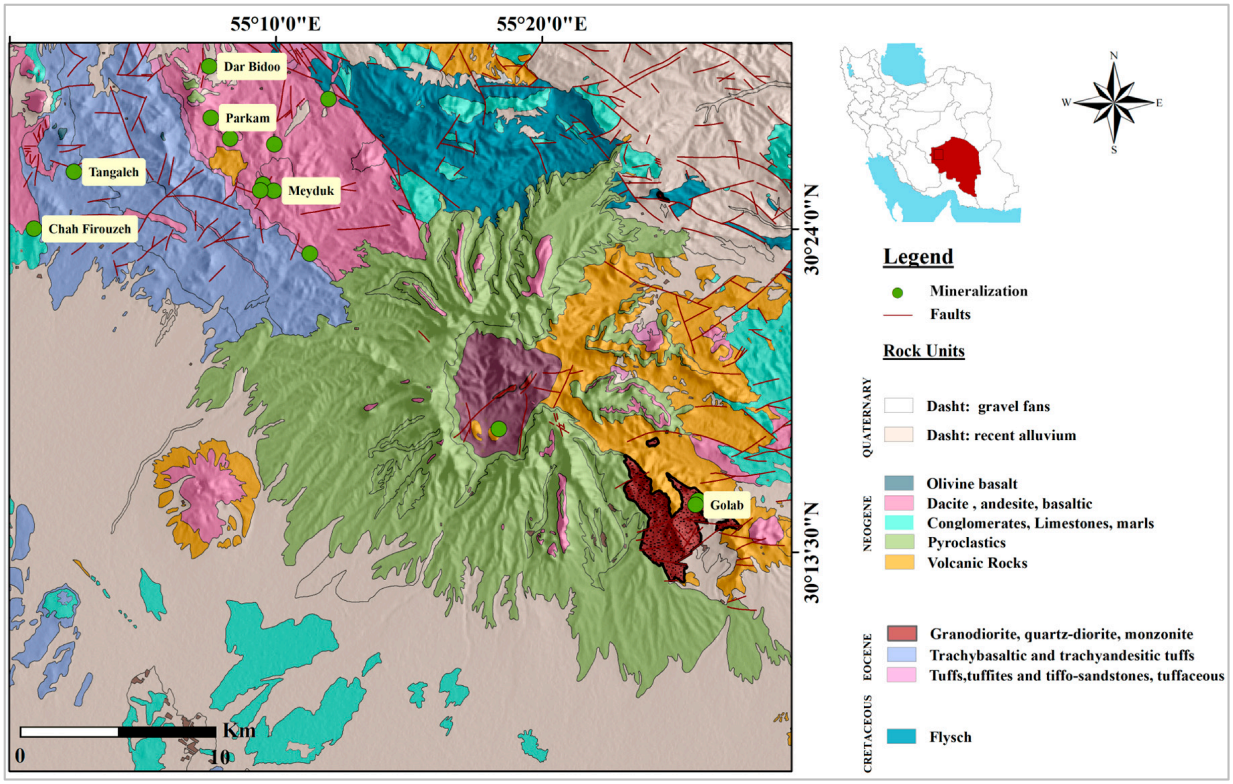

Figure 1. Geological map of the Shahr-e-Babak area, south Iran.

2. Materials and Methods

2.1. Geology of the Study Area

The study area is located within the 1:100,000-scale geological map of the Shahr-e-Babak district, situated in the Urumieh-Dokhtar Volcanic Belt (UDVB). This belt hosts several major porphyry copper deposits, such as Kahang [31], Sungun [32], Sarcheshmeh [33], and Meiduk [34]. The UDVB is subdivided into three segments: the Arasbaran porphyry copper belt in the northwest, the Saveh-Yazd porphyry copper belt in the central portion, and the Kerman porphyry copper belt in the southeast [35]. The Kerman porphyry copper belt, which extends roughly 500 Km in length and 80 Km in width, is the main copper-bearing region in Iran [36]. It comprises two distinct plutonic sub-belts: the northwestern Dehaj-Sarduieh zone, which hosts numerous porphyry copper deposits, and the southeastern Jebal-e-Barez zone [36]. Geologically, this belt consists predominantly of calc-alkaline volcanic, subvolcanic, and intrusive rocks.

Figure 1 illustrates the simplified geological map of the study area, corresponding to the Shahr-e-Babak 1:100,000 sheet in southeastern Iran. The western sector of the region is characterized by exposures of colored mélange units. The eastern portion is mainly composed of a thick Eocene conglomerate unit, accompanied by well-bedded, fine-grained sandstone and massive limestone. In the northern portion of the map, Eocene flysch, together with volcanic units, extends the area. The volcanic complex includes andesitic basalt, red tuff, and trachy-andesite and trachy-basalt in the lower

parts. In the southeastern part of the study area, Eocene quartz monzonite and granodiorite rocks are exposed. Although compositionally similar, the quartz monzonite contains a higher amount of K-feldspar compared to plagioclase. Younger andesitic agglomerates overlie volcanic units. Neogene volcanic rocks outcrop around the Kuh-e-Masahim volcano in the north. The northern portion of the caldera is occupied by an altered microdiorite intrusion, the youngest subvolcanic unit identified in the region. The mafic units within this complex are also altered and locally contain sulfide mineralization [37,38].

2.2. Exploration Datasets

The exploration datasets used in this study include a 1:100,000-scale geological map of the Shareh-e-Babak district, located in the UDVB, provided by the Geological Survey of Iran (GSI). This map contains digitized geological units, faults and fractures, and the locations of 14 known mineral occurrences, which were used as positive ground-truth labels for validation of prospectivity models. To identify geochemical anomalies related to porphyry copper deposits, we used a multi-element geochemical dataset comprising 585 stream sediment samples collected and analyzed by GSI using ICP-MS. In addition, remote sensing images were employed to delineate hydrothermal alteration zones related to porphyry copper mineralization.

Mappable Targeting Criteria

In this research, fifteen evidence layers and exploration proxies were constructed based on the regional-scale conceptual model of classic porphyry-type copper systems, following the approach used in previous works [37,39]. Porphyry copper deposits in Iran commonly show a close spatial association with felsic to intermediate intrusive rocks, which serve as the heat engine of the system, act as metal sources, and in some cases provide favorable geological sites for mineralization [39,40]. According to this spatial and genetic relationship, the proximity (Euclidean distance) to intrusive rocks was calculated (Figure 2), as areas situated near these intrusions generally offer a higher likelihood for the occurrence of mineralization. Moreover, structural features like faults are regarded as key components in modeling mineral exploration targets [41]. In many hydrothermal mineralization systems, including porphyry copper deposits, faults and fractures provide high-permeability pathways that allow high-temperature, metal-bearing hydrothermal fluids to migrate into the intrusive bodies and surrounding favorable rocks like volcanic units [22,42]. Following this rationale, and given that areas with abundant faults and fault intersections are generally more favorable for mineralization and thus valuable for exploration, proximity (Euclidean distance) maps to faults and fault intersections were also generated (Figure 3). Furthermore, to incorporate geochemical information into the modeling process, geochemical maps of Cu, As, W, Pb, Zn, Sb, and Mo were generated using IDW interpolation with a pixel size of 100 × 100 m to represent the post-mineralization subsystem (Figure 4). These elements are known to be effective indicators for identifying areas that are likely to host Cu-Mo, Cu-Mo-Au, and Cu-Au mineralization in the UDVB. In addition, porphyry copper deposits commonly show a strong spatial association with different types of hydrothermal alteration zones, as these zones provide essential exploration footprints, reflect physicochemical processes related to mineralization, and, in some cases, serve as sites of metal deposition [43,44]. These zones were mapped using ASTER and Landsat-8 OLI images through band ratio and relative absorption band depth (RBD) techniques (for more details, see [37]). Therefore, predictor maps representing proximity (Euclidean distance) to silicification, phyllic, argillic, propylitic, and iron oxide alteration zones

were generated (Figure 5). All generated evidence layers were finally rescaled using logistic fuzzification functions prior to integration [45–47].

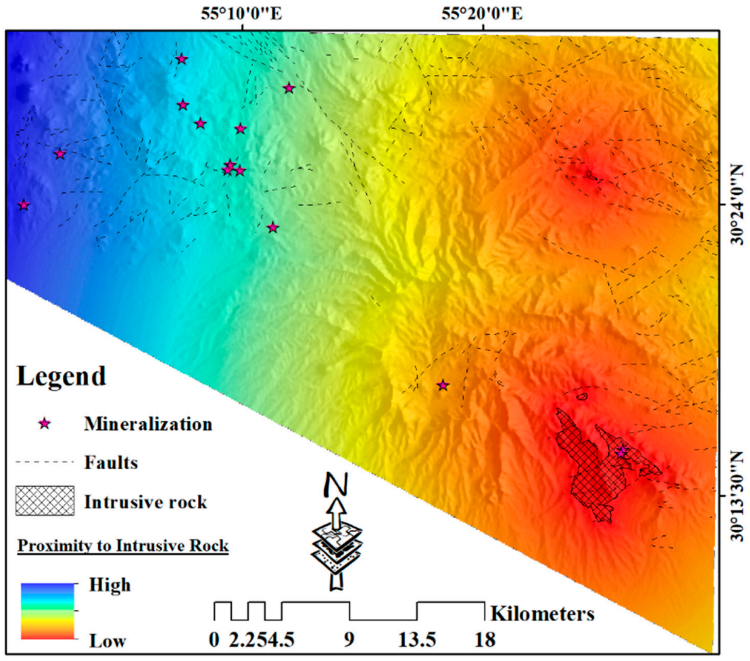

Figure 2. Proximity map to intrusive rocks in the study area.

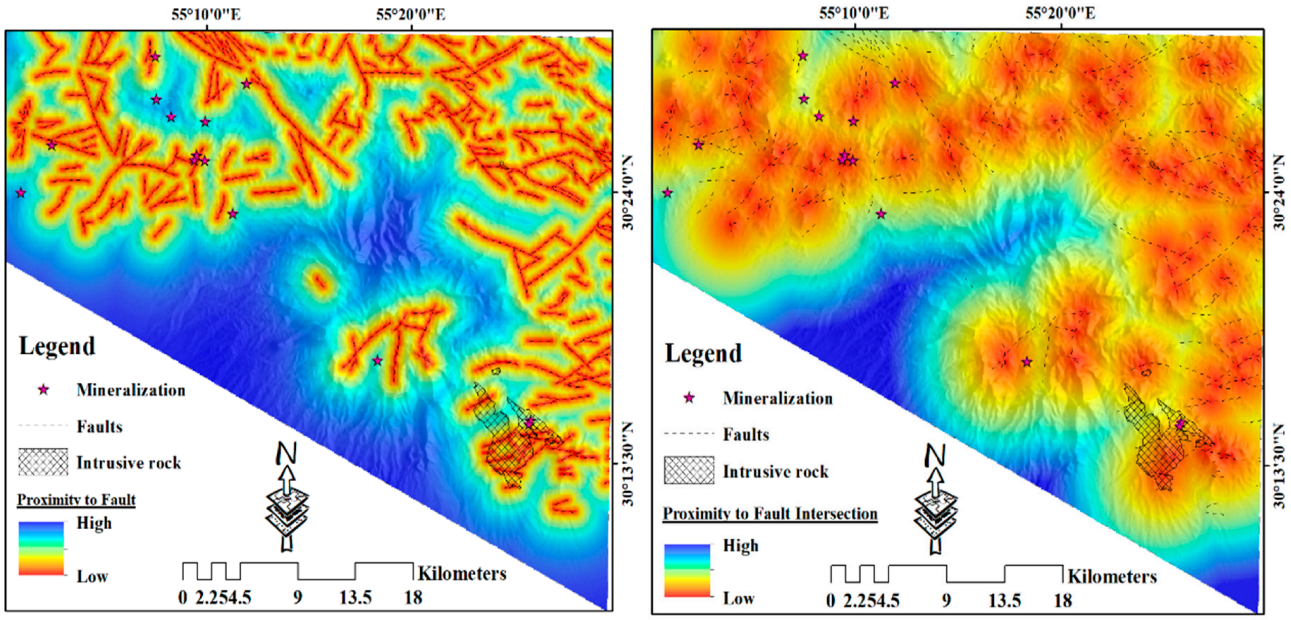

Figure 3. Proximity to faults and fault intersections in the study area.

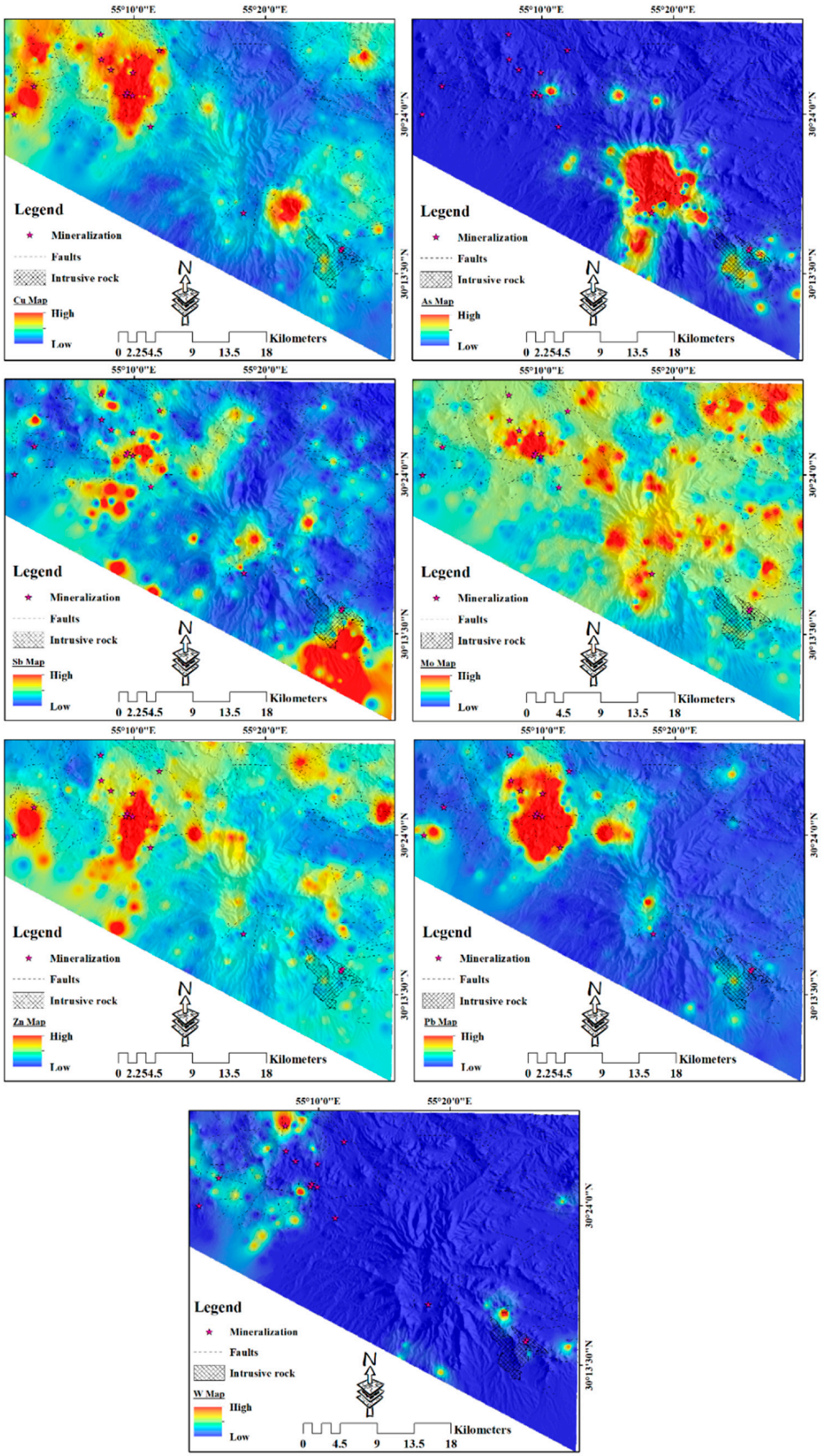

Figure 4. Geochemical maps of important elements in the study area.

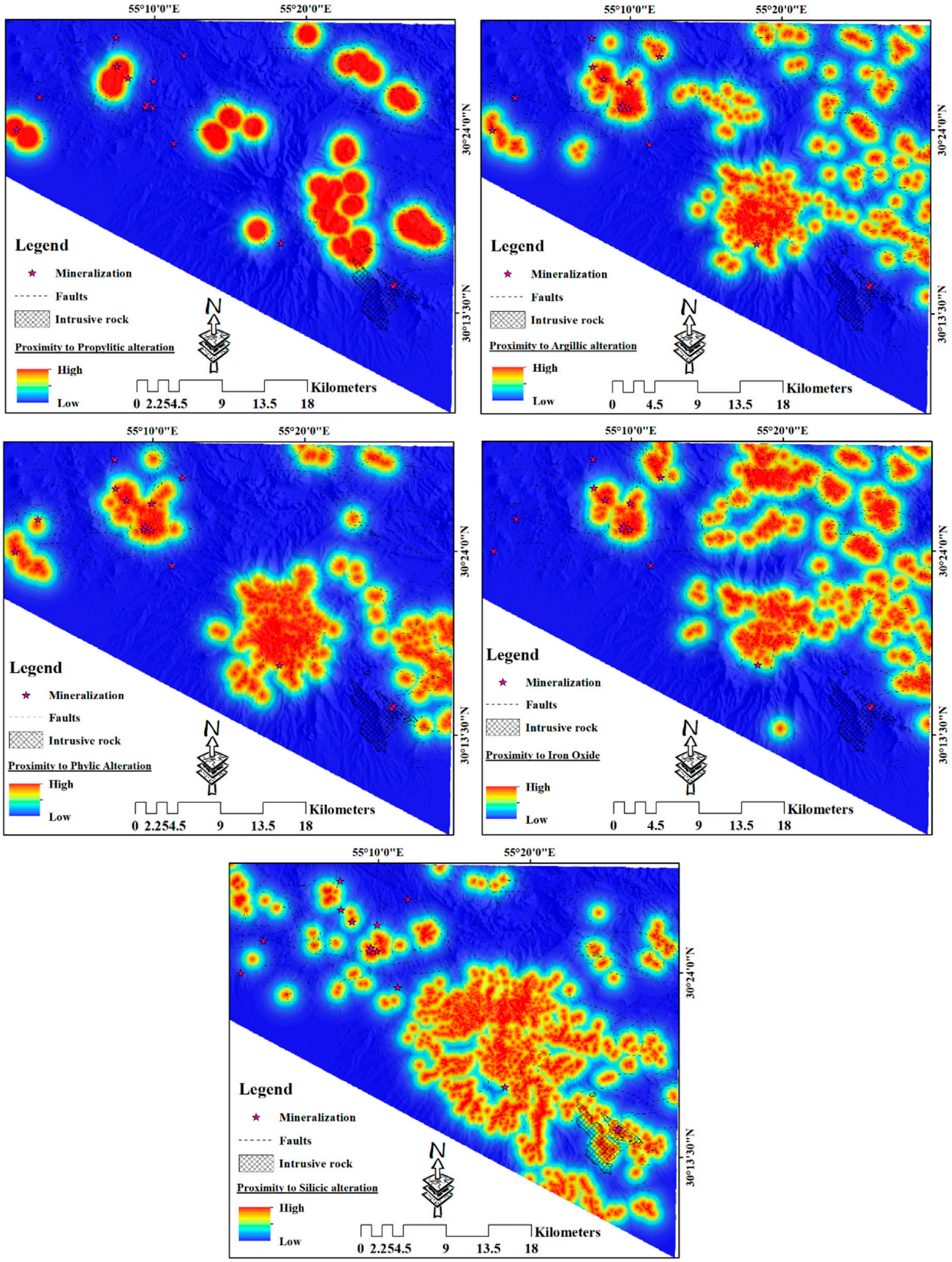

Figure 5. Proximity maps to hydrothermal alteration zones in the study area.

2.3. Methodology

This study adopted a five-step analytical procedure for mineral exploration targeting. First, the multi-source exploration datasets were processed to generate evidence layers representing different ore-forming subsystems. Next, a VAE was applied to all evidence layers

to rank them, identify the most informative exploration features, and reduce redundancy. Subsequently, the AE algorithm was applied to combine the evidence layers under two scenarios and generate the corresponding prospectivity maps. Finally, the prediction rate of the resulting prospectivity models was quantitatively evaluated using a P-A plot.

Two different modeling strategies were designed to investigate how the VAE-based strategy for selection of exploration features influences the prospectivity mapping results. In the first strategy, the AE model was applied to the full set of evidence layers derived from the conceptual model, allowing all available geo-information to contribute to the integration process. In the second strategy, a VAE was used to assess and rank the evidence layers and to provide guidance for selecting the most informative features. Based on this evaluation, a subset of evidence layers was chosen and then integrated using an AE model to generate another prospectivity map. Figure 6 illustrates the flowchart of the proposed approach.

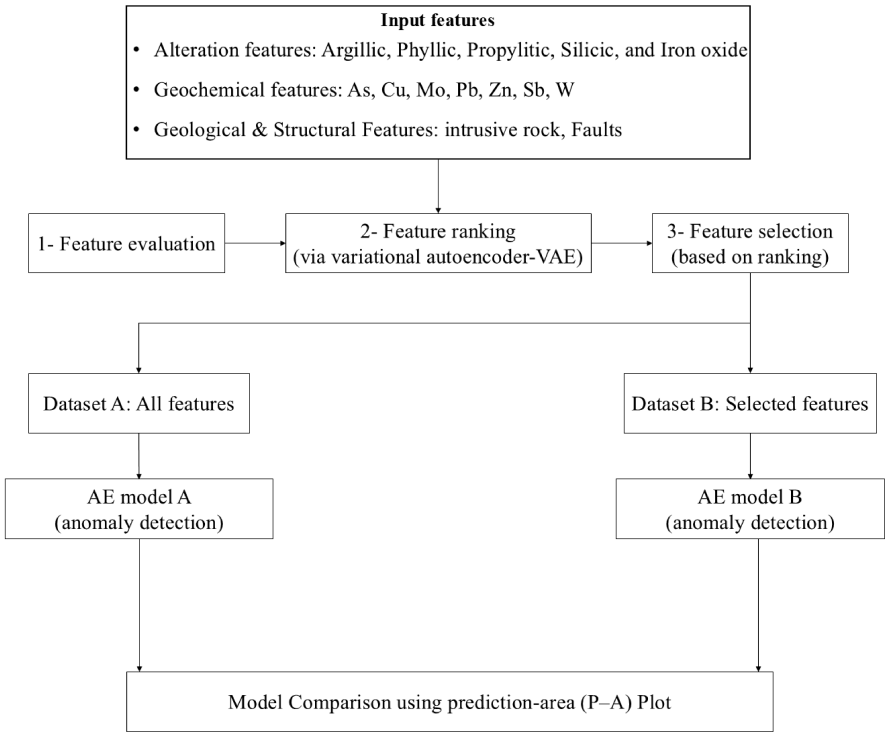

Figure 6. Overall workflow of the proposed DL-based mineral exploration targeting framework.

All analyses, including the VAE-based evaluation of evidence layers and the integration of features using an AE method, were implemented in Python (v3.9.25). The entire workflow was performed on a laptop equipped with a 7th-Generation Intel (R) Core (TM) i5 @ 2.50 GHz processor, 12 GB of installed physical memory (RAM), and an integrated NDVIDIA GTX graphics card. The operating system used was Microsoft Windows 10 Enterprise (64-bit).

2.3.1. Autoencoder (AE)

An AE is a type of neural network designed to learn a compact representation of input data by trying to reproduce the input as its outputs [48,49]. In its typical form, an AE consists of an encoder, a bottleneck layer, and a decoder. The encoder transforms the input data into a reduced, lower-dimensional representation that aims to capture the essential patterns. The bottleneck serves as the compressed code. The decoder then uses this code to reconstruct the input data. Both encoder and decoder are trained together to minimize the difference between the original input and its reconstruction.

The bottleneck's smaller size forces the network to focus on the most essential features of the input data while ignoring less important information. As a result, inputs that differ significantly from the typical patterns in the dataset are reconstructed poorly, leading to larger reconstruction errors. This property makes AE especially useful for identifying anomalies, as anomalous samples generally produce higher reconstruction errors than normal ones [17].

2.3.2. Variational Autoencoder (VAE)

Mathematically, in a VAE [30], the latent variable z is used to model the distribution of the original data set $\mathbf{X} = \{X_i\}_{i=1}^{N}$. The latent variable z, which follows a Gaussian distribution, can generate data that, through a neural network, can model any target distribution. By optimizing the generative parameters θ, the latent variable z can produce a data set $\hat{X} = \{\hat{x}_i\}_{i=1}^{N}$ that is very similar to the original data $X = \{X_i\}_{i=1}^{N}$. Therefore, our goal is to maximize the marginal likelihood $P_\theta(x)$ [50]:

$$P_\theta(x) = \int P_\theta(z)P_\theta(x|z)dz, \; with \; z \sim N(0,1), \tag{1}$$

The VAE introduces an inference (recognition) model $q_\varphi(z \mid x)$ to approximate the true posterior density, since the true posterior $P_\theta(z \mid x)$ cannot be computed exactly. The VAE measures the similarity between the recognition model $q_\varphi(z \mid x)$ and the true posterior distribution $P_\theta(z \mid x)$ using the Kullback–Leibler (KL) divergence. We have:

$$\log P_\theta\left(x^{(i)}\right) = D_{KL}\left(q_\varnothing\left(z\middle|x^{(i)}\right)\middle|\middle|P_\theta\left(z\middle|x^{(i)}\right)\right) + L\left(\theta, \varnothing; x^{(i)}\right), \tag{2}$$

Since the KL divergence is always non-negative, it follows that:

$$logP_\theta\left(x^{(i)}\right) > L\left(\theta, \varnothing; x^{(i)}\right) \tag{3}$$

The term $L\left(\theta, \varphi; x^{(i)}\right)$, which is known as the variational lower bound on the marginal likelihood of a data point i, can be written as

$$L\left(\theta, \varnothing; x^{(i)}\right) = -D_{KL}(q_\varnothing\left(z\middle|x^{(i)}\right)\middle|\middle|P_\theta(z) + E_{q_{(z|x^{(i)})}}\left[\log P_\theta\left(x^{(i)}\middle|z\right)\right], \tag{4}$$

To optimize $\log P_\theta\left(x^{(i)}\right)$, the variational lower bound on the marginal likelihood becomes the main optimization objective of the VAE. The first term on the right-hand side of Equation (4) acts as a regularization term, and the second term corresponds to a negative reconstruction error in the autoencoder formulation. Therefore, $q_\varphi\left(z \mid x^{(i)}\right)$ can be interpreted as a probabilistic encoder. The encoder is parameterized by the variational parameter φ, and $P_\theta\left(x^{(i)} \mid z\right)$ can be seen as a probabilistic decoder with a generative parameter θ. In general, the conditional distribution $P_\theta\left(x^{(i)} \mid z\right)$ is modeled either as a Bernoulli or a Gaussian distribution. In this study, because the network input consists of multivariate continuous data rather than binary data, the distribution $P_\theta\left(x^{(i)} \mid z\right)$ is assumed to be Gaussian. We then compute a stochastic gradient estimator of the variational lower bound $L\left(\theta, \varphi; x^{(i)}\right)$. After introducing the recognition model $q_\varphi\left(z \mid x^{(i)}\right)$, we use the reparameterization trick. Let z be a continuous random variable and $z \sim q_\varphi\left(z \mid x^{(i)}\right)$ a conditional distribution. By introducing an auxiliary noise variable $\varepsilon \sim p(\varepsilon)$, where $p(\varepsilon)$ has a known marginal distribution, and applying a transformation to $q_\varphi\left(z \mid x^{(i)}\right)$, we obtain $N(z;\, u,\, \sigma^2 I) = q_\varnothing\left(z\middle|x^{(i)}\right)$.

Since we assume that $q_\varphi\left(z \mid x^{(i)}\right)$ is Gaussian and that $p(z) = \mathcal{N}(z; 0, I)$, when the approximation $q_\varphi\left(z \mid x^{(i)}\right) = \mathcal{N}\left(z; \mu^{(i)}, \left(\sigma^{(i)}\right)^2 I\right)$ holds, the regularization term becomes:

$$-D_{KL}\left(q_\varphi\left(z\middle|x^{(i)}\right)\middle||p(z)\right) = \frac{1}{2}\sum_{j=1}^{j} 1 + \log\left(\sigma^{(i)^2}\right) - u^{(i)^2} - \sigma^{(i)^2}\Big), \tag{5}$$

where J is the dimension of z. When computing the reconstruction term using Monte Carlo estimation, we obtain:

$$E_{q_{\varnothing(z|x^{(i)})}}[\log p_\theta\left(x^{(i)}\middle|z\right) = \frac{1}{L}\sum_{l=}^{L}\log\left(p(x^{(i)}|z^{(i,j)})\right)), \tag{6}$$

2.3.3. Prediction-Area (P-A) Plot

A wide range of evaluation methods has been developed to assess the performance and prediction ability of prospectivity models. A key component of nearly all these assessment techniques is the spatial association between known mineral deposits and the modelled targets. Since minimizing the search area is a major objective in mineral exploration efforts, many assessment tools also account for the spatial extent occupied by exploration targets, because the likelihood of discovering new mineral deposits is generally higher within smaller target zones than in larger ones [16]. Among these methods, the P-A plot [21] is one of the most widely used and informative tools. The plot includes two curves: one representing the predication rate based on known mineral occurrences, and the other illustrating the occupied area. The intersection point of these curves provides a quantitative measure of model performance. A higher intersection point indicates a more powerful predictive capability, meaning that the prospectivity model has a stronger intrinsic association with the known mineralization.

3. Result and Analysis

3.1. Hyperparameter Optimization

In this study, the hyperparameters of the VAE model were tuned through a gradual and comparative evaluation process in order to obtain an optimal balance between training stability, reconstruction quality, and the ability to extract meaningful latent patterns. To determine the network architecture, several candidate structures, including symmetric and asymmetric configurations with different layer sizes, were tested. In the end, the configuration [64, 32] for the encoder and [32, 64] for the decoder was selected, because compared to smaller architectures (such as [32, 16]) and larger ones (such as [128, 64]), it provided the best trade-off between reducing reconstruction error and avoiding overfitting in the experimental tests. The results showed that this level of capacity is sufficient for the data used in this study and leads to smaller fluctuations in the KL loss. The latent space dimension was also chosen by testing several values, including 4, 8, 16, and 32. A value of latent_dim = 8 produced the lowest reconstruction error without causing an excessive increase in model complexity and led to more stable convergence in the loss curves. Therefore, this value was selected as the optimal setting.

For the training parameters, the number of epochs was set to 250 and the batch size to 32, based on monitoring the decrease in the reconstruction loss and KL loss. Initial experiments showed that the model had not fully converged before about 220 epochs, while after about 260 epochs, there was no noticeable improvement in performance. Thus, 250 epochs were chosen as a compromise between computational cost and model performance. In addition, the model was trained with different batch sizes (16, 32, and 64), and a batch size of 32 provided the best stability in the KL term.

To control the strength of the KL term and to obtain a suitable structure in the latent space, different values of β, including 0.5, 1.0, and 2.0, were examined. A value of β = 1.0 offered the best balance between preserving reconstruction quality and regularizing the latent space, whereas larger values led to an excessive reduction in reconstruction accuracy. The capacity parameter was also increased gradually (warm-up) to prevent collapse in the latent distribution. To ensure the reproducibility of the results, random_state was set to 42. All of the above values were chosen based on multiple runs, analysis of the loss curves, and comparison of model performance under different configurations.

In this method, our goal is to evaluate the importance of each data feature; that is, to understand which features have a stronger effect on the performance of the model when they are removed or perturbed. For this purpose, we use a VAE. This model is a type of AE that maps the input data into a compressed latent space and then reconstructs it. A key property of the VAE is that this compression is performed in a probabilistic manner, which makes it very suitable for feature selection and feature importance analysis. First, the VAE model is trained using the complete data set. During training, the model learns how to reconstruct the input data. After training, we feed the original data into the model and obtain the reconstructed outputs. The reconstruction error is then computed using metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and the KL divergence. These values are considered the baseline.

In the next step, we remove one input feature at a time by setting that feature to zero for all samples. We then feed this modified (incomplete) data set into the trained model and perform reconstruction again. The reconstruction errors are recalculated in this setting. For each feature, the difference between the new reconstruction error and the baseline error is measured. This difference indicates the importance of that feature: if removing a feature led to a noticeable increase in the error, that feature is considered important for the model. Finally, the results are visualized using plots to show which features have a stronger impact on the model accuracy. The main advantage of this approach is that it can evaluate feature importance in a fully automatic way, without the need for labels or supervised modeling. Therefore, this method is highly useful and effective for unsupervised problems and complex data sets.

3.2. Exploratory Data Analysis (EDA)

Before implementing the main model and applying the VAE for feature evaluation, the dataset was first examined using basic statistical methods in order to obtain an initial understanding of the structure of the data and the relationships among the features. This stage, known as EDA, plays a crucial role in guiding the subsequent modeling steps. In this phase, a set of statistical and graphical techniques was applied to identify patterns, correlations between variables, feature groupings, and the overall structure of the geospatial data. In particular, procedures such as computing correlation coefficients, plotting the correlation matrix as a heatmap, and using a dendrogram to represent the hierarchical clustering structure of the features were carried out.

The heatmap of the correlation matrix is shown in Figure 7. Argillic alteration exhibits a strong correlation with hydrothermal alterations such as iron oxide, phyllic, and propylitic alteration, with correlation coefficients of 0.72, 0.70, and 0.65, respectively, as reported in Figure 7. The spatial association of phyllic, argillic, propylitic, and iron oxide alterations is commonly a characteristic feature of porphyry copper systems, and together they form important geological footprints for vectoring toward mineralization zones.

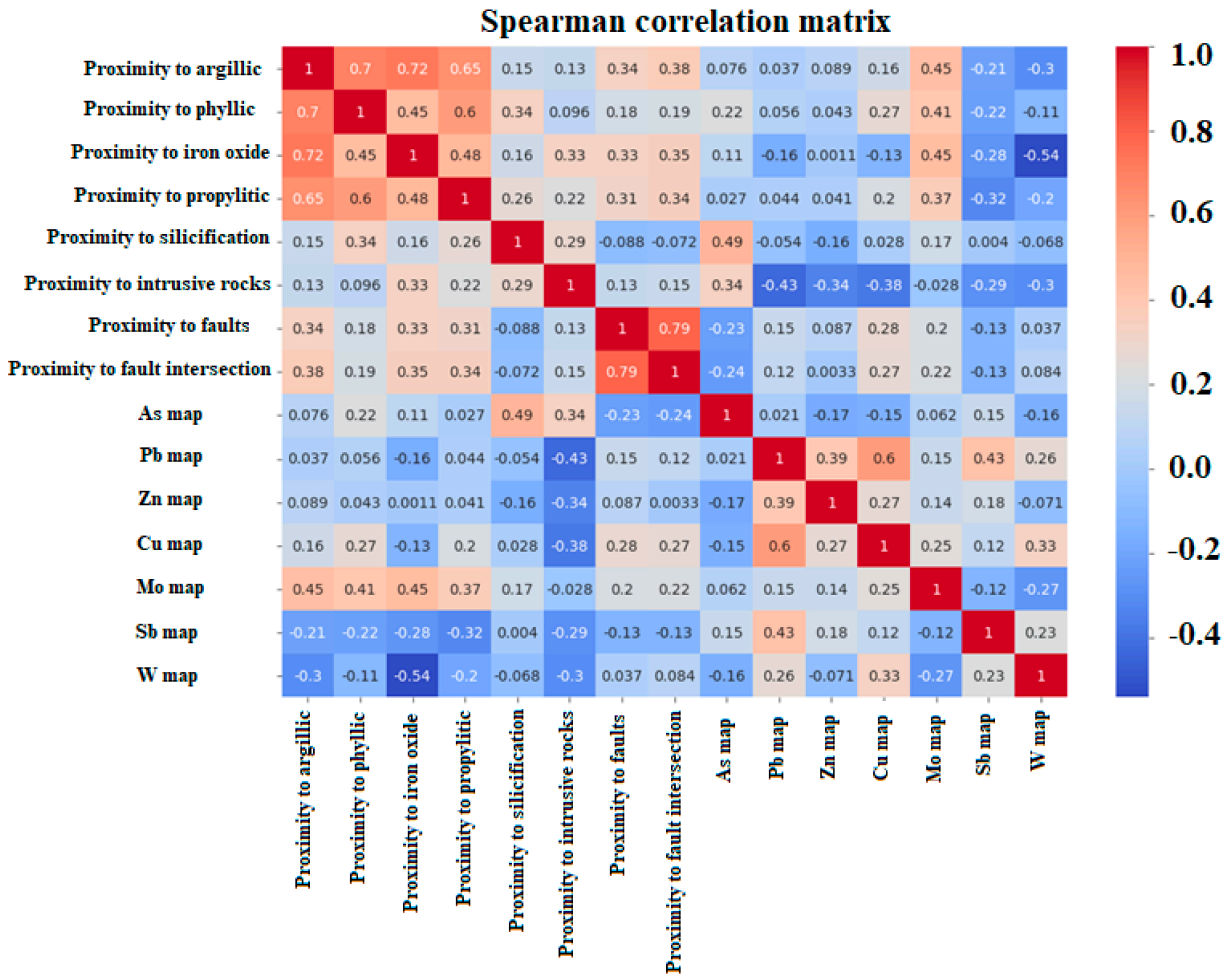

Figure 7. Correlation matrix heatmap of all features.

In addition, clear relationships are observed between Mo and propylitic, argillic, phyllic, and iron oxide alteration zones. This pattern indicates that Mo is closely associated with altered zones, which is consistent with the characteristics of porphyry copper systems. Moreover, the elements Pb, Cu, Zn, and Sb exhibit mutual correlations and are considered suitable pathfinder elements for targeting of porphyry copper systems.

Figure 8 presents the dendrogram corresponding to the evidence layers used in our analysis. According to this diagram, the layers are divided into three main clusters. The first cluster, shown in yellow, represents the alteration zones in the study area. The second cluster, shown in red, separates the geochemical elements from the other layers. The last cluster, shown in blue, includes the structural layers and the intrusive rocks, where the intrusive body is further distinguished from the structural layers in its own sub-branch.

3.3. Algorithm Implementation

After performing the basic statistical analyses and gaining an initial understanding of the data structure, an unsupervised learning model based on the VAE was used for feature evaluation and selection. The main objective of this analysis is to identify the features that play the most important role in the accurate reconstruction of the data. In this method, each feature is removed from the dataset in a stepwise manner, and the change in the reconstruction error, including MSE, MAE, and KL divergence, is measured to evaluate the importance of that feature. In practice, the larger the increase in reconstruction error after removing a feature, the more important that feature is for the model.

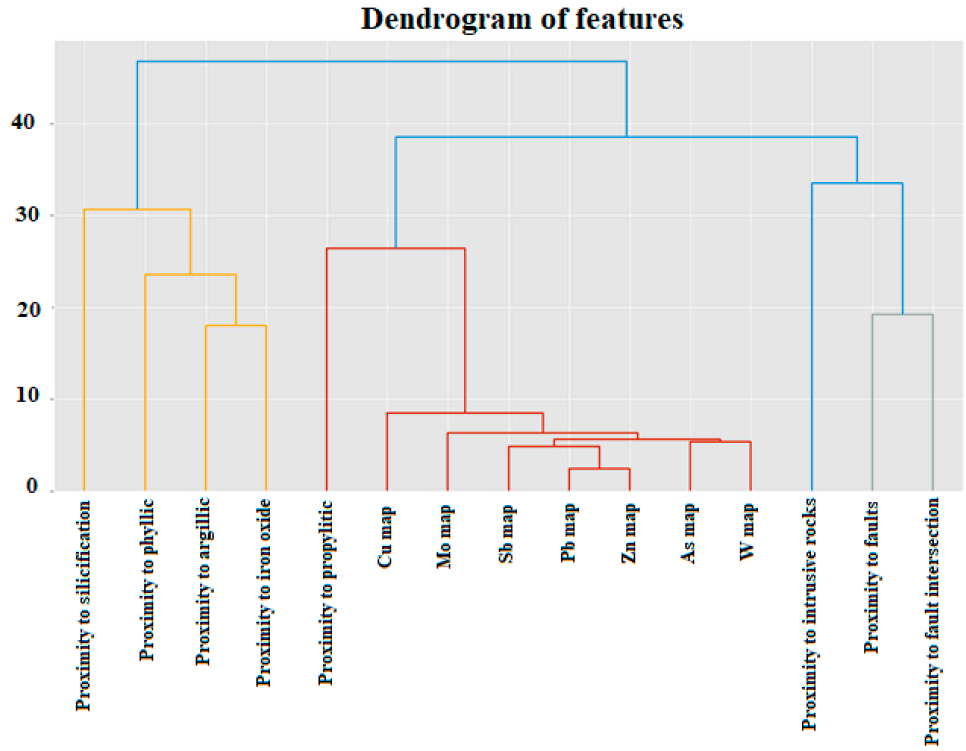

Figure 8. Dendrogram of the evidence layers used in the analysis.

Therefore, the difference between the errors in the full model (baseline) and in the model with the removed feature is considered as an indicator of the importance of that feature. In the following, comparison tables and plots of the results of this analysis are presented, in order to provide a clear view of the contribution of each feature to the reconstruction of the data by the VAE model. The presented table reports the results of feature evaluation based on the VAE model. In this analysis, the aim is to determine how the removal of each feature affects the accuracy of data reconstruction by the model. To this end, the VAE model is first trained using the complete data set, and the reconstruction errors, including MSE, MAE, and the KL divergence, are computed in the baseline state.

Next, in an iterative process, each feature is removed from the input data one at a time, and the new reconstruction errors are calculated. The absolute differences between the baseline values and the values obtained after removing each feature are then used as a measure of that feature's impact on the model performance. The three main columns of the table are defined as follows:

- $|\Delta$ MSE$|$: The absolute change in the MSE after removing the feature compared to the baseline. This value reflects the effect of feature removal on the reconstruction of fine details in the data.
- $|\Delta$ MAE$|$: The absolute change in the MAE, which is less sensitive to outliers and represents the overall reconstruction accuracy.
- $|\Delta$ KL$|$: The change in the KL divergence, indicating how the latent distribution of the data is altered after removing the feature.

The values in Table 1 show that some features play a more prominent role in data reconstruction. For example, the feature proximity to silicification alteration, with a $|\Delta$ MAE$|$ of 0.010638 and a $|\Delta$ KL$|$ of 1.556276, has the largest impact on the reconstruction error and can therefore be considered one of the most important features. Features such as proximity to iron oxide alteration and proximity to phyllic alteration also exhibit relatively large changes in the error metrics, indicating their significant contribution to the learned data representation. In contrast, features like Sb, Cu, and Pb

produce only small changes in the error values, suggesting that their role in the VAE-based reconstruction is limited and that they may contain redundant or overlapping information relative to other features (Figure 9). Overall, this analysis enables the ranking of features (Figure 10) according to their impact on the model and can be very useful for dimensionality reduction, feature selection, and the interpretation of the internal structure learned by the VAE.

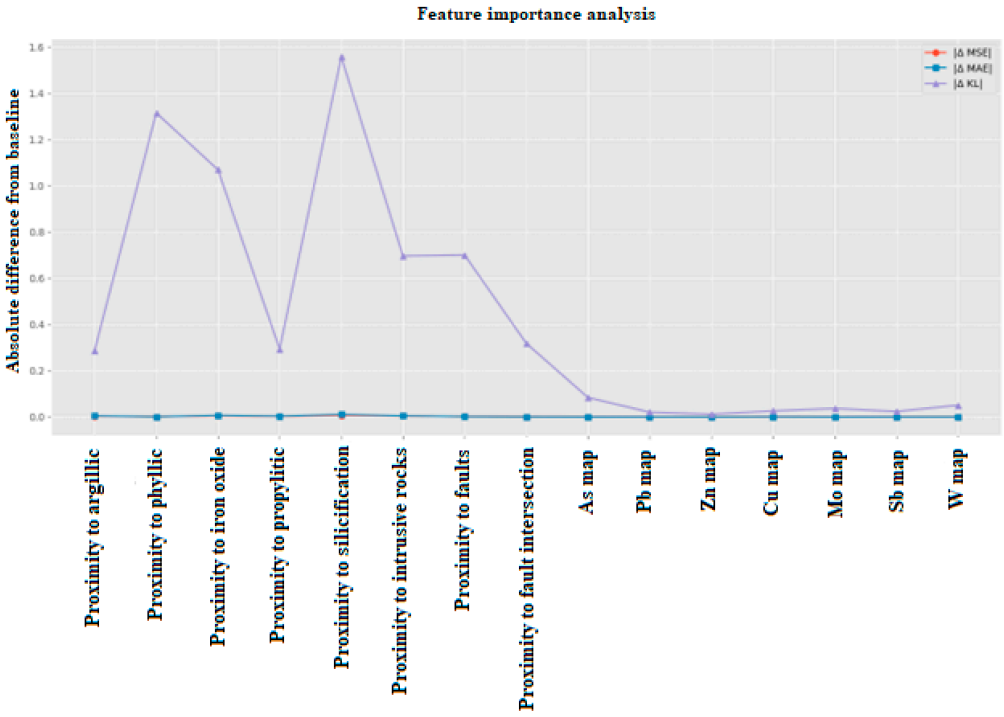

Figure 9. Feature importance visualization based on reconstruction error changes.

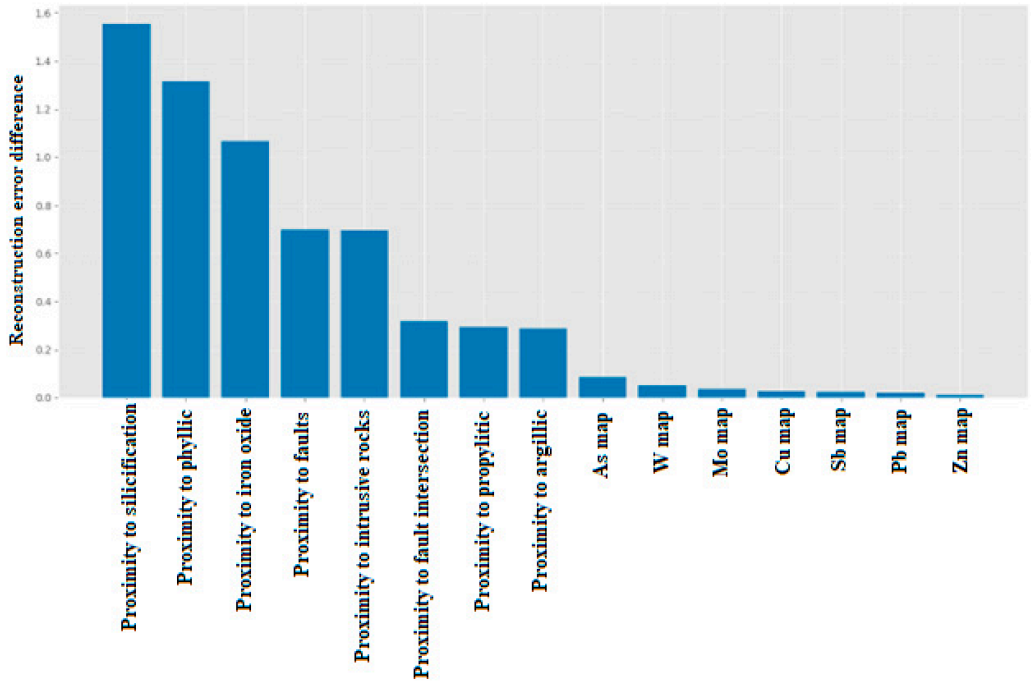

Figure 10. Ranked features based on their importance.

Table 1. Feature importance evaluation based on different criteria.

Feature	| Δ MSE |	| Δ MAE |	| Δ KL |
Proximity to argillic alteration	0.003255	0.004544	0.287477
Proximity to phyllic alteration	0.001065	0.000860	1.314490
Proximity to iron oxide alteration	0.005026	0.006900	1.068729
Proximity to propylitic alteration	0.002717	0.003516	0.293995
Proximity to silicification alteration	0.007186	0.010638	1.556276
Proximity to intrusive rocks	0.004565	0.004957	0.695872
Proximity to fault	0.001496	0.001531	0.699963
Proximity to fault intersection	0.000423	0.000610	0.318101
As	0.000143	0.000156	0.083974
Pb	0.000148	0.000253	0.020467
Zn	0.000092	0.000186	0.012031
Cu	0.000474	0.000840	0.026064
Mo	0.000072	0.000054	0.036769
Sb	0.000222	0.000339	0.023590
W	0.000194	0.000319	0.050811

In the plot shown in Figure 10, the features are sorted according to their importance, and based on this ranking, the most important features were selected for mineral exploration targeting.

3.4. Predictive Modeling and Performance Appraisal

The AE used in this study followed a 64–32–16 architecture. ReLU was used as the activation function. Batch normalization was applied, and the dropout rate was set to 0.1. The model was trained using the Adam optimizer with a learning rate of 0.001 and 250 epochs, with L2 regularization. A fixed random seed (42) was also used. After the feature evaluation stage, two modeling scenarios were adopted to generate prospectivity maps for vectoring toward porphyry copper mineralization zones. In the first scenario, all available evidence layers were integrated using the AE method (Figure 11). In the second scenario, the features were selected using a simple approach. First, an elbow-like drop in the importance scores was observed in Figure 10 and used as a guide. Then, the final features were chosen based on expert knowledge. These layers included proximity to argillic, phyllic, iron oxide, propylitic, and silicification alterations, proximity to intrusive rocks, proximity to faults, and proximity to fault intersections, and they were integrated using the same AE approach (Figure 12).

As shown in Figures 11 and 12, both models predict high prospectivity values primarily in the northwestern, northeastern, central, and southeastern parts of the study area. Most of these high-probability zones exhibit strong spatial associations with volcanic and intrusive rocks, as well as faults. Importantly, they show good spatial correspondence with the locations of known mineral occurrences. However, visual inspection indicates that the prospectivity map generated using all features often fails to predict high-probability areas for some known mineral occurrences in the northwestern part of the study area. In contrast, the model constructed using the selected features demonstrates improved performance in highlighting these known occurrences. Additionally, several new exploration targets were identified in different parts of the study area, offering new opportunities for exploration geologists to discover new mineral deposits. These newly delineated targets are often

located far from known mineral occurrences and can be justified using the survival bias concept [51].

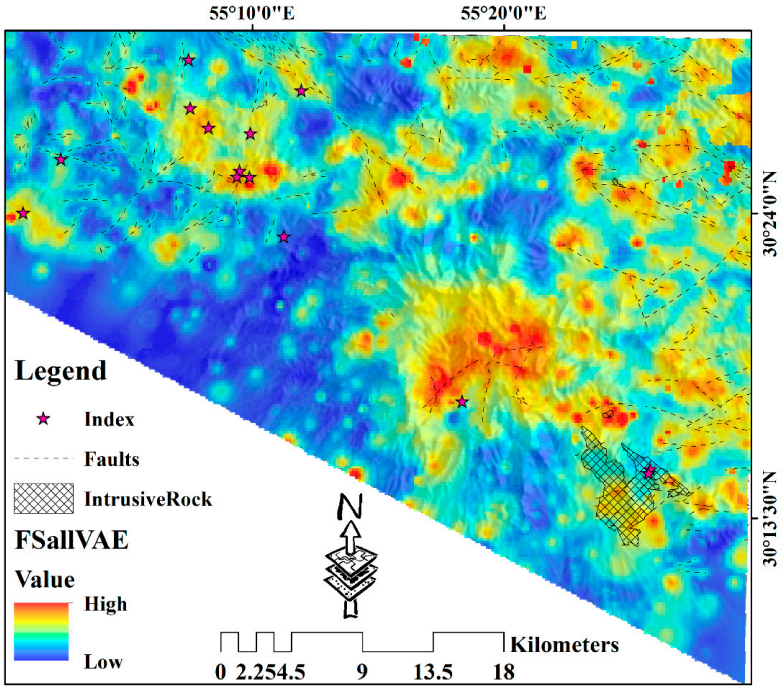

Figure 11. Porphyry copper prospectivity model generated using all features based on the first scenario for the study area.

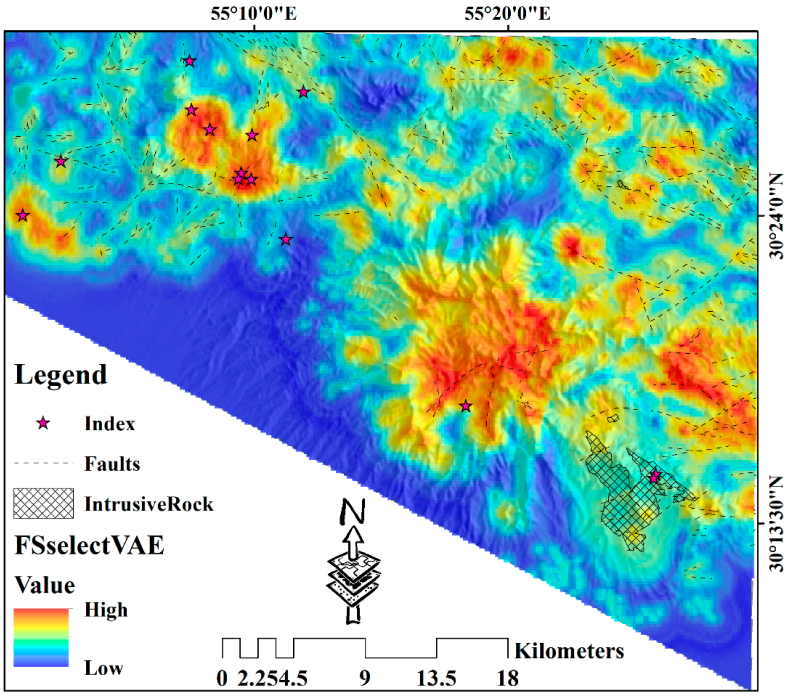

Figure 12. Porphyry copper prospectivity model produced using the selected features based on the second scenario for the study area.

To quantitatively compare the predictive performance of two models, P-A plots were constructed using 14 known mineral occurrences (Figures 13 and 14). The model using all features achieved a prediction rate of 72, whereas the feature-selection-based model achieved a higher prediction rate of 80. These results confirm that incorporating feature

evaluation leads to a more effective prospectivity model for copper exploration targeting in the study area.

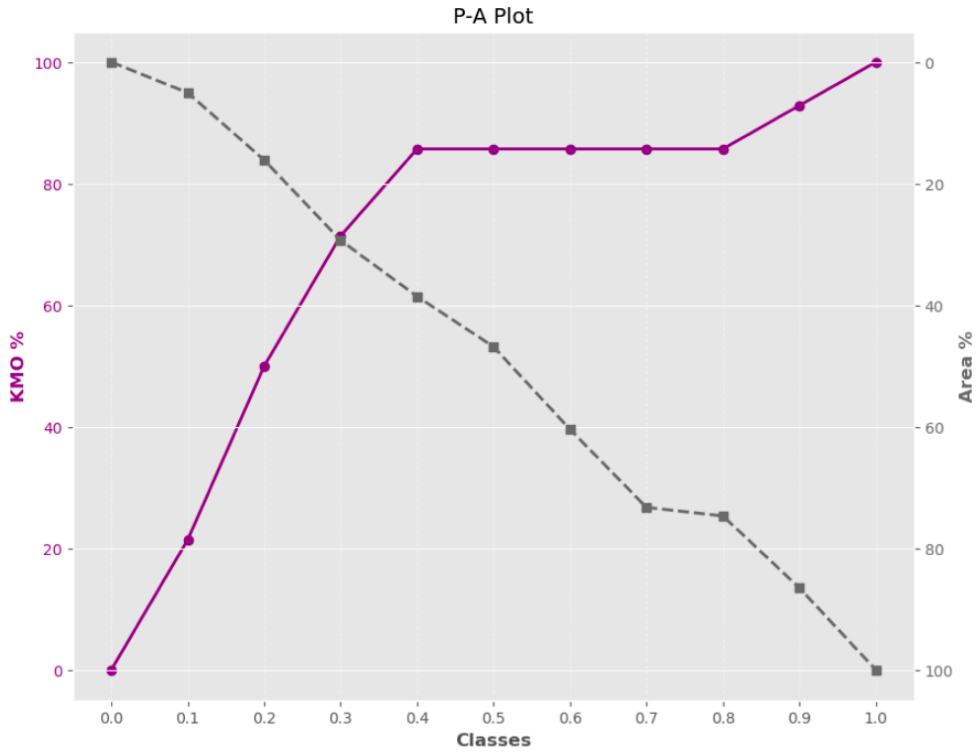

Figure 13. P-A plot for prospectivity model generated using all evidence layers (The dashed gray curve represents the prediction rate (KMO%), whereas the purple curve represents the area).

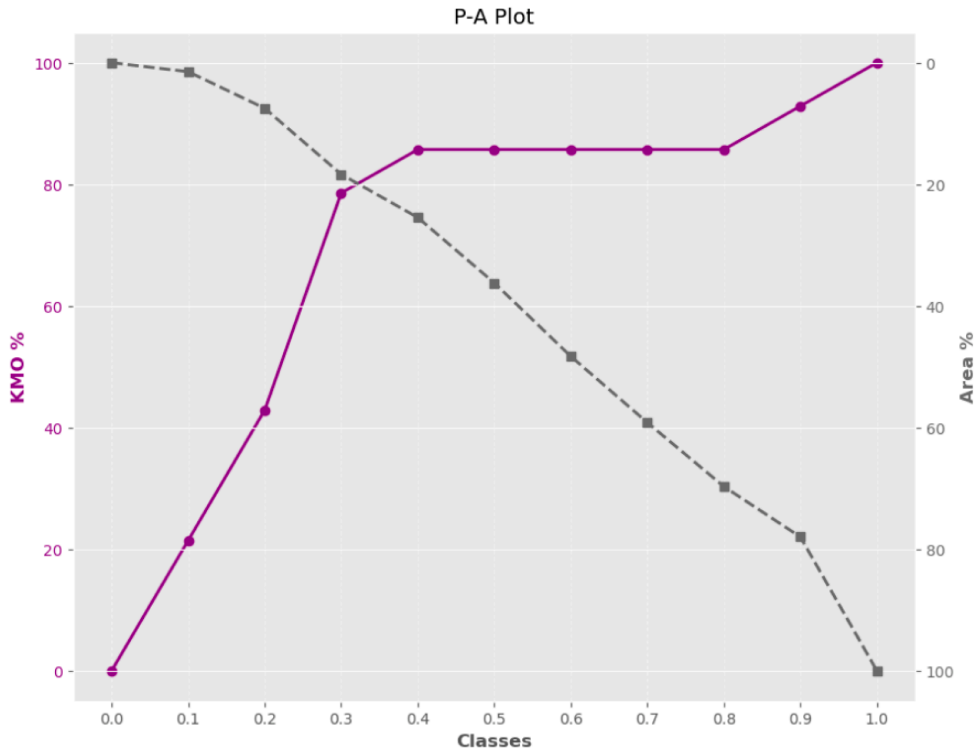

Figure 14. P-A plot for the prospectivity model generated using only the selected evidence layers (The dashed gray curve represents the prediction rate (KMO%), whereas the purple curve represents the area).

4. Discussion

The formation of mineral deposits is a rare and highly complex process [52], which makes the targeting of these deposits inherently challenging. This challenge arises from the intricate nature of ore-forming processes operating within the Earth's crust, the limited or even absent mineralization-related signals at the surface, and the superposition of multiple geological processes over time [53]. Together, these factors significantly complicate the prediction of mineral deposit locations. As a result, modern exploration targeting relies on the ability of exploration geologists to recognize the subtle signature of mineralization, particularly those associated with deep-seated systems [51]. Under these conditions, effective exploration targeting requires improved prospectivity analysis methods capable of detecting complex and nonlinear relationships within geoscience datasets [51].

In past decades, exploration geologists and mathematical geoscientists have adopted a variety of approaches, including ML and DL [4,54,55], as well as conceptual frameworks like exploration information systems (EIS) [8], to enhance mineral exploration targeting and to detect even subtle indicators of mineralization. Within this context, the present study employed a feature selection procedure based on VAE, which is capable of capturing intrinsic and complex relationships among the exploration features. This approach was utilized to make more effective use of the evidence layers and to enhance the predictive performance of the mineral prospectivity model.

The exploration features generated in this study are geologically meaningful and are well aligned with the conceptual model of porphyry copper deposits. The results of the feature analysis indicate that the hydrothermal alteration layers represent some of the most influential inputs for prospectivity modeling. This result is consistent with previous studies conducted in the study area, which have highlighted the critical role of hydrothermal alterations in identifying favorable areas for porphyry copper exploration [37,56]. Similarly, fault-related layers were recognized as important features. Tectonic activities play a fundamental role in the migration of ore-bearing hydrothermal fluids, and faults and fracture zones provide suitable permeable pathways for fluid flow and create favorable structural traps for metal deposition. For this reason, exploration geologists attach great importance to fault-related features in mineral exploration targeting, as widely documented in numerous mineral exploration studies [39]. Geological layers, such as proximity to intrusive rocks, were also identified as important inputs for prospectivity modeling. The role of such geological features has long been recognized, and they are known to exert a strong controlling factor on the spatial distribution of mineralization-related anomalies [40].

It should be noted that potassic alteration, one of the most important alteration types in porphyry copper systems, was not included among the evidence layers used in this study. In many porphyry copper systems, potassic alteration commonly occurs in the deeper parts of the mineralized system and may have limited surface expression, making its regional-scale identification difficult using remote sensing data. In regions where potassic alteration can be reliably identified, it could be incorporated into the proposed framework as an additional exploration feature. Given its close spatial association with porphyry copper mineralization, the inclusion of potassic alteration may further improve feature selection results and prospectivity mapping performance.

An interesting outcome of the proposed approach is that geochemical evidence layers of pathfinder elements were assigned relatively lower importance compared to hydrothermal alterations and structural features. This result indicates that although these evidence layers are important components of the mineral system (representing the post-mineralization subsystem) and have been widely and successfully used in previous mineral exploration studies, they are not the primary determinants of prospectivity analysis in this region. Several factors may explain this pattern, although these interpretations should

be regarded as hypotheses rather than definitive conclusions. A possible geological explanation for the higher ranking of hydrothermal alteration layers is that hydrothermal systems commonly exhibit a larger spatial footprint than geochemical anomalies detected in stream sediment. In porphyry copper mineralization, hydrothermal alteration halos may extend well beyond the boundaries of mineralization and can remain preserved even where surface geochemical signatures have been dispersed by geological and environmental processes. Recent studies have highlighted hydrothermal alteration footprints as important exploration vectors in the study area [57]. Stream sediment geochemical anomalies may be influenced by a variety of secondary geological and environmental processes, including but not limited to transport, drainage characteristics, weathering, sampling density, and dispersion processes. These factors can reduce their relative contribution to predictive modeling. The lower ranking of geochemical layers should not be interpreted as evidence that these layers are unimportant for mineral exploration targeting; rather, it suggests that, within the current geospatial dataset and study area, alteration-related and structural features provide more robust predictors.

Although integrating all available evidence layers is conceptually justifiable, such models do not necessarily achieve a higher prediction rate. The results of this study demonstrate that, within a given region, only a subset of exploration features plays a dominant role in predictive modeling, and identifying these key features is critical for effective exploration targeting. In this context, the proposed feature selection framework provides a practical and effective solution. By filtering out less informative variables and retaining the most influential features, the proposed method produces an improved prospectivity model compared to the scenario in which all evidence layers are used. The resulting prospectivity model shows better spatial correspondence with the locations of known mineral occurrences and also achieves a higher prediction rate, as confirmed by the P-A plot. Also, an important novelty of the proposed framework lies in its unsupervised nature, which enables its application in both brownfield and greenfield exploration settings. In contrast to conventional expert-based approaches, the proposed approach in this study is capable of learning complex and nonlinear relationships among exploration features. It can capture hidden patterns that may not be readily identifiable through expert interpretation alone. This capability is particularly valuable in greenfield areas, where information on known mineral occurrences is limited or unavailable. In such environments, the proposed method provides a data-driven approach for identifying the most informative exploration features. It can assist exploration geologists in selecting relevant evidence layers and developing more effective data-integration strategies for mineral exploration targeting.

In a previous study conducted in this area [3], the evidence layers were generated and selected based on the conceptual model of porphyry copper mineralization, but proximity to intrusive rocks was not incorporated into their prospectivity analysis. However, in the present study, this evidence layer was re-examined within the VAE framework. Although the host rock layer does not exhibit a strong spatial correlation with known mineral occurrences, the feature-evaluation results indicate that it is one of the most influential exploration features in the model. This finding shows that exploration geologists should first construct a comprehensive suite of evidence layers based on the mineral system framework and then evaluate them quantitatively in order to determine the most suitable layers for prospectivity modeling.

Numerous studies have been carried out in the study area. The results of the present study indicate that the spatial extent of the identified anomaly zones is smaller than that reported in previous studies [37,38]. From an exploration perspective, this is a favorable outcome, as it reduces the exploration search space and enables more focused and cost-effective mineral exploration targeting. These results confirm that our proposed feature

selection strategy can enhance the predictive performance of mineral prospectivity models, making it a valuable tool for implementation within EIS [8] for the exploration of the next generation of mineral deposits.

Despite these promising results, several limitations should be considered. Since the proposed framework was developed and validated within a specific study area, its application to other geological settings requires further investigation. The framework was assessed using a single porphyry copper district. Although the results are promising, additional implementation across different deposit types and geological environments is required. Applying this model to new geo-datasets will likely necessitate renewed hyperparameter optimization, as the optimal architecture and training settings are inherently data-dependent. In addition, a notable limitation of the current feature-ranking strategy is the absence of a mathematical threshold for feature selection. For scenario 2, feature selection followed a hybrid approach, integrating the observed sharp decline in the importance scores (Figure 10), identifiable as an elbow, with expert knowledge. While effective, this approach may introduce a degree of subjectivity. Consequently, defining a data-driven criterion for feature selection remains an essential objective for future studies to enhance the proposed framework. Also, due to the limited availability of certain exploration datasets in the study area, particularly geophysical data, evidence layers derived from these data were not included in the present analysis. The absence of such information may have influenced the final feature rankings and prospectivity mapping results. Future studies should investigate the incorporation of additional exploration datasets to further evaluate the effectiveness of the proposed framework.

5. Conclusions

In this study, the effectiveness of an unsupervised DL-based feature selection approach for mineral exploration targeting was evaluated. Based on the obtained results and analyses, the following key findings and implications can be highlighted:

- The results indicate that our VAE-based feature selection strategy can effectively rank, evaluate, and identify the most influential exploration features for mineral exploration targeting.
- The feature evaluation results indicate that hydrothermal alteration, geological, and fault-related layers are more effective for prospectivity modeling in the study area, whereas the geochemical layers show lower effectiveness.
- By prioritizing the most informative evidence layers, the proposed approach allows exploration geologists to develop more efficient exploration strategies aimed at locating porphyry copper systems.
- The exploration targets for porphyry copper mineralization identified in the study area using this approach can serve as favorable areas for follow-up prospecting.
- Future research should focus on applying the proposed unsupervised feature selection framework to different mineral deposit types and geological settings, as well as comparing its performance with other unsupervised and related feature selection approaches to evaluate its capacities further.

Author Contributions: Conceptualization, M.S. and Z.H.; Software, M.S. and Z.H.; Formal analysis, M.S. and A.S. (Aref Shirazi); Investigation, Z.H.; Resources, M.S.; Data curation, Z.H.; Writing—original draft, M.S., Z.H. and A.S. (Aref Shirazi); Writing—review & editing, A.S. (Adel Shirazy), A.S. (Aref Shirazi) and A.B.P.; Visualization, Z.H.; Supervision, A.B.P.; Project administration, A.S. (Adel Shirazy); Funding acquisition, A.S. (Adel Shirazy) and A.S. (Aref Shirazi). All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding authors.

Acknowledgments: The authors would like to thank Seyyed Ataollah Agha Seyyed Mirzabozorg for his valuable help in improving this manuscript.

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

References

1. Mostafaei, K.; Yousefi, M.; Kreuzer, O.; Kianpour, M.N. Simulation-based mineral prospectivity modeling and Gray Wolf optimization algorithm for delimiting exploration targets. *Ore Geol. Rev.* **2025**, *177*, 106458. [CrossRef]
2. Mirzabozorg, S.A.A.S.; Saremi, M.; DehghanNiri, R.; Abedi, M.; Yousefi, M.; Pour, A.B.; Hezarkhani, A.; Maghsoudi, A. Hyperparameter optimization in unsupervised anomaly detection for mineral prospectivity mapping. *Ore Geol. Rev.* **2025**, *181*, 106627. [CrossRef]
3. Saremi, M.; Hoseinzade, Z.; Yousefi, M. A deep embedded clustering algorithm in conjunction with an ensemble technique for mineral prospectivity mapping. *Sci. Rep.* **2025**, *15*, 38086. [CrossRef] [PubMed]
4. Li, Q.; Chen, G.; Wang, D. Mineral prospectivity mapping using semi-supervised machine learning. *Math. Geosci.* **2024**, *57*, 275–305. [CrossRef]
5. Harris, J.R.; Wilkinson, L.; Heather, K.; Fumerton, S.; Bernier, M.A.; Ayer, J.; Dahn, R. Application of GIS processing techniques for producing mineral prospectivity maps—A case study: Mesothermal Au in the Swayze Greenstone Belt, Ontario, Canada. *Nat. Resour. Res.* **2001**, *10*, 91–124. [CrossRef]
6. Bonham-Carter, G.F.; Agterberg, F.P.; Wright, D.F. Weights of evidence modeling: A new approach to mapping mineral potential. In *Statistical Applications in the Earth Sciences*; Agterberg, F.P., Bonham-Carter, G.F., Eds.; Geological Survey of Canada: Ottawa, ON, Canada, 1989; pp. 171–183.
7. Yousefi, M.; Carranza, E.J.M.; Kreuzer, O.P.; Nykänen, V.; Hronsky, J.M.; Mihalasky, M.J. Data analysis methods for prospectivity modelling as applied to mineral exploration targeting: State-of-the-art and outlook. *J. Geochem. Explor.* **2021**, *229*, 106839.
8. Yousefi, M.; Kreuzer, O.P.; Nykänen, V.; Hronsky, J.M. Exploration information systems—A proposal for the future use of GIS in mineral exploration targeting. *Ore Geol. Rev.* **2019**, *111*, 103005. [CrossRef]
9. Bahrami, Y.; Hassani, H.; Maghsoudi, A. BWM-ARAS: A new hybrid MCDM method for Cu prospectivity mapping in the Abhar area, NW Iran. *Spat. Stat.* **2019**, *33*, 100382. [CrossRef]
10. Hajihosseinlou, M.; Maghsoudi, A.; Ghezelbash, R. A semi-supervised learning framework for intelligent mineral prospectivity mapping: Incorporation of the CatBoost and Gaussian mixture model algorithms. *J. Geochem. Explor.* **2025**, *274*, 107755. [CrossRef]
11. Qaderi, S.; Maghsoudi, A.; Yousefi, M.; Pour, A.B. Translation of mineral system components into time step-based ore-forming events and evidence maps for mineral exploration: Intelligent mineral prospectivity mapping through adaptation of recurrent neural networks and random forest algorithm. *Ore Geol. Rev.* **2025**, *179*, 106537. [CrossRef]
12. Soltani, Z.; Hassani, H.; Esmaeiloghli, S. Multi-model decision system: An ensemble deep learning model to enhance predictive power in mineral prospectivity mapping. *Ore Geol. Rev.* **2025**, *184*, 106768. [CrossRef]
13. McCuaig, T.C.; Beresford, S.; Jon Hronsky, M.A. Translating the mineral systems approach into an effective exploration targeting system. *Ore Geol. Rev.* **2010**, *38*, 128–138. [CrossRef]
14. Keykhay-Hosseinpoor, M.; Porwal, A.; Kalimuthu, R.; Panja, S. Mineral system modeling of Lithium-Cesium-Tantalum (LCT)-type pegmatites: Regional-scale exploration targeting and uncertainty analysis in the Sanandaj-Sirjan zone, Western Iran. *J. Geochem. Explor.* **2025**, *279*, 107885. [CrossRef]
15. Yousefi, M.; Carranza, E.J.M. Geometric average of spatial evidence data layers: A GIS-based multi-criteria decision-making approach to mineral prospectivity mapping. *Comput. Geosci.* **2015**, *83*, 72–79. [CrossRef]
16. Yousefi, M.; Lindsay, M.D.; Kreuzer, O. Mitigating uncertainties in mineral exploration targeting: Majority voting and confidence index approaches in the context of an exploration information system (EIS). *Ore Geol. Rev.* **2024**, *165*, 105930. [CrossRef]
17. Mirzabozorg, S.A.A.S.; Abedi, M. Recognition of mineralization-related anomaly patterns through an autoencoder neural network for mineral exploration targeting. *Appl. Geochem.* **2023**, *158*, 105807. [CrossRef]
18. Yousefi, M.; Carranza, E.J.M. Data-driven index overlay and Boolean logic mineral prospectivity modeling in greenfields exploration. *Nat. Resour. Res.* **2015**, *25*, 3–18. [CrossRef]
19. Nykänen, V.; Lahti, I.; Niiranen, T.; Korhonen, K. Receiver operating characteristics (ROC) as validation tool for prospectivity models—A magmatic Ni–Cu case study from the Central Lapland Greenstone Belt, Northern Finland. *Ore Geol. Rev.* **2015**, *71*, 853–860.
20. Agterberg, F.P.; Bonham-Carter, G.F. Measuring the performance of mineral-potential maps. *Nat. Resour. Res.* **2005**, *14*, 1–17. [CrossRef]

21. Yousefi, M.; Carranza, E.J.M. Prediction–area (P–A) plot and C–A fractal analysis to classify and evaluate evidential maps for mineral prospectivity modeling. *Comput. Geosci.* **2015**, *79*, 69–81.
22. Daviran, M.; Maghsoudi, A.; Ghezelbash, R.; Pradhan, B. A new strategy for spatial predictive mapping of mineral prospectivity: Automated hyperparameter tuning of random forest approach. *Comput. Geosci.* **2021**, *148*, 104688. [CrossRef]
23. Lachaud, A.; Marcus, A.; Vučetić, S.; Mišković, I. Study of the Influence of Non-Deposit Locations in Data-Driven Mineral Prospectivity Mapping: A Case Study on the Iskut Project in Northwestern British Columbia, Canada. *Minerals* **2021**, *11*, 597. [CrossRef]
24. Sadr, M.; Nazeri, M. Random forests algorithm in podiform chromite prospectivity mapping in Dolatabad area, SE Iran. *J. Min. Environ.* **2018**, *9*, 403–416. [CrossRef]
25. Parsa, M.; Maghsoudi, A.; Yousefi, M. Spatial analyses of exploration evidence data to model skarn-type copper prospectivity in the Varzaghan district, NW Iran. *Ore Geol. Rev.* **2018**, *92*, 97–112. [CrossRef]
26. Ribeiro, B.O.L.; Barbuena, D.; de Melo, G.H.C.; Motta, J.G.; Marques, E.D.; Marinho, M.d.S. How do non-deposit sites influence the performance of machine learning-based gold prospectivity mapping? A study case in the Pitangui Greenstone Belt, Brazil. *J. Geochem. Explor.* **2024**, *264*, 107543. [CrossRef]
27. Feng, S.; Duarte, M.F. Graph autoencoder-based unsupervised feature selection with broad and local data structure preservation. *Neurocomputing* **2018**, *312*, 310–323. [CrossRef]
28. Han, K.; Wang, Y.; Zhang, C.; Li, C.; Xu, C. Autoencoder inspired unsupervised feature selection. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*; IEEE: Piscataway, NJ, USA, 2018.
29. Sharifipour, S.; Fayyazi, H.; Sabokrou, M.; Adeli, E. Unsupervised feature ranking and selection based on autoencoders. In *ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*; IEEE: Piscataway, NJ, USA, 2019.
30. Kingma, D.P.; Welling, M. Auto-encoding variational bayes. *arXiv* **2013**, arXiv:1312.6114.
31. Afshooni, S.; Mirnejad, H.; Esmaeily, D.; Haroni, H.A. Mineral chemistry of hydrothermal biotite from the Kahang porphyry copper deposit (NE Isfahan), Central Province of Iran. *Ore Geol. Rev.* **2013**, *54*, 214–232. [CrossRef]
32. Hezarkhani, A.; Williams-Jones, A.E. Controls of alteration and mineralization in the Sungun porphyry copper deposit, Iran; evidence from fluid inclusions and stable isotopes. *Econ. Geol.* **1998**, *93*, 651–670. [CrossRef]
33. Boomeri, M.; Nakashima, K.; Lentz, D.R. The Sarcheshmeh porphyry copper deposit, Kerman, Iran: A mineralogical analysis of the igneous rocks and alteration zones including halogen element systematics related to Cu mineralization processes. *Ore Geol. Rev.* **2010**, *38*, 367–381. [CrossRef]
34. Aliani, F.; Alirezaei, A.; Moradian, A.; Abbasloo, Z. Geochemistry and petrography of the Meiduk porphyry copper deposit, Kerman, Iran. *Aust. J. Basic Appl. Sci.* **2009**, *3*, 3786–3800.
35. Aghazadeh, M.; Hou, Z.; Badrzadeh, Z.; Zhou, L. Temporal–spatial distribution and tectonic setting of porphyry copper deposits in Iran: Constraints from zircon U–Pb and molybdenite Re–Os geochronology. *Ore Geol. Rev.* **2015**, *70*, 385–406. [CrossRef]
36. Etemadi, A.; Karimpour, M.H. Geological constraints on magmatic evolution in subduction zones and cumulative factors effective on the fertility of Cenozoic host porphyritic rocks associated with major porphyry copper deposits in the Lut Block and Kerman porphyry copper belt, Iran. *J. Asian Earth Sci. X* **2022**, *7*, 100081. [CrossRef]
37. Hoseinzade, Z.; Saremi, M.; Shojaei, M.; Mokhtari, A.R.; Pour, A.B.; Mirzabozorg, S.A.A.S.; Hezarkhani, A.; Maghsoudi, A.; Yousefi, S. Fusion of remote sensing and geochemical data using hybrid Variational Autoencoder- BIRCH deep learning algorithm for copper prospectivity mapping. *Remote Sens. Appl. Soc. Environ.* **2025**, *40*, 101738. [CrossRef]
38. Jahantigh, M.; Ramazi, H. Porphyry copper prospectivity modelling using data driven and hybrid outranking methods: A case study of Shahr-e-Babak study area, South Eastern Iran. *J. Afr. Earth Sci.* **2025**, *232*, 105792.
39. Yousefi, M.; Hronsky, J.M. Translation of the function of hydrothermal mineralization-related focused fluid flux into a mappable exploration criterion for mineral exploration targeting. *Appl. Geochem.* **2023**, *149*, 105561. [CrossRef]
40. Hezarkhani, A. Petrology of the intrusive rocks within the Sungun porphyry copper deposit, Azerbaijan, Iran. *J. Asian Earth Sci.* **2006**, *27*, 326–340. [CrossRef]
41. Saremi, M.; Yousefi, S.; Yousefi, M. Combination of Geochemical and Structural Data to Determine Exploration Target of Copper Hydrothermal Deposits in Feizabad District. *J. Min. Environ.* **2024**, *15*, 1089–1101.
42. Ghasemzadeh, S.; Maghsoudi, A.; Yousefi, M.; Kreuzer, O. Spatially weighted singularity mapping in conjunction with random forest algorithm for mineral prospectivity modeling. *Int. J. Min. Geo-Eng.* **2023**, *57*, 455–460.
43. Pour, A.B.; Hashim, M. Identification of hydrothermal alteration minerals for exploring of porphyry copper deposit using ASTER data, SE Iran. *J. Asian Earth Sci.* **2011**, *42*, 1309–1323. [CrossRef]
44. Bahrami, Y.; Hassani, H.; Maghsoudi, A. Employing both full and partial sub-pixel mapping methods to delineate hydrothermal alteration zones associated with porphyry copper deposits. *Egypt. J. Remote Sens. Space Sci.* **2025**, *28*, 303–321. [CrossRef]
45. Sabbaghi, H.; Tabatabaei, S.H. Data-driven logistic function for weighting of geophysical evidence layers in mineral prospectivity mapping. *J. Appl. Geophys.* **2023**, *212*, 104986. [CrossRef]

46. Almasi, A.; Yousefi, M.; Carranza, E.J.M. Prospectivity analysis of orogenic gold deposits in Saqez-Sardasht goldfield, Zagros orogen, Iran. *Ore Geol. Rev.* **2017**, *91*, 1066–1080. [CrossRef]
47. Yousefi, M.; Nykänen, V. Data-driven logistic-based weighting of geochemical and geological evidence layers in mineral prospectivity mapping. *J. Geochem. Explor.* **2016**, *164*, 94–106. [CrossRef]
48. Fan, C.; Xiao, F.; Zhao, Y.; Wang, J. Analytical investigation of autoencoder-based methods for unsupervised anomaly detection in building energy data. *Appl. Energy* **2018**, *211*, 1123–1135. [CrossRef]
49. Cheng, Z.; Wang, S.; Zhang, P.; Wang, S.; Liu, X.; Zhu, E. Improved autoencoder for unsupervised anomaly detection. *Int. J. Intell. Syst.* **2021**, *36*, 7103–7125. [CrossRef]
50. Doersch, C. Tutorial on variational autoencoders. *arXiv* **2016**, arXiv:1606.05908.
51. Yousefi, M.; Nykänen, V.; Harris, J.; Hronsky, J.M.; Kreuzer, O.P.; Bertrand, G.; Lindsay, M. Overcoming survival bias in targeting mineral deposits of the future: Towards null and negative tests of the exploration search space, accounting for lack of visibility. *Ore Geol. Rev.* **2024**, *172*, 106214. [CrossRef]
52. Cheng, Q. Mapping singularities with stream sediment geochemical data for prediction of undiscovered mineral deposits in Gejiu, Yunnan Province, China. *Ore Geol. Rev.* **2007**, *32*, 314–324. [CrossRef]
53. Nykänen, V.; Yousefi, M.; Sadeghi, M. Exploration information system: Mineral systems anatomy linked to computational techniques for mineral exploration targeting. *Ore Geol. Rev.* **2025**, *187*, 106961. [CrossRef]
54. Zuo, R.; Xu, Y. Graph deep learning model for mapping mineral prospectivity. *Math. Geosci.* **2022**, *55*, 1–21. [CrossRef]
55. Esmaeiloghli, S.; Lima, A.; Sadeghi, B. Lithium exploration targeting through robust variable selection and deep anomaly detection: An integrated application of sparse principal component analysis and stacked autoencoders. *Geochemistry* **2024**, *84*, 126111. [CrossRef]
56. Safari, M.; Maghsoudi, A.; Pour, A.B. Application of Landsat-8 and ASTER satellite remote sensing data for porphyry copper exploration: A case study from Shahr-e-Babak, Kerman, south of Iran. *Geocarto Int.* **2017**, *33*, 1186–1201. [CrossRef]
57. Mahboobi, M.; Katibeh, H.; Bahrami, Y.; Pour, A.B. Metaheuristic-optimized machine learning framework for remote sensing-based alteration mapping of porphyry copper systems. *Sci. Rep.* **2026**. [CrossRef] [PubMed]