

Ensemble machine learning strategies for mineral prospectivity mapping under data scarcity

Received: 4 November 2025

Accepted: 10 February 2026

Published online: 15 February 2026

Cite this article as: Amirajlo P., Hassani H., Pour A.B. *et al.* Ensemble machine learning strategies for mineral prospectivity mapping under data scarcity. *Sci Rep* (2026). <https://doi.org/10.1038/s41598-026-40125-1>

Poorya Amirajlo, Hossein Hassani, Amin Beiranvand Pour & Narges Habibkhah

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

Ensemble Machine Learning Strategies for Mineral Prospectivity Mapping Under Data Scarcity

Poorya Amirajlo¹, Hossein Hassani^{1*}, Amin Beiranvand Pour¹ and Narges Habibkhah¹

¹ Department of Mining Engineering, Amirkabir University of Technology (Tehran Polytechnic), Tehran 1591634311, Iran

Pooryaamiri298@aut.ac.ir; hhassani@aut.ac.ir;
Beiranvand.amin80@gmail.com; habibkhah@aut.ac.ir

Corresponding author: hhassani@aut.ac.ir

Abstract

In mineral prospectivity mapping (MPM), the scarcity of labeled data and severe class imbalance often undermine the stability and reliability of machine learning models. This study advances a reliability-centered framework that prioritizes calibration and reproducibility over marginal accuracy gains when training data are limited. Two ensemble configurations, Light Gradient Boosting Machine combined with AdaBoost, and Support Vector Machine combined with Gaussian Naive Bayes, were systematically evaluated using three hyperparameter optimization strategies: Grid Search, Random Search, and Bayesian Optimization. The Synthetic Minority Oversampling Technique (SMOTE) and five-fold cross-validation were applied to counteract data imbalance and improve model robustness. Results from the Dehaq Pb-Zn district, within the Sanandaj-Sirjan Zone of Iran, reveal that while Bayesian Optimization can yield slightly higher Receiver Operating Characteristic

(ROC)-Area Under the Curve (AUC) scores, Grid Search consistently delivers more stable, well-calibrated probabilities under extreme data scarcity. The SVM+GNB ensemble tuned via Grid Search demonstrated superior balance between discrimination and reliability, achieving an AUC of 0.90 with the most consistent calibration curve, whereas the Bayesian Optimization configuration marginally reached the highest AUC (0.95). These findings highlight that stability and probability calibration are decisive for trustworthy mineral prospectivity modeling in data-scarce environments. The proposed framework provides a reproducible and interpretable pathway for translating machine learning predictions into reliable exploration decisions, supporting risk-aware targeting in early-stage mineral exploration.

Keywords: Mineral Prospectivity Mapping; Ensemble Learning; Data Scarcity; Model Calibration; Grid Search; SMOTE; Bayesian Optimization

1. Introduction

Mineral prospectivity mapping (MPM) is a cornerstone of modern mineral exploration, enabling geoscientists to delineate areas with a high probability of hosting mineral deposits. By integrating geological, geochemical, and geophysical datasets, MPM supports data-informed decision-making in both well-studied and frontier terrains, ultimately reducing exploration cost and uncertainty¹⁻³. Over the past decade, artificial intelligence (AI) and data-driven analytics have transformed

mineral prospectivity mapping, replacing heuristic and subjective targeting with systematic, reproducible modeling frameworks. Machine learning (ML) and deep learning (DL) algorithms now play central roles in uncovering nonlinear relationships between mineral occurrences and diverse exploration variables. Studies such as ⁴ demonstrated the robustness of random forests under incomplete datasets, while ⁵ developed a hybrid anomaly detection model combining isolation forests and elliptic envelope techniques to identify multivariate geochemical anomalies. Recent advances have also expanded the interpretability of ML outputs, such as the Bayesian decomposition framework proposed by ⁶, and improved prediction efficiency using extreme learning machines ⁷. Comprehensive reviews ^{8,9} have underscored that machine-learning (ML) frameworks, particularly deep-learning (DL) architectures, consistently outperform classical statistical methods, especially in high-dimensional geological environments.

Among commonly applied ML approaches in MPM are logistic regression ¹⁰, support vector machines ^{11,12}, random forests ^{13,14}, and artificial neural networks ¹⁵. More advanced architectures, such as convolutional neural networks ^{16,17}, autoencoders ¹⁸, and graph-based neural networks ¹⁹, have demonstrated impressive capability in detecting subtle spatial and geochemical patterns linked to mineralization. Despite these advances, one fundamental challenge persists: most ML algorithms rely heavily on large quantities of labeled data, which are rarely available in early-stage exploration, where confirmed mineral occurrences are

sparse. This scarcity leads to model instability, overfitting, and poor generalizability ^{20,21}.

Semi-supervised learning (SSL) has emerged as a promising bridge between supervised and unsupervised approaches by exploiting both labeled and unlabeled datasets to enhance learning efficiency. Techniques such as transductive support vector machines ²², graph-based learning ^{23,24}, and generative models ^{25,26} have achieved notable success in geoscientific prediction tasks ^{27,28}. More recently, hybrid SSL models, such as SemiGAN ²⁹, have demonstrated strong potential in capturing complex spatial distributions of mineralization by combining adversarial learning with probabilistic inference. Nevertheless, comprehensive comparisons of model reliability, calibration, and reproducibility under conditions of data scarcity remain limited, particularly within MPM applications. Table 1 summarizes key developments in machine-learning (ML) models applied to MPM, including deep-learning (DL) architectures and semi-supervised learning (SSL) approaches. While deep learning dominates in data-rich scenarios, semi-supervised and ensemble-based methods show greater resilience in sparse data environments by leveraging complementary strengths across algorithms. This shift reflects a broader transition in exploration geoscience: from a focus on maximizing accuracy to ensuring reliability and stability of predictions under uncertainty. Yet, few studies have systematically assessed the reliability and stability of ML-based MPM frameworks under extreme data scarcity, where calibration and reproducibility may matter more than marginal accuracy gains.

Building on this gap, the present study shifts focus from accuracy to reliability in mineral prospectivity modeling. We develop and evaluate two ensemble strategies: (i) Support Vector Machine combined with Gaussian Naive Bayes, and (ii) LightGBM combined with AdaBoost using three optimization approaches: Grid Search, Random Search, and Bayesian Optimization. Each configuration is assessed across three complementary criteria: predictive accuracy (F1-score), discrimination power (ROC-AUC), and probability calibration. Using Pb-Zn mineralization data from the Dehaq area within the Sanandaj-Sirjan Zone, Iran, we demonstrate that grid search produces the most stable and reproducible results under data-limited conditions. By systematically examining the trade-offs between model complexity, tuning strategy, and calibration behavior, this study introduces a reproducible ensemble framework tailored to data-scarce geological environments. The findings provide methodological guidance for developing reliable AI-driven MPM workflows that support risk-aware and interpretable exploration decisions.

Table 1. Comparative table of MPM studies using machine-learning (ML) models, including deep-learning (DL) and semi-supervised learning (SSL) approaches.

Ref.	Authors (Year)	Model Type	Learning	Data Used	Key Contribution	Study Area
4	Carranza & Laborte (2015)	Random Forest	Supervised	Sparse, incomplete geological data	Robust with small and missing data	Abra, Philip-pines
5	Chen et al. (2021)	Isolation Forest + Elliptic Envelope	Unsupervised	Multivariate geochemical anomalies	Accurate detection of anomalies	China
9	Chen et al. (2023a)	Review (Theoretical)	-	-	Opportunities in AI-driven geosciences	-
30	Chen et al. (2023b)	Deep Learning (General)	Supervised	Crustal tectonic features	Insight into early Earth tectonics	Global
7	Chen & Wu (2017)	Extreme Learning Machine (ELM)	Supervised	Geochemical data	High-speed, efficient regression	China
6	Mao et al. (2023)	Bayesian Decomposition	Supervised	MPM data	Nonlinear, interpretable model	Not specified
8	Sun et al. (2020)	Multiple ML / DL Models	Supervised	Geological, geophysical, geochemical	Broad ML vs classic model comparison	Not specified

31	Tao et al. (2021)	Fuzzy Logic + WofE	Hybrid	3D mineral data	Probabilistic 3D-prospectivity modeling	China
10	Xiong & Zuo (2018)	Logistic Regression (Rare Event)	Supervised	GIS + deposit points	Probabilistic prediction with rare data	China
11	Abedi et al. (2012)	SVM (Multi-class)	Supervised	Structural geological data +	Multi-target classification	Iran
12	Zuo & Carranza (2011)	Support Vector Machine	Supervised	Multi-layered geological data	Effective nonlinear classification	China
32	Chen & Xiao (2023)	Projection Pursuit RF	Supervised	Geology + remote sensing	Feature space optimization + interpretability	China
13	Parsa & Maghsoudi (2021)	Random Forest	Supervised	Mineral systems criteria	Sensitivity to expert-driven features	Iran
14	Talebi et al. (2022)	Spatial RF	Supervised	Spatially autocor related geology	RF adapted for spatial correlation	Iran
33	Chen et al. (2022)	Wavelet Neural Network + Monte Carlo	Supervised	Geophysical + geochemical	Probabilistic modeling with uncertainty	China
15	Maepa et al. (2021)	ANN + SVM	Supervised	Gold prospectivity data	High-performance hybrid prediction	South Africa

34	Bigdeli et al. (2022)	SOM + K-Means	Unsupervised	Geochemical anomalies	Pattern detection in unsupervised clustering	Iran
35	Zuo et al. (2019)	Deep Learning	Supervised	Geochemical survey data	First review of DL for geochemistry	Global
36	Zuo et al. (2021)	DL Framework Review	-	Big geological datasets	Challenges in DL-based prospecting	Global
37	McMillan et al. (2021)	VNet CNN	Supervised	Remote sensing	DL-based 3D mineral mapping	Australia
38	Yang et al. (2022a)	CNN + Data Augmentation	Supervised	Mineral + remote sensing	Performance gain with augmented data	China
17	Xu et al. (2021)	Deep Neural Network	Supervised	Remote sensing + geology	CNN applied to regional mapping	China
16	Li et al. (2023)	Attention-based CNN	Supervised	Multivariate MPM data	Enhancing interpretability in CNNs	China
39	Yin et al. (2023)	Self-Attention DL	Supervised	Multisource prospecting data	Novel transformer-based architecture	China
18	Yang et al. (2022b)	Convolutional Autoencoder + RF	Supervised	Multisource prospecting data	Feature extraction + ensemble learning	China
19	Zuo & Xu (2023)	Graph Neural Network	Supervised	Spatial and relational features	Capturing geological topology	China
27	Fatehi & Asadi (2017)	Semi-supervised SVM	Semi-supervised	Drilling data	Optimized drill hole targeting	Africa

28	Wang et al. (2020)	Semi-supervised RF	Semi-supervised	Sparse labeled + unlabeled data	Combines RF with SSL for better generalization	China
29	Li et al. (2025)	Semi-supervised Ensemble	Semi-supervised	GIS + geochemical features	Novel model with high accuracy	Not specified

ARTICLE IN PRESS

2. Geology of the study area

The Malayer-Esfahan belt, situated in central Iran, is one of the most significant tectono-metallogenic zones in the region, particularly renowned for its abundant lead and zinc deposits. It lies within the Sanandaj-Sirjan metamorphic-sedimentary zone, a structurally complex domain influenced by the proximity of the Urumieh-Dokhtar magmatic arc to the north and the actively deforming Zagros orogenic belt to the south. This tectonic configuration, shaped by successive Alpine orogenic events, has generated intense deformation, metamorphism, and magmatism, collectively fostering favorable conditions for hydrothermal fluid circulation and base-metal mineralization. Within this framework, the 1:100,000 Kuhe-Dehaq map sheet, part of the Golpayegan quadrangle, provides an excellent representation of these geological processes. Positioned along the northern margin of the Sanandaj-Sirjan Zone, this area exhibits geological similarities to Central Iran and possesses considerable potential for further exploration⁴⁰. The alteration of fold orientations from east-west to northwest-southeast reflects the overprinting of multiple tectonic regimes, suggesting the existence of concealed fault systems that may have acted as fluid conduits. The northern subzone (Zarakan-Lay Bid), dominated by Upper Paleozoic, Triassic-Jurassic, and Cretaceous rock units, is divided by the Dehaq-Saleh Kuh main fault—an extensive structural feature associated with secondary faults and fractures that likely facilitated hydrothermal fluid migration.

Metamorphic, magmatic, and volcanic activities during the Triassic-Jurassic and Cretaceous periods established ideal conditions for Pb-Zn mineralization. Lead and zinc occur in layered, vein-type, and lenticular forms, predominantly as strata-bound deposits hosted in limestone, dolomitic limestone, shale, and Cretaceous sandstone. The K1sh.m and K1dl formations constitute the principal host rocks. Mineralization in the Dehaq area is mainly of the Mississippi Valley-Type (MVT), commonly associated with dolomitization and stratabound ore bodies ⁴¹. Widespread dolomitization, silicification, and the development of iron oxides (hematite, limonite, and goethite) indicate hydrothermal alteration processes linked to fluid-rock interaction. Fractures, faults, and karstic zones provide critical structural controls that guided fluid flow and ore deposition, making them important exploration indicators. The accumulation of mineralization within both dolomitized carbonate units and the underlying clastic sequences points to a multi-phase mineralizing system. These features collectively highlight the need for integrated geological, geochemical, and remote sensing studies, supported by three-dimensional modeling, to better understand the spatial distribution and genesis of Pb-Zn deposits in the region. The geographic location and geological framework of the Dehaq study area are shown in Figure 1.

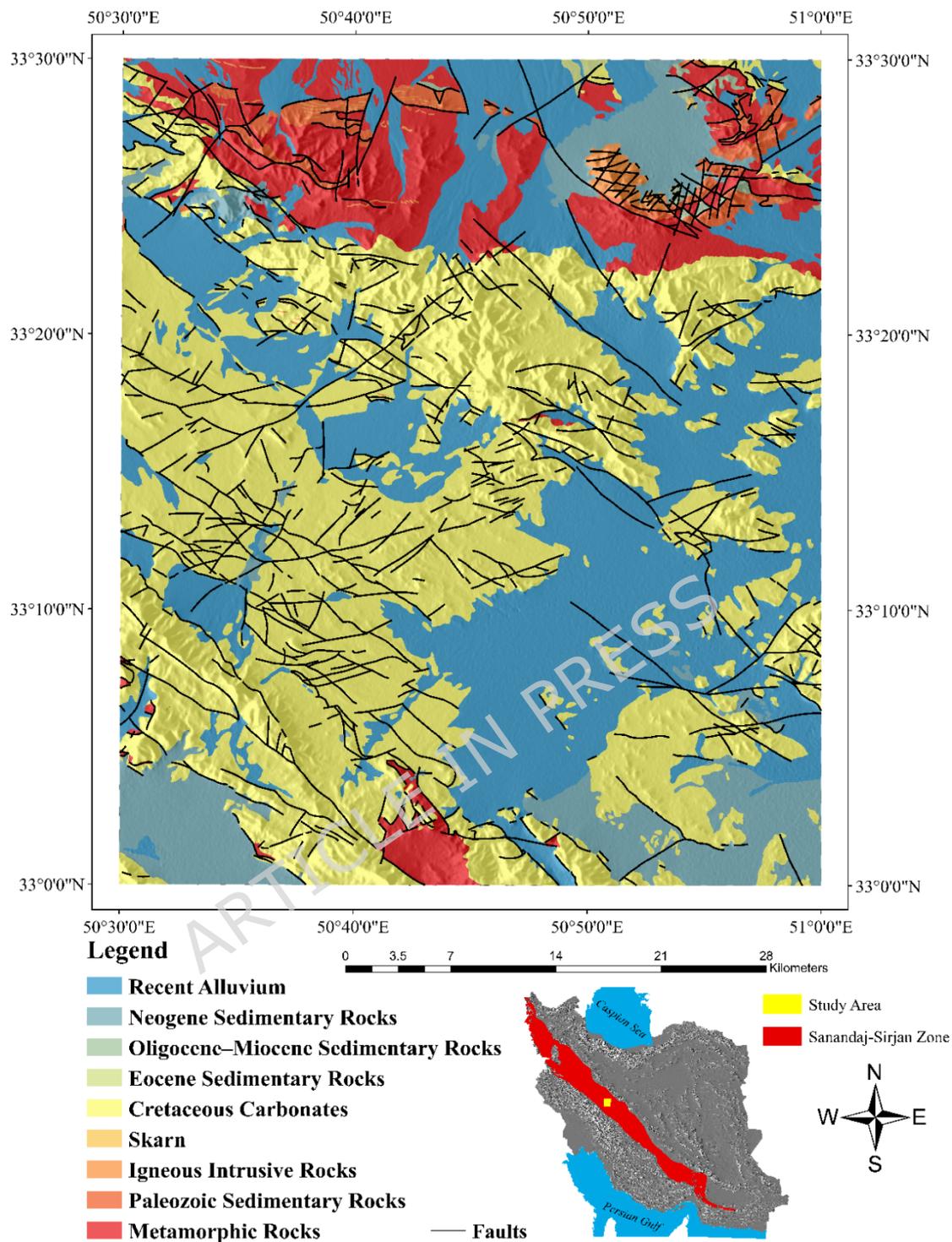


Figure 1. Geographical location and geology map of the Dehaq region. (Maps were generated using ArcGIS Desktop version 10.8 (<https://www.esri.com/arcgis>))

4. Materials and methods

4.1 Geological data

The geological evidence utilized in this study consists of lithological map units (polygons) and structural features, including faults. These characteristics offer essential insights into the geological processes that drive mineralization and are employed to establish criteria that can be mapped for mineral prospectivity analysis. The geological information for this research was sourced from the 1:100,000 bedrock geology map published by the Geological Survey & Mineral Explorations of Iran (GSI). Analyzing the geological units and faults depicted on these maps allows exploration geologists to enhance their understanding and interpretation of geophysical and geochemical data related to Pb-Zn mineralization.

Based on the conceptual framework for Pb-Zn Mississippi Valley Type (MVT) deposits and earlier studies on mineralization, the carbonates and dolomitic limestones from the Lower Cretaceous period serve as the primary host rocks for Pb-Zn mineral deposits within the Malayer-Esfahan belt ⁴². To establish a dependable geological basis for predicting MVT Pb-Zn mineralization, the geological map of the Dehaq area at a scale of 1:100,000 was digitized. Subsequently, a proximity map highlighting the Lower Cretaceous limestone formations was created, and its values were transformed into a fuzzy [0-1] scale using the logistic function (Figure 2a).

In tasks related to mineral exploration and MPM studies, examining faults and fractures is vital, as these geological structural elements are

essential for pinpointing the best locations for exploring various types of mineral deposits ^{43,44}. Faults serve as important conduits for the movement and transmission of ore-bearing fluids ⁴⁵. Therefore, creating evidence layers associated with faults, such as fault density, is a critical structural factor in modeling the prospectivity of MVT Pb-Zn ore deposits. Regions with elevated fault density values are more inclined to contain mineralization, rendering them desirable targets for forecasting the locations of ore deposits. The fault density layer was produced (Figure 2b), and its values were normalized to a [0-1] range using the logistic function ⁴⁶.

4.2 Geochemical data

A 1:100,000 Geochemical Data: A total of 624 samples were collected, testing for 44 chemical elements such as Pb, Zn, Cu, Au, Mo, Mn, Cd, Co, Cr, and Hg. The geochemical datasets underwent preprocessing to eliminate extreme outliers, ensuring data accuracy. The primary objective of regional-scale geochemical exploration is to recognize promising areas related to mineralization through analyzing and processing stream sediment data. Based on the characteristics of MVT mineralization and previous MPM research in the Dehaq area ⁴⁷, it has been demonstrated that Pb, Zn, Cu, As, Ti and Cd elements show a strong correlation with mineralization. Figure 2c illustrates the geochemical anomalies of lead and zinc in the Dehaq area, derived using the RPCA+IF approach.

4.3 Remote sensing data

ASTER Remote Sensing Images were selected based on clear imagery, prioritizing those with minimal cloud cover and sparse vegetation. The study area is situated in the arid Sanandaj-Sirjan zone of Isfahan, Iran, which is characterized by a temperate mountainous climate featuring cold, dry winters and hot, dry summers. Precipitation primarily occurs during the spring and autumn seasons. Data from ASTER for the year 2007 was chosen, specifically with a processing level of L1-T, cloud cover of less than 3%, and relatively low vegetation density, free of snow. The L1-T data underwent atmospheric, radiometric, and geometric corrections during preprocessing, making them immediately usable after clipping. In carbonate-hosted Pb-Zn deposits, zones of silicification and Fe-oxide alteration often exhibit a spatial correlation with mineralization. Consequently, utilizing remote sensing data can significantly narrow the search area and improve the accuracy of identifying high-potential zones for subsequent mineral exploration programs. This study employed a mathematical-based band ratio technique for satellite image processing. Based on the characteristics of the targeted alterations, the band ratios of 14/12 and 2/1 were utilized to emphasize silicic and iron oxide alterations, respectively. Proximity maps to these alterations were then generated, and their values were converted to a fuzzy [0-1] scale using the logistic function (Figure 2d).

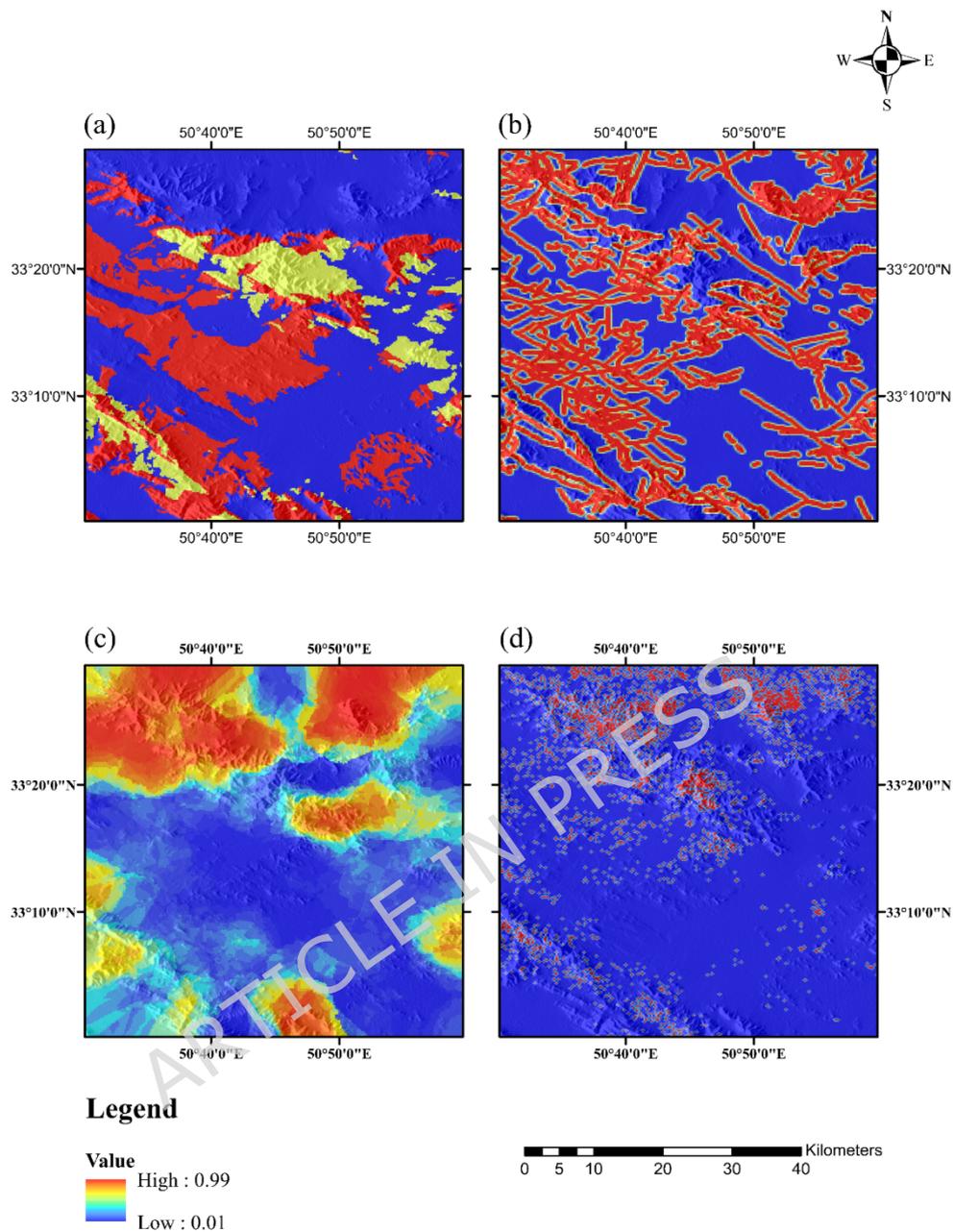


Figure 2. Fuzzy value of the (a) proximity to host rock (b) fault density (c) geochemical anomaly (d) proximity to alteration. (Maps were generated using ArcGIS Desktop version 10.8 (<https://www.esri.com/arcgis>))

4.4 Methodology

The methodological framework adopted in this study (Figure 3) provides a systematic, data-driven workflow for mineral prospectivity mapping of Pb-Zn mineralization in the Dehaq area, situated within the Sanandaj-

Sirjan Zone of Isfahan Province, Iran. The workflow integrates geochemical, geological, and remote sensing datasets to capture multiple mineralization indicators within a unified spatial analysis environment. In the first stage, geochemical datasets were processed to delineate Pb-Zn geochemical anomalies, while geological data were analyzed to derive fault density, lithological composition, and host rock parameters. Concurrently, remote sensing data were employed to map alteration zones associated with dolomitization and silicification, which are diagnostic of Mississippi Valley-Type (MVT) mineralization. These derived features were standardized, georeferenced, and compiled into a comprehensive spatial database. The resulting dataset was partitioned into training, validation, and testing subsets to facilitate model calibration and evaluation. In total, the dataset comprised 108 labeled samples, including 27 Pb-Zn deposit points (positive class) and 81 non-deposit points (negative class). Using stratified random sampling, the data were split into training (70%), validation (15%), and testing (15%) subsets while preserving the original class imbalance. This yielded 76 samples (19 positive and 57 negative) for training, 16 samples (4 positive and 12 negative) for validation, and 16 samples (4 positive and 12 negative) for testing.

During model calibration, five-fold cross-validation was applied exclusively to the training subset. For each ensemble configuration and each hyperparameter optimization strategy (Grid Search, Random Search, and Bayesian Optimization), the 76 training samples were

randomly partitioned into five stratified folds. In each iteration, the model was fitted on four folds and evaluated on the remaining fold, and the average performance across the five folds was used as the cross-validated score for that hyperparameter configuration. The validation subset (16 samples) was used only for post-hoc assessment and threshold analysis, and the independent test subset (16 samples) was never included in the cross-validation loop and was reserved solely for final performance reporting.

Two hybrid ensemble machine learning models were constructed to predict mineralization potential: (i) Light Gradient Boosting Machine (LightGBM) integrated with AdaBoost, and (ii) Gaussian Naive Bayes (GNB) combined with a Support Vector Machine (SVM) classifier. Each ensemble configuration was subjected to three distinct hyperparameter optimization strategies, Grid Search, Random Search, and Bayesian Optimization, to systematically explore model parameter space and enhance performance.

Following model training, outputs were validated using standard classification metrics and calibration analysis to assess discrimination ability and probability reliability. Comparative evaluation of model predictions enabled identification of the most stable and reproducible ensemble approach for mapping prospective Pb-Zn zones under data-scarce conditions. This methodological design emphasizes not only predictive accuracy but also interpretability and reproducibility, key

principles for developing reliable artificial intelligence workflows in mineral exploration.

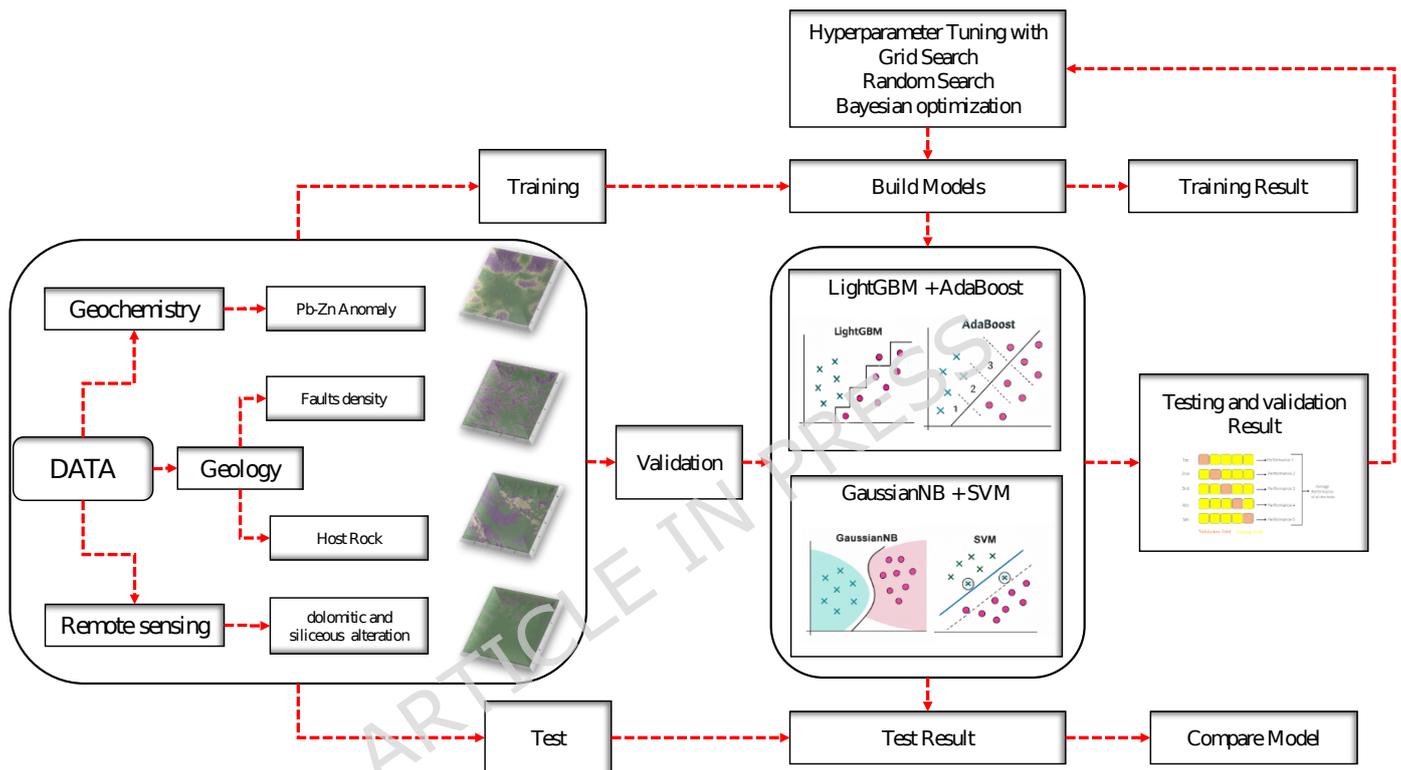


Figure 3. Diagram displaying the process recommended and applied in this research.

4.5 Algorithms implemented

4.5.1 Light Gradient Boosting Machine (LightGBM)

LightGBM is a highly efficient and scalable implementation of Gradient Boosting Decision Trees (GBDT), developed by Microsoft. It is specifically designed for rapid training, minimal memory consumption, and high

accuracy, making it particularly well-suited for large-scale and high-dimensional datasets. In contrast to traditional boosting algorithms that utilize level-wise tree growth, LightGBM employs an innovative leaf-wise strategy along with a histogram-based learning mechanism to enhance performance.

Given a training dataset $D = \{(x_i, y_i)\}_{i=1}^N$ with features $x_i \in \mathbb{R}^d$ and binary labels $y_i \in \{0, 1\}$, the goal of LightGBM is to learn an additive model of the form:

$$\hat{y}_i = \sum_{t=1}^T f_t(x_i), \quad f_t \in F$$

(1)

where each f_t represents a regression tree, and F denotes the functional space of all possible regression trees. The learning process aims to minimize a regularized objective function:

$$L^{(t)} = \sum_{i=1}^N l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t)$$

(2)

where l is a differentiable loss function (e.g., binary cross-entropy), and $\Omega(f)$ is a regularization term that penalizes the complexity of the tree. A typical regularization form is:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2$$

(3)

where T is the number of leaves in the tree and w_j is the weight of the j -th leaf. LightGBM uses a second-order approximation of the loss function to fit each tree. Let g_i and h_i denote the first and second derivatives of the loss function with respect to the prediction \hat{Y}_i :

$$g_i = \frac{\partial l(y_i, \hat{y}_i)}{\partial \hat{y}_i}, \quad h_i = \frac{\partial^2 l(y_i, \hat{y}_i)}{\partial \hat{y}_i^2}$$

(4)

Then the objective becomes:

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

(5)

Unlike XGBoost, which grows trees level-wise, LightGBM grows trees leaf-wise by expanding the leaf with the maximum loss reduction. This strategy reduces more loss but may lead to deeper trees, which can be mitigated by setting a maximum depth or number of leaves.

LightGBM speeds up training by discretizing continuous feature values into fixed-size bins (typically 255). Histogram-based split finding reduces computation and memory costs, making it highly efficient for large-scale datasets.

LightGBM supports native handling of categorical features, parallel learning, and GPU acceleration. Its key advantages include faster

training, higher accuracy due to leaf-wise growth, and lower memory usage through histogram-based computation.

4.5.2 Adaptive Boosting (AdaBoost) Classifier

Adaptive Boosting (AdaBoost) is one of the most influential ensemble learning algorithms primarily designed to improve the performance of weak learners, most commonly decision stumps. The core idea of AdaBoost is to sequentially train a series of weak classifiers, where each subsequent classifier is trained with an increased focus on the previously misclassified instances. This adaptive reweighting of instances enables AdaBoost to effectively minimize classification error.

In AdaBoost, for a given training set $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, where $x_i \in \mathbb{R}^d$ and $y_i \in \{-1, +1\}$, the algorithm initializes a weight distribution $D_1(i) = \frac{1}{n}$ over all samples. At each boosting round $t = 1, 2, \dots, T$, a weak classifier $h_t(x)$ is trained to minimize the weighted classification error:

$$\varepsilon_t = \sum_{i=1}^n D_t(i) \cdot 1(h_t(x_i) \neq y_i) \quad (6)$$

The classifier's weight α_t is computed as:

$$\alpha_t = \frac{1}{2} \ln \left(\frac{1 - \varepsilon_t}{\varepsilon_t} \right) \quad (7)$$

The weights of the samples are then updated for the next iteration:

$$D_{t+1}(i) = \frac{D_t(i) \cdot e^{-\alpha_t y_i h_t(x_i)}}{Z_t}$$

(8)

where Z_t is a normalization factor ensuring that D_{t+1} forms a probability distribution. Finally, the ensemble classifier is formed as a weighted majority vote of the weak learners:

$$H(x) = \text{sign}\left(\sum_{t=1}^T \alpha_t h_t(x)\right)$$

(9)

AdaBoost is particularly effective in reducing bias and variance simultaneously and tends to perform well on both small and large datasets. However, it is sensitive to noisy data and outliers due to its emphasis on hard-to-classify examples.

4.5.3 Support Vector Machine (SVM) Classifier

Support Vector Machine (SVM) is a powerful supervised learning algorithm primarily used for classification tasks. The core idea behind SVM is to find an optimal hyperplane that maximally separates the data points of different classes. In the case of linearly separable data, SVM aims to find the hyperplane that has the maximum margin, which is the largest distance between the hyperplane and the nearest data points from each class, called support vectors.

Given a training data $\{(x_i, y_i)\}_{i=1}^n$ where $x_i \in \mathbb{R}_d$ is the i -th input vector and $y_i \in \{-1, +1\}$ is its corresponding class label, the primal optimization problem for a hard-margin SVM is formulated as:

$$\min_{w, b} \frac{1}{2} \|w\|^2$$

(10)

subject to: $y_i(w^T x_i + b) \geq 1, \forall i = 1, \dots, n$

(11)

For non-linearly separable data, the soft-margin SVM introduces slack variables ξ_i to allow some misclassifications, modifying the objective as:

$$\min_{w, b, \xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i$$

(12)

subject to: $y_i(w^T x_i + b) \geq 1 - \xi_i, \xi_i \geq 0, \forall i$

(13)

Here, $C > 0$ is the regularization parameter that controls the trade-off between maximizing the margin and minimizing classification errors. In practice, kernel functions $K(x_i, x_j)$ are used to transform the input space into higher dimensions, allowing SVMs to perform non-linear classification efficiently via the kernel trick.

Commonly used kernels include:

- Linear: $K(x, x') = x^T x'$
- Polynomial: $K(x, x') = (x^T x' + r)^d$
- Radial Basis Function (RBF): $K(x, x') = \exp(-\gamma \|x - x'\|^2)$

SVMs have demonstrated excellent performance in high-dimensional spaces and are widely used in various applications, including bioinformatics, image recognition, and text classification.

4.5.4 Gaussian Naive Bayes (GNB) Classifier

The Gaussian Naive Bayes (GNB) classifier is a probabilistic machine learning model based on Bayes' Theorem with the strong (naive) assumption that features are conditionally independent given the class label. In the Gaussian variant, it is assumed that the likelihood of the features follows a Gaussian (normal) distribution.

Bayes' Theorem forms the foundation of the model and is expressed as:

$$P(Y|X) = \frac{P(X|Y) \cdot P(Y)}{P(X)}$$

(14)

where $P(y | x)$ is the posterior probability of class Y given the input features $x = (x_1, x_2, \dots, x_n)$, $P(y)$ is the prior probability of the class, $P(x | y)$ is the likelihood of the features given the class, and $P(x)$ is the marginal likelihood.

Under the naive independence assumption:

$$P(X|Y) = \prod_{i=1}^n P(x_i|Y)$$

(15)

In the Gaussian version, each feature x_i is modeled as a normal distribution conditioned on the class label:

$$P(x_i|Y) = \frac{1}{\sqrt{2\pi\sigma_{Y,i}^2}} \exp\left(-\frac{(x_i - \mu_{Y,i})^2}{2\sigma_{Y,i}^2}\right)$$

(16)

where $\mu_{Y,i}$ and $\sigma_{Y,i}^2$ are the mean and variance of feature x_i for class Y , estimated from the training data.

The GNB classifier predicts the class \hat{Y} that maximizes the posterior:

$$\hat{Y} = \underset{Y}{\operatorname{argmax}} (P(Y) \prod_{i=1}^n P(x_i|Y))$$

(17)

Gaussian Naive Bayes is particularly effective when the assumptions of feature independence and Gaussian distribution hold approximately true. It performs well in high-dimensional spaces and is widely used in text classification, spam detection, and medical diagnosis due to its simplicity, interpretability, and computational efficiency.

4.5.5 Ensemble Classifiers: LightGBM + AdaBoost and SVM + GaussianNB

To enhance predictive performance and promote better generalization on imbalanced and small-scale datasets, two ensemble learning configurations were devised. The first framework combines the Light Gradient Boosting Machine (LightGBM) with AdaBoost, whereas the second integrates the Support Vector Machine (SVM) with Gaussian Naive Bayes (GNB). Both ensembles employ a soft voting strategy, wherein the final prediction is derived by averaging the class probability estimates produced by each base classifier. This methodology capitalizes on the complementary strengths of heterogeneous learners, thereby facilitating the effective modeling of both high-bias and high-variance patterns.

Formally, the final predicted label \hat{Y} using soft voting is determined as:

$$\hat{Y} = \operatorname{argmax}_k \left(\frac{1}{J} \sum_{j=1}^J P_k^{(j)} \right)$$

(18)

where $P_k^{(j)}$ is the probability of class k predicted by the j -th classifier, and J denotes the total number of models in the ensemble.

Ensemble 1: LightGBM + AdaBoost LightGBM is a gradient boosting framework that grows trees leaf-wise and leverages histogram-based feature binning for computational efficiency. Its objective function after

$$T \text{ boosting rounds is: } L = \sum_{i=1}^n l(Y_i, \hat{Y}_i^{(t)}) + \sum_{t=1}^T \Omega(f_t) \quad (19)$$

where $l(\cdot)$ is the loss function, and $\Omega(f_t)$ is a regularization term for controlling model complexity. AdaBoost iteratively trains weak learners, giving more weight to misclassified samples. The final prediction of the ensemble is a weighted vote:

$$H(x) = \text{sign}\left(\sum_{t=1}^T \alpha_t h_t(x)\right) \quad (20)$$

where $h_t(x)$ denotes the t -th weak learner and α_t is its associated weight derived from its accuracy.

Ensemble 2: SVM + Gaussian Naive Bayes This hybrid ensemble integrates the margin-maximizing nature of SVM with the probabilistic generative capacity of GNB. The decision function of a linear SVM is given by:

$$f(x) = \text{sign}\left(\sum_{i=1}^n \alpha_i Y_i K(x, x_i) + b\right) \quad (21)$$

where $K(x, x_i)$ is the kernel function, and α_i are the Lagrange multipliers. GNB computes the posterior using Bayes' theorem assuming feature independence:

$$P(Y = k | \mathbf{x}) \propto P(Y = k) \prod_{i=1}^d N(x_i | \mu_{ki}, \sigma_{ki}^2)$$

(22)

This ensemble combines a discriminative classifier with a generative one to achieve higher calibration and robustness.

4.6 Methodology developed in this study

Due to the limited number of available samples and the presence of class imbalance in the dataset, a multi-stage and optimized framework was designed to prevent overfitting and enhance prediction accuracy. The key stages of this framework are as follows:

(1) Feature Standardization: to eliminate the effect of variable scale on model performance especially for algorithms like SVM that are sensitive to feature magnitude standardization based on the mean and standard deviation was applied:

$$X_{\text{scaled}} = \frac{X - \mu}{\sigma}$$

(23)

where μ and σ represent the mean and standard deviation of each feature, respectively.

(2) Handling class imbalance using SMOTE: To mitigate the strong imbalance between the deposit (minority) and non-deposit (majority) classes, the Synthetic Minority Oversampling Technique (SMOTE) was

employed. In this study, the full dataset contained 27 Pb-Zn deposit points and 81 non-deposit points, and this imbalance was also reflected in the training subset. SMOTE generates synthetic samples for the minority class by interpolating between each original deposit sample and its nearest neighbours in the multidimensional predictor space, according to:

$$x_{\text{new}} = x_i + \delta \cdot (x_{\text{nn}} - x_i), \delta \sim U(0,1)$$

(24)

where x_i is a real minority sample and x_{nn} is its nearest neighbor. In our workflow, SMOTE was applied exclusively to the training subset. Before oversampling, the training set contained 19 deposit and 57 non-deposit samples; SMOTE was used to generate additional deposit samples and obtain a more balanced training distribution, while the majority class was left unchanged. The validation and test sets were not modified and retained their original class distributions, providing an unbiased evaluation under realistic conditions.

(3) Ensemble Learning: two ensemble strategies were implemented to enhance classification performance:

- Ensemble 1: Combination of LightGBM and AdaBoost using soft voting.
- Ensemble 2: Combination of GaussianNB and SVM using soft voting.

In soft voting, the final predicted probability is the average of the individual model predictions:

$$P_{\text{final}} = \frac{1}{n} \sum_{i=1}^n P_i$$

(25)

where P_i is the predicted probability from model i , and n is the number of base classifiers.

(4) Evaluation via 5-Fold Cross-Validation: To ensure fair and stable model evaluation under limited data conditions, five-fold cross-validation was applied to the training subset. The 76 training samples were split into five stratified folds; in each iteration, four folds were used for fitting the model and one fold for internal validation. This process was repeated five times, and the average performance across the five folds was used as the cross-validated score for model selection and hyperparameter tuning:

$$\text{Accuracy}_{\text{CV}} = \frac{1}{5} \sum_{i=1}^5 \text{Accuracy}_i$$

(26)

(5) Hyperparameter Tuning: to enhance model generalization and prevent overfitting hyperparameter tuning was carried out for both ensemble classifiers. Three well-known optimization strategies were employed: Grid Search, Randomized Search, and Bayesian Optimization

using Optuna. Each method explores the parameter space in a distinct way and offers a balance between computational cost and accuracy.

Hyperparameter Tuning for Ensemble (LightGBM + AdaBoost):

For the ensemble combining LightGBM and AdaBoost classifiers, the following parameters were subject to tuning:

lgbm n estimators $\in \{50, 100, 200\}$

- lgbm learning rate $\in \{0.01, 0.05, 0.1\}$
- ada n-estimators $\in \{50, 100, 200\}$
- ada learning rate $\in \{0.5, 1.0, 1.5\}$

The Bayesian optimization objective function was defined as:

$$\max_{\theta} f(\theta) = \text{CrossValScore}(\text{VotingClassifier}_{\text{LGBM+Ada}}(\theta)) \quad (27)$$

where θ denotes the hyperparameter vector, and $f(\theta)$ is the cross-validated accuracy averaged over 5 folds.

(6) Hyperparameter Tuning for Ensemble (SVM + GaussianNB): in the ensemble composed of Support Vector Machine and Gaussian Naive Bayes, only the regularization parameter of the SVM component was optimized:

- svm $C \in [10^{-3}, 10^3]$

Three search strategies were adopted:

- Grid Search: Evaluates C at fixed intervals in log-scale.
- Randomized Search: Samples 30 values randomly from the same log-scale range.
- Bayesian Optimization: Applies sequential model-based optimization using Optuna.

The Bayesian optimization function in this case was:

$$\max_{C} f(C) = \text{CrossValScore}(\text{VotingClassifier}_{\text{SVM}+\text{GNB}}(C))$$

(28)

5. Results and analysis

Figure 4 illustrates the calibration curves for the ensemble model combining the Support Vector Machine (SVM) and Gaussian Naive Bayes (GNB) classifiers, evaluated under three hyperparameter optimization strategies: Bayesian Optimization, Grid Search, and Random Search. The calibration curve represents the relationship between predicted probabilities and the actual observed frequencies of the positive class.

The diagonal dashed line indicates perfect calibration, where the model's predicted probabilities precisely match the observed outcomes. A well-calibrated model produces a curve that closely follows this diagonal, signifying that its predicted confidence levels accurately reflect true probabilities.

Among the evaluated tuning strategies, Bayesian Optimization exhibits notable fluctuations in the low-probability range (below 0.4), suggesting instability and a tendency toward over- or underconfident predictions in that interval. Although its performance improves at higher probability values, this inconsistency reduces overall calibration reliability. In contrast, the Grid Search approach produces a smoother and more stable calibration curve across the entire probability spectrum. The curve closely follows the ideal line from approximately 0.4 to 0.9, demonstrating consistent agreement between predicted and observed probabilities. This behavior indicates superior probability calibration and greater reliability of predictions. Random Search, while adequate, shows intermediate stability with minor deviations from ideal calibration. Overall, the Grid Search-optimized SVM+GNB ensemble delivers the

most reliable and well-calibrated probabilistic output among the tested configurations.

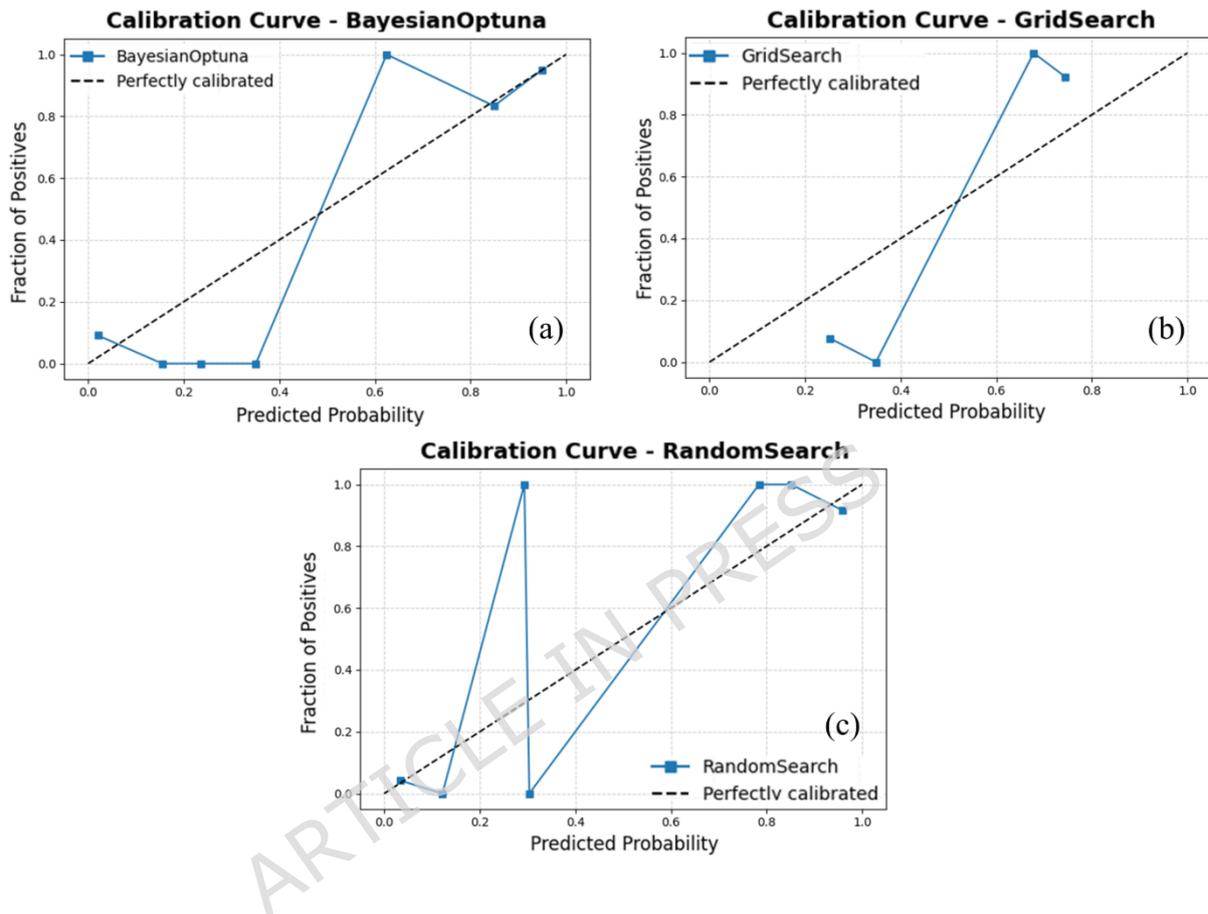


Figure 4. Calibration curves for the ensemble model combining Support Vector Machine (SVM) and Gaussian Naive Bayes (GNB), using Bayesian Optimization, Grid Search, and Random Search. (This figure was generated using Python for data visualization and analysis)

The Random Search strategy exhibits a performance pattern broadly similar to that of Bayesian Optimization, with moderate improvements in the mid- to high-probability range. However, it remains unstable at lower probability values and does not surpass Grid Search in terms of overall

calibration quality. Among the three optimization strategies, Grid Search produces the most well-calibrated ensemble model, delivering a stable and accurate correspondence between predicted probabilities and observed class frequencies. This reliability is especially critical for probability-dependent decision-making tasks, such as mineral prospectivity assessment, risk analysis, or other applications requiring trustworthy confidence estimates.

Figure 5 presents the ROC curves for the SVM+GNB ensemble classifier optimized using Bayesian Optimization, Grid Search, and Random Search. The ROC curve illustrates each model's ability to discriminate between positive and negative classes by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) across varying classification thresholds. The Area Under the ROC Curve (AUC) quantifies this performance, with values approaching 1.0 indicating excellent discriminative capability. The ensemble model tuned via Bayesian Optimization achieved the highest AUC of 0.95, reflecting superior classification performance across all thresholds. Its curve rises sharply toward the top-left corner of the plot, signifying high sensitivity with minimal false alarms. This indicates that Bayesian Optimization effectively navigated the hyperparameter space to identify a configuration with strong generalization capacity. The Random Search variant produced a slightly lower AUC of 0.94, representing similarly strong predictive power. Although its ROC curve closely parallels that of the Bayesian-optimized model, subtle deviations appear at mid-range

FPR values, suggesting marginally less consistent threshold behavior. Despite this, its overall performance remains robust and reliable for practical mineral prospectivity applications.

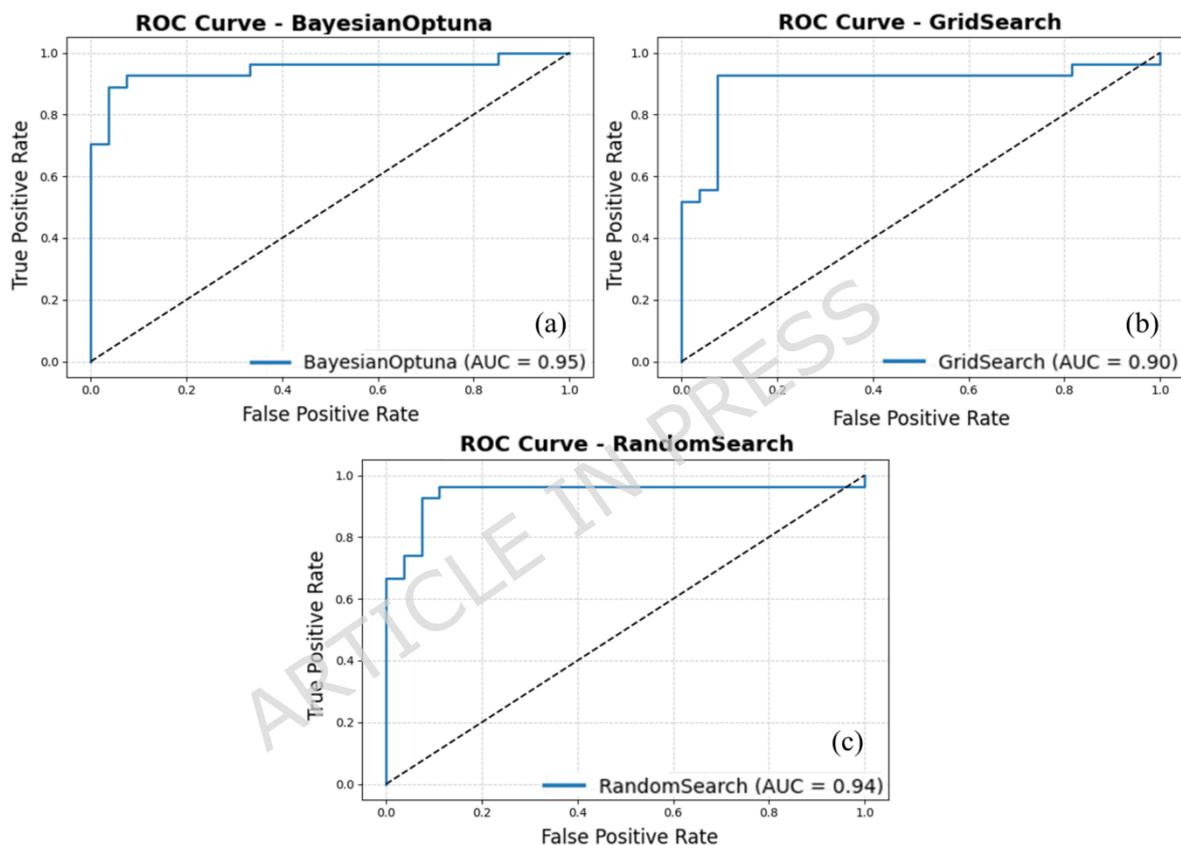


Figure 5. ROC curves for the ensemble model combining Support Vector Machine (SVM) and Gaussian Naive Bayes (GNB), using Bayesian Optimization, Grid Search, and Random Search. (This figure was generated using Python for data visualization and analysis)

The Grid Search approach, while achieving a respectable AUC of 0.90, performs slightly below the Bayesian Optimization and Random Search

methods. Its ROC curve rises more gradually and is less steep near the origin, indicating a higher false positive rate at lower classification thresholds. This limitation may result from the exhaustive yet discretized nature of Grid Search, which restricts exploration to predefined parameter combinations and may overlook optimal configurations beyond the grid boundaries. Despite these differences, all three ROC curves lie well above the diagonal reference line representing random classification, confirming that the SVM+GNB ensemble effectively captures meaningful decision boundaries under all optimization strategies. Among them, Bayesian Optimization demonstrates the best overall balance between sensitivity and specificity, yielding the highest AUC and the most desirable ROC curvature. This consistent performance underscores its suitability for high-stakes or precision-critical applications, where minimizing false positives is essential for reliable mineral prospectivity assessment.

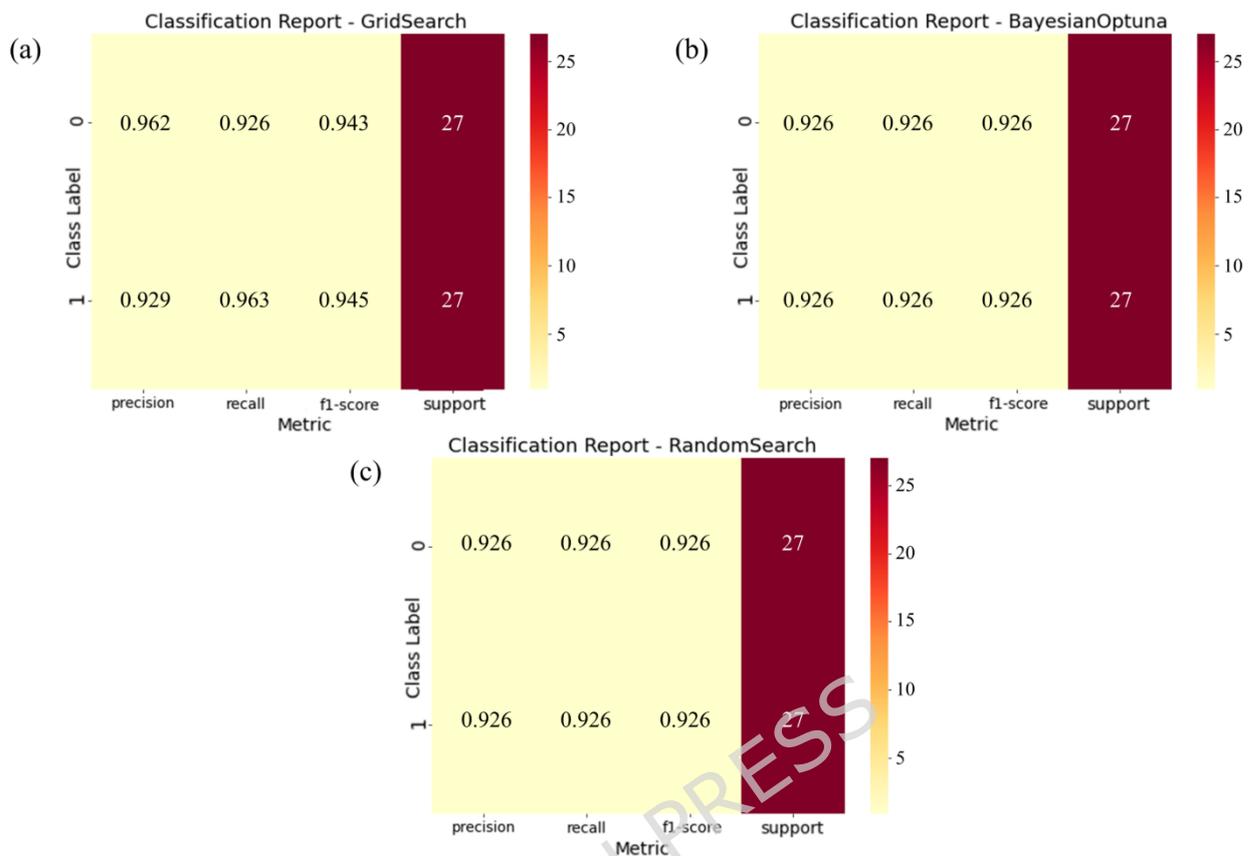


Figure 6. Classification reports of the ensemble model combining Support Vector Machine (SVM) and Gaussian Naive Bayes (GNB) using Bayesian Optimization, Grid Search, and Random Search. (This figure was generated using Python for data visualization and analysis)

Figure 6 presents the classification reports for the SVM+GNB ensemble model optimized using Bayesian Optimization, Grid Search, and Random Search. The results reveal striking consistency across all three optimization strategies. Precision, recall, and F1-score values are identical (0.926) for both classes in each configuration, demonstrating balanced and symmetric classification performance. This uniformity indicates that the ensemble model effectively distinguishes between positive and negative samples without bias, maintaining consistent predictive behavior regardless of the optimization method. While

Bayesian Optimization achieved these results with fewer evaluations due to its efficient exploration of the parameter space, the Grid Search and Random Search approaches attained equivalent metrics, suggesting a well-structured and relatively smooth search landscape for this task. The comparable outcomes across tuning methods highlight the inherent robustness of the SVM+GNB ensemble, confirming its stability under varying optimization conditions. In all three configurations, each classification report in Figure 6 is based on all 27 Pb-Zn deposit samples and 27 non-deposit samples randomly selected from the original 81 non-deposit observations, yielding a balanced evaluation subset of 54 instances (support = 27 per class).

However, identical classification metrics do not necessarily imply identical model reliability. Differences observed in the calibration curves and ROC-AUC analyses reveal that Grid Search produces the most stable and well-calibrated probability estimates, particularly in the mid-probability range where classification thresholds are typically applied. This alignment between predicted probabilities and observed outcomes indicates superior probabilistic reliability, an essential characteristic for risk-aware mineral prospectivity mapping. In contrast, the Bayesian and Random Search variants, though exhibiting slightly higher AUC scores, showed greater variability at low probability levels, suggesting a tendency toward overconfident predictions. In general, the Grid Search-optimized ensemble emerges as the most dependable configuration for mineral prospectivity modeling under data-scarce conditions. Its balance

of accuracy, calibration stability, and interpretability supports more reliable decision-making in exploration geoscience.

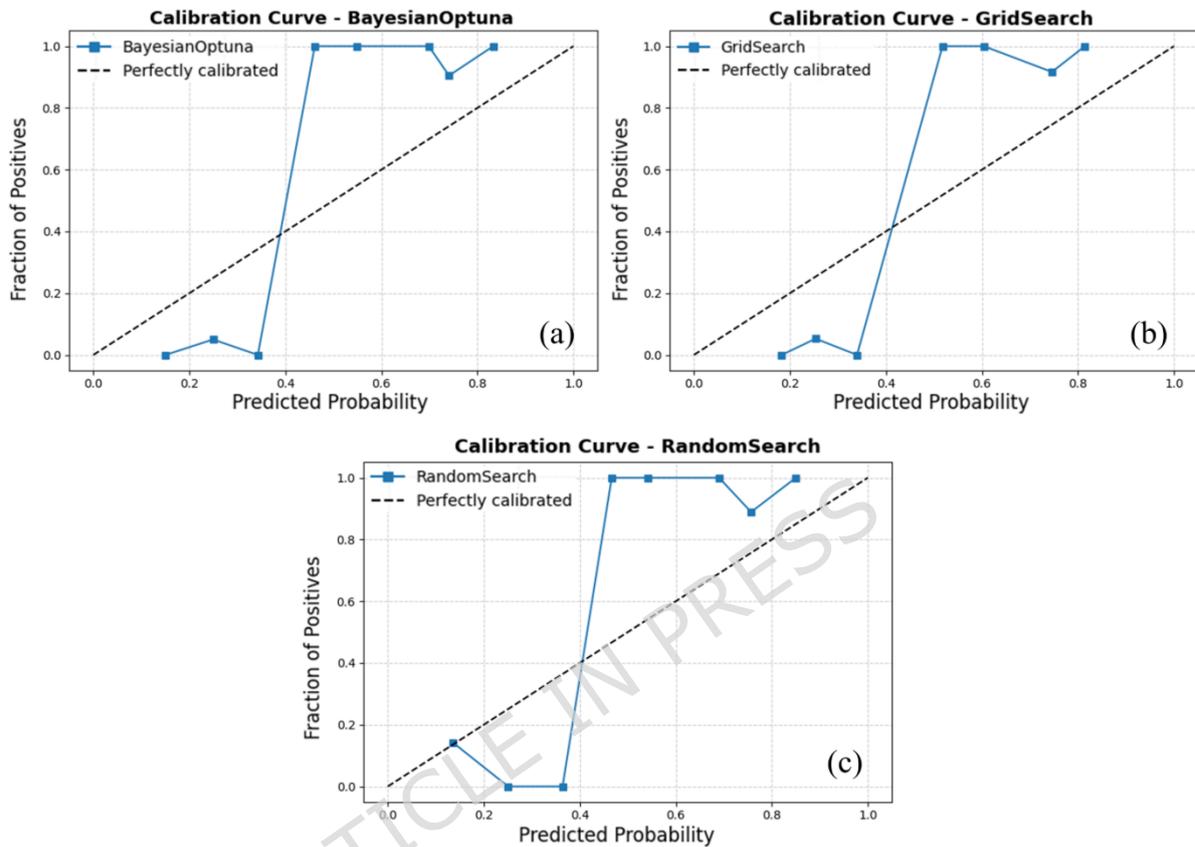


Figure 7. Calibration curves for the ensemble model combining LightGBM and AdaBoost, using Bayesian Optimization, Grid Search, and Random Search. (This figure was generated using Python for data visualization and analysis)

Figure 7 presents the calibration curves for the Light Gradient Boosting Machine (LightGBM) and AdaBoost ensemble model optimized using Bayesian Optimization, Grid Search, and Random Search. These plots evaluate the reliability of the model's predicted probabilities by comparing them to the actual observed frequencies of the positive class. All three configurations exhibit strong calibration performance, particularly in higher probability intervals (above 0.5), indicating that the

ensemble's predicted confidence levels closely reflect real outcomes. Among the tested approaches, the Grid Search-optimized model demonstrates the most consistent calibration across the full probability range. Its curve aligns smoothly with the diagonal reference line, especially in the mid- to high-probability regions, suggesting that the predicted probabilities are both stable and well-calibrated. The Bayesian Optimization variant shows minor underconfidence in the lower probability bins but quickly converges toward the ideal calibration line as probability values increase. The Random Search configuration follows a similar trend, though with slightly greater fluctuations in the low-probability range. Despite these subtle variations, all three models display satisfactory calibration behavior, with minimal deviation from perfect alignment. This consistency underscores the robustness of the LightGBM + AdaBoost ensemble across different hyperparameter optimization strategies. Nevertheless, the smoother and more stable curve of the Grid Search variant suggests a marginal advantage in terms of probabilistic reliability and interpretability, key attributes for dependable mineral prospectivity modeling under uncertain and data-limited conditions.

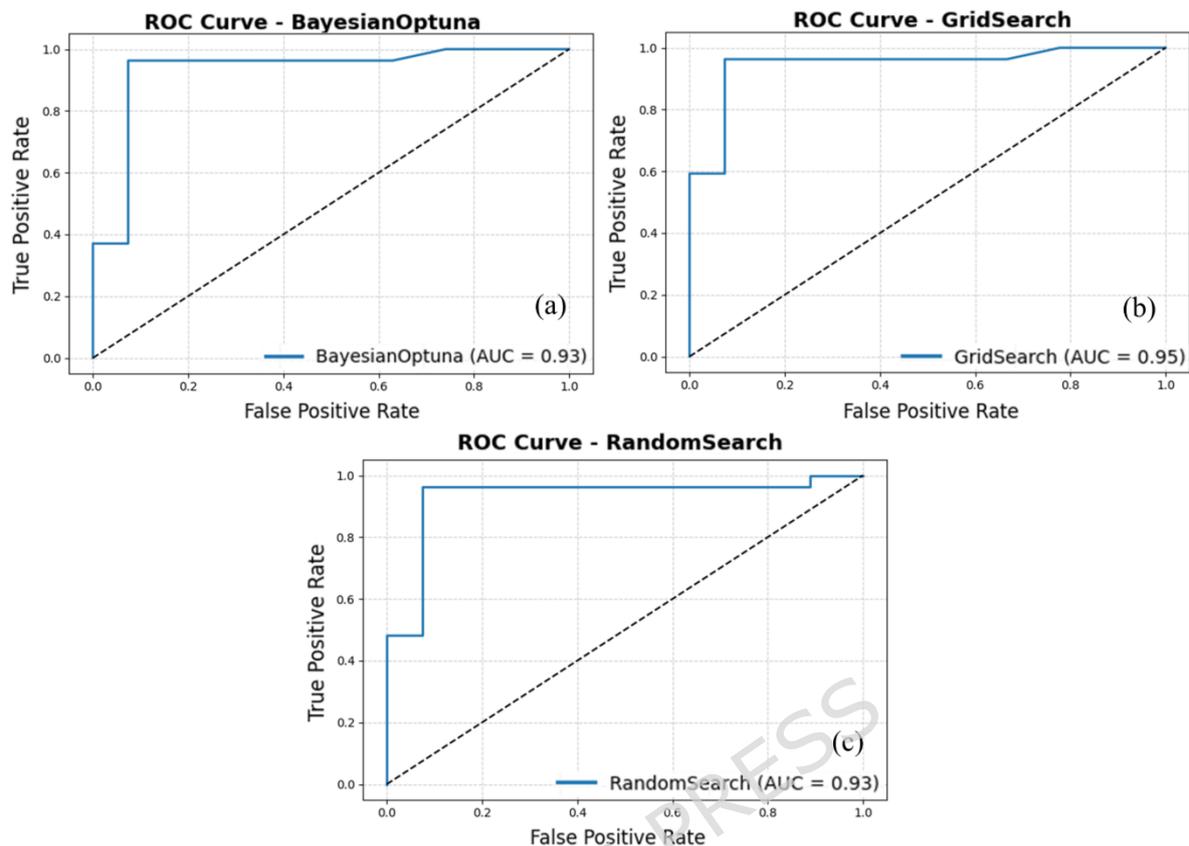


Figure 8. ROC curves for the LightGBM + AdaBoost ensemble model under three hyperparameter tuning methods: Bayesian Optimization, Grid Search, and Random Search. (This figure was generated using Python for data visualization and analysis)

Figure 8 presents the ROC curves for the Light Gradient Boosting Machine (LightGBM) and AdaBoost ensemble model optimized using Bayesian Optimization, Grid Search, and Random Search. The ROC curve evaluates the model's ability to discriminate between classes across varying classification thresholds, while the AUC quantifies this discriminative performance. Among the three optimization strategies, the Grid Search-tuned model achieved the highest AUC of 0.95, reflecting the most effective balance between true positive and false

positive rates. Both Bayesian Optimization and Random Search produced AUC values of 0.93, indicating similarly strong, though slightly lower, classification performance.

The similarity of the ROC curves suggests that all three optimization strategies yield well-performing models; however, the marginal improvement achieved through Grid Search implies that its exhaustive parameter exploration identified a slightly more optimal hyperparameter configuration. While computationally intensive, Grid Search's systematic search of the parameter space ensures thorough evaluation and reliable convergence. In contrast, Bayesian Optimization achieved competitive results with far fewer evaluations, underscoring its efficiency for scenarios where computational resources are constrained. The Random Search approach also performed strongly, demonstrating that even a randomly sampled subset of the parameter space can approximate near-optimal solutions when the underlying ensemble architecture is robust. In particular, the ROC analysis confirms the stability and high discriminative capability of the LightGBM + AdaBoost ensemble across all optimization strategies, with Grid Search showing a slight edge in ranking performance. The selection among these approaches should therefore consider practical factors such as computational budget, time efficiency, and interpretability.

Figure 9 presents the corresponding classification reports for the LightGBM + AdaBoost ensemble under the same optimization strategies,

summarizing key performance metrics, precision, recall, F1-score, and support for each class. Both Bayesian Optimization and Random Search yield perfectly balanced results, with identical values of 0.926 for all metrics, indicating a well-generalized and unbiased classifier. In contrast, the Grid Search variant shows marginally higher precision and F1-scores for Class 0 (0.962 and 0.943) and slightly improved recall and F1-scores for Class 1 (0.963 and 0.945). This minor enhancement likely results from the exhaustive parameter evaluation inherent to Grid Search, which may have captured a more optimal model configuration. Collectively, these findings highlight the robustness and consistency of the LightGBM + AdaBoost ensemble across diverse optimization techniques. The model's balanced performance, combined with reliable probability calibration and high discriminative accuracy, demonstrates its strong potential for data-scarce mineral prospectivity mapping applications. For both the SVM+GNB ensemble (Figure 6) and the LightGBM+AdaBoost ensemble (Figure 9), the classification reports were generated using all 27 Pb-Zn deposit samples together with 27 non-deposit samples randomly selected from the original 81 non-deposit observations, yielding a balanced evaluation subset of 54 instances (support = 27 per class).

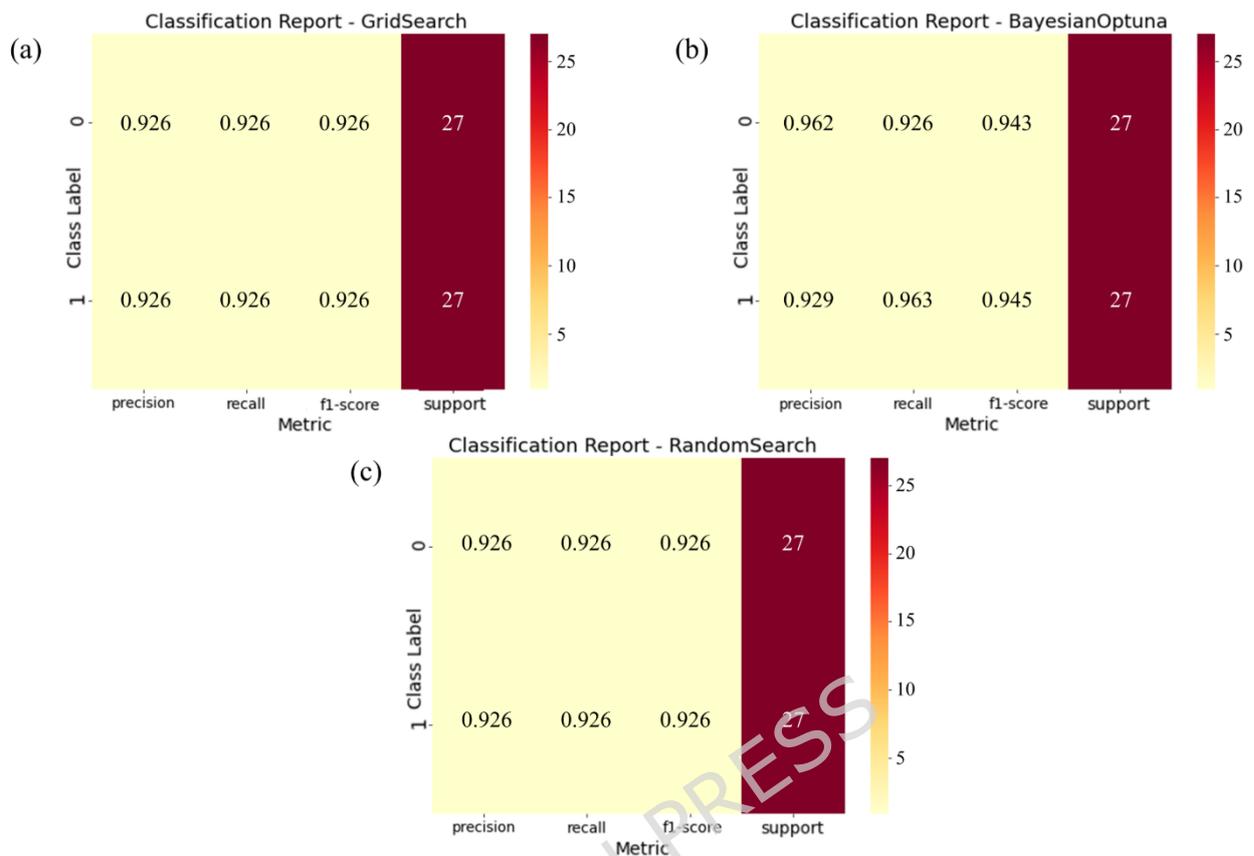


Figure 9. Classification reports of the LightGBM + AdaBoost ensemble under three hyperparameter tuning strategies: Bayesian Optimization, Grid Search, and Random Search. (This figure was generated using Python for data visualization and analysis)

A comparative assessment of the two ensemble models, SVM+GNB and LightGBM+AdaBoost, was conducted using calibration curves, ROC curves, and classification reports (Figures 4, 8, and 9). For the SVM+GNB ensemble, the Grid Search-optimized model exhibited the best calibration performance, with predicted probabilities closely matching observed frequencies across the entire probability range. Its ROC curve achieved an AUC of 0.90, slightly below Bayesian Optimization (AUC = 0.95) and Random Search (AUC = 0.94), but its superior calibration performance makes the Grid Search-optimized

SVM+GNB ensemble the most reliable configuration overall. The classification reports further revealed identical metrics for all three tuning strategies (precision, recall, and F1-score = 0.926), highlighting the model's strong stability and balanced classification behavior.

Similarly, the LightGBM+AdaBoost ensemble demonstrated consistently high discriminative performance across all optimization strategies, with Grid Search again producing the top AUC value of 0.95. This configuration also achieved marginally superior precision and F1-scores for both classes (F1 = 0.945), suggesting that the exhaustive parameter exploration inherent to Grid Search yielded a more optimal hyperparameter combination. While Bayesian Optimization provided slightly smoother calibration curves, Grid Search delivered the best overall balance between probability calibration, discrimination, and class-wise accuracy.

Typically, both ensemble frameworks proved robust across different tuning methods, underscoring their generalizability under data-scarce conditions. However, Grid Search consistently produced the most reliable and well-calibrated models, offering the most effective trade-off between predictive accuracy and probabilistic stability. These findings emphasize the importance of model reliability, rather than mere accuracy, in developing dependable machine learning workflows for mineral prospectivity mapping in limited-data environments.

Table 2. Comparison of the two ensemble models across evaluation criteria.

Evaluation Criterion	Ensemble (SVM,GNB)	Ensemble (LightGBM,AdaBoost)
Best suited for small datasets	✓ (with Grid Search)	✗
Best classification performance (F1-score)	✗	✓ (with Grid Search)
Well-calibrated probability estimates	✓	Slightly weaker
Highest ROC-AUC	✓ (AUC = 0.95, with Bayesian Optimization)	✓ (AUC = 0.95, with Grid Search)
Model simplicity and interpretability	✓	✗
Low computational cost	✓	✗ (computationally intensive)
Robustness in data-scarce MPM tasks	✓	Only with enough data and resources

Based on the comprehensive evaluation of calibration curves, ROC analyses, and classification reports, two ensemble configurations, SVM+GNB and LightGBM+AdaBoost, were systematically compared under three hyperparameter optimization strategies: Bayesian Optimization, Grid Search, and Random Search. Both ensembles demonstrated strong predictive and calibration performance; however, distinct trade-offs influence their suitability for practical mineral prospectivity mapping applications. As summarized in Table 2, the SVM+GNB ensemble optimized via Grid Search achieved the most stable and well-calibrated results, with minimal variance across evaluation

metrics The SVM+GNB ensemble optimized via Grid Search achieved the most stable and well-calibrated results, with minimal variance across evaluation metrics. It attained an AUC of 0.90, while the Bayesian Optimization configuration marginally achieved the highest AUC (0.95). This confirms high discriminative capability, but with the advantage of more stable and well-calibrated probability estimates when using Grid Search. These characteristics make it particularly suitable for data-scarce exploration environments, where model transparency and reliability are critical.

The LightGBM+AdaBoost ensemble, likewise optimized through Grid Search, achieved a comparable AUC of 0.95 and slightly higher F1-scores, reflecting its stronger learning capacity. However, its increased computational cost and reduced interpretability may limit its practicality in early-stage exploration with limited data. Accordingly, for small or sparse datasets typical of early mineral prospectivity assessments, the SVM+GNB ensemble with Grid Search is recommended as the most balanced and practical modeling strategy. As data availability and complexity increase, transitioning to more advanced learners such as LightGBM can further enhance predictive performance. Based on the two ensemble configurations, SVM+GNB and LightGBM+AdaBoost, mineral prospectivity prediction maps for Mississippi Valley-Type (MVT) Pb-Zn mineralization in the Dehaq area were generated (Figure 10a,b). Both maps delineate extensive high-potential zones that overall correspond well with the spatial distribution of known Pb-Zn occurrences,

demonstrating the ability of the two ensemble models to generalize effectively and identify geologically meaningful targets for future exploration. Visual inspection of Figure 10 shows that several deposits coincide with the highest-probability (red) anomalies in both models, whereas a few known deposits are located in adjacent high- to moderate-probability (orange/yellow) cells or along the margins of the red zones. This behaviour is consistent with the probabilistic and generalized nature of the models: the red class represents only the upper tail of the predicted probability distribution, deposit locations are represented as single points within mineralized bodies that extend over multiple grid cells, and some occurrences may be controlled by local geological features that are not fully captured by the regional-scale evidence maps used as predictors.

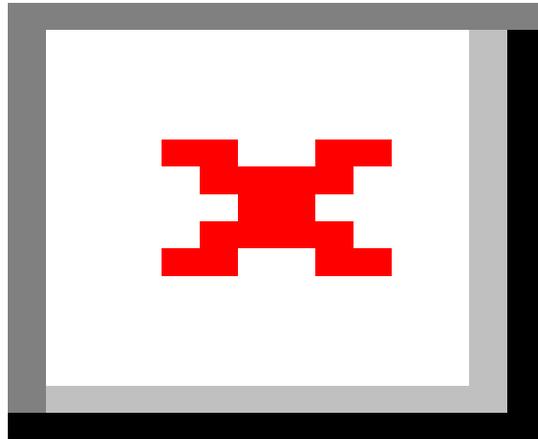


Figure 10. MPM prediction models: (a) SVM+GNB (b) light GBM+Adaboost.
(Maps were generated using ArcGIS Desktop version 10.8
(<https://www.esri.com/arcgis>))

6. Discussion

This study introduces a multi-stage framework for MPM under conditions of data scarcity and class imbalance. By integrating the Synthetic Minority Oversampling Technique (SMOTE) with two complementary ensemble configurations, LightGBM+AdaBoost and SVM+GNB, and validating them through five-fold cross-validation, the research demonstrates that Grid Search provides the most stable and reproducible optimization strategy in low-label environments. The SVM+GNB model

tuned via Grid Search yields well-calibrated probability estimates without compromising discriminative performance (AUC = 0.90), confirming the central hypothesis that calibration and methodological robustness are more critical than marginal accuracy gains when training data are limited.

Positioned within the growing literature on data-driven mineral exploration, this study extends previous work in three key respects. First, whereas most prior MPM studies emphasize ranking metrics or pursue increasingly complex tree-based learners, our results highlight that probability calibration is essential for transforming model scores into defensible exploration thresholds. Second, we demonstrate that a parsimonious hybrid ensemble (SVM+GNB) can match, and under severe data scarcity, even surpass, the practical utility of deeper and more complex architectures such as LightGBM. Third, the results provide evidence that exhaustive Grid Search can outperform Bayesian Optimization in small-sample regimes, where surrogate models tend to overfit and distort tail probabilities. Collectively, these findings reframe hyperparameter tuning as a data-regime-dependent design decision rather than a universal procedure.

6.1 Geological interpretation and exploration implications

The generated prospectivity map for the Dehaq Pb-Zn district displays strong spatial coherence with known mineral occurrences, reinforcing a geological model in which stratabound mineralization is structurally

controlled and enhanced by metasomatic alteration. The superior calibration of the SVM+GNB ensemble enables explicit management of Type I and Type II errors in decision-making, facilitating cost-sensitive threshold selection for follow-up surveys such as geochemistry, geophysics, trenching, and drilling. This approach replaces ad-hoc score cutoffs with likelihood-based decision criteria, supporting transparent and risk-aware resource allocation, an important advancement for field operations under logistical constraints. These results clarify the relationship between algorithmic complexity, data regime, and model reliability. Consistent with prior research, tree-based ensembles generally outperform margin-based classifiers as the quantity of labeled data increases. However, under extreme data scarcity, SVM-based ensembles exhibit superior resilience due to their margin-maximizing formulation and lower sensitivity to overfitting. Moreover, while Bayesian Optimization is often praised for its efficiency in large parameter spaces, our findings indicate that it may become unstable in small datasets, where iterative surrogate updates amplify noise and bias. In contrast, Grid Search, despite its higher computational cost, uniformly explores the parameter space, producing smoother likelihood landscapes and stronger run-to-run reproducibility. These characteristics align with our core argument that reliability should supersede accuracy as the guiding criterion for MPM in low-data contexts.

6.2 Strengths, limitations, and future directions

The proposed framework offers several strengths: (i) integration of heterogeneous evidence layers (geology, structure, remote sensing, and geochemistry) within a unified spatial database; (ii) controlled evaluation of hyperparameter tuning strategies with explicit attention to calibration and decision-theoretic thresholding; and (iii) validation of practical exploration utility through a prospectivity map that aligns closely with known mineralization zones. Nevertheless, limitations remain. The dataset's small size and residual class imbalance introduce wider confidence intervals even after SMOTE resampling. Additionally, the model may retain sensitivity to feature selection and engineering choices that were not exhaustively tested. Computational demands also increase substantially when applying tree-based ensembles over larger spatial extents. These factors motivate external validation across independent Pb-Zn districts and other metamorphic belts to assess model generalizability.

For early-stage exploration campaigns with limited labeled data, the SVM+GNB ensemble optimized via Grid Search provides the best balance between discrimination, calibration stability, and computational efficiency. As data volume and geological complexity increase, gradual adoption of more powerful learners such as LightGBM+AdaBoost, combined with post-hoc calibration, can further enhance prediction accuracy. Future research should integrate explicit uncertainty quantification (e.g., temperature scaling, Bayesian isotonic regression), active learning to prioritize informative sampling locations, and semi-

supervised or graph-based methods to exploit unlabeled spatial structure. Cross-province validation and multi-scale integration will be critical to establishing a generalizable, reliability-focused framework for machine learning-assisted mineral exploration.

7. Conclusions

This study proposed and validated a multi-stage ensemble learning framework for mineral prospectivity mapping (MPM) under conditions of limited and imbalanced datasets. By integrating the SMOTE with two complementary ensemble classifiers, LightGBM+AdaBoost and SVM+GaussianNB, and evaluating their performance through five-fold cross-validation, the framework achieved strong and consistent predictive results despite data scarcity. A key outcome of this research is the demonstrated importance of hyperparameter optimization strategy selection in data-constrained environments. While Bayesian Optimization is generally favored for its efficiency in large parameter spaces, our results reveal that it can exhibit instability and overfitting tendencies when applied to small datasets. In contrast, the Grid Search approach, though computationally exhaustive, provided superior stability, calibration consistency, and reproducibility, making it the preferred choice for low-data scenarios commonly encountered in early-stage mineral exploration. These findings highlight that model reliability and probability calibration are more critical than marginal accuracy gains

when datasets are sparse or imbalanced. For geoscientific applications characterized by limited labeled data, prioritizing stable optimization frameworks such as Grid Search can significantly improve the interpretability and trustworthiness of predictive models. Future research should explore hybrid optimization strategies that combine the exploratory efficiency of Bayesian methods with the stability of exhaustive search. Expanding this framework to include uncertainty quantification, semi-supervised learning, and cross-district validation will further enhance its generalizability and value for data-driven, risk-aware mineral exploration.

Author Contributions: Conceptualization, P.A.; Methodology, P.A.; Software, P. A.; Validation, P. A., H.H., N.H; Writing - Original Draft Preparation, P.A., H.H; Writing - Review & Editing, P.A., H.H., A.B.P.; Visualization, N.H.; Supervision, H.H.; Project administration, P.A. All authors have read and agreed to the published version of the manuscript.

Funding: The authors did not receive support from any organization for the submitted manuscript.

Data Availability Statement: The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Acknowledgments: We would appreciate the Department of Mining Engineering at Amirkabir University of Technology (Tehran Polytechnic). This study is part of the research activity carried out during doctoral research (By the first author) at the Department of Mining Engineering at Amirkabir University of Technology (Tehran Polytechnic).

Conflicts of Interest: The authors declare no conflict of interest

References

1. Bonham-Carter G. *Geographic Information Systems for Geoscientists: Modelling with GIS*. Elsevier; 1994.
2. Carranza EJM. *Geochemical Anomaly and Mineral Prospectivity Mapping in GIS*. Vol 11. Elsevier; 2008.
3. Parsa M, Carranza EJM. Modulating the impacts of stochastic uncertainties linked to deposit locations in data-driven predictive mapping of mineral prospectivity. *Nat Resour Res*. 2021;30(5):3081-3097.
4. Carranza EJM, Laborde AG. Random forest predictive modeling of mineral prospectivity with small number of prospects and data with missing values in Abra (Philippines). *Comput Geosci*. 2015;74:60-70.
5. Chen Y, Wang S, Zhao Q, Sun G. Detection of multivariate geochemical anomalies using the bat-optimized isolation forest and bat-optimized elliptic envelope models. *J Earth Sci*. 2021;32(2):415-426.
6. Mao X, Wang J, Deng H, et al. Bayesian decomposition modelling: An interpretable nonlinear approach for mineral prospectivity mapping. *Math Geosci*. 2023;55(7):897-942.
7. Chen Y, Wu W. Mapping mineral prospectivity using an extreme learning machine regression. *Ore Geol Rev*. 2017;80:200-213.
8. Sun T, Li H, Wu K, Chen F, Zhu Z, Hu Z. Data-driven predictive modelling of mineral prospectivity using machine learning and deep learning methods: A case study from southern Jiangxi Province, China. *Minerals*. 2020;10(2):102.
9. Chen G, Cheng Q, Puetz S. Data-driven discovery in geosciences: Opportunities and challenges. *Math Geosci*. 2023;55(3):287-293.
10. Xiong Y, Zuo R. GIS-based rare events logistic regression for mineral prospectivity mapping. *Comput Geosci*. 2018;111:18-25.
11. Abedi M, Norouzi GH, Bahroudi A. Support vector machine for multi-classification of mineral prospectivity areas. *Comput Geosci*. 2012;46:272-283.
12. Zuo R, Carranza EJM. Support vector machine: A tool for mapping mineral prospectivity. *Comput Geosci*. 2011;37(12):1967-1975.
13. Parsa M, Maghsoudi A. Assessing the effects of mineral systems-derived exploration targeting criteria for random Forests-based

- predictive mapping of mineral prospectivity in Ahar-Arasbaran area, Iran. *Ore Geol Rev.* 2021;138:104399.
14. Talebi H, Peeters LJM, Otto A, Tolosana-Delgado R. A truly spatial random forests algorithm for geoscience data analysis and modelling. *Math Geosci.* 2022;54(1):1-22.
 15. Maepa F, Smith RS, Tessema A. Support vector machine and artificial neural network modelling of orogenic gold prospectivity mapping in the Swayze greenstone belt, Ontario, Canada. *Ore Geol Rev.* 2021;130:103968.
 16. Li Q, Chen G, Luo L. Mineral prospectivity mapping using attention-based convolutional neural network. *Ore Geol Rev.* 2023;156:105381.
 17. Xu Y, Li Z, Xie Z, Cai H, Niu P, Liu H. Mineral prospectivity mapping by deep learning method in Yawan-Daqiao area, Gansu. *Ore Geol Rev.* 2021;138:104316.
 18. Yang N, Zhang Z, Yang J, Hong Z. Mineral prospectivity prediction by integration of convolutional autoencoder network and random forest. *Nat Resour Res.* 2022;31(3):1103-1119.
 19. Zuo R, Xu Y. Graph deep learning model for mapping mineral prospectivity. *Math Geosci.* 2023;55(1):1-21.
 20. 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(1-2):314-324.
 21. Van Engelen JE, Hoos HH. A survey on semi-supervised learning. *Mach Learn.* 2020;109(2):373-440.
 22. Bruzzone L, Chi M, Marconcini M. A novel transductive SVM for semisupervised classification of remote-sensing images. *IEEE Trans Geosci Remote Sens.* 2006;44(11):3363-3373.
 23. Camps-Valls G, Marsheva TVB, Zhou D. Semi-supervised graph-based hyperspectral image classification. *IEEE Trans Geosci Remote Sens.* 2007;45(10):3044-3054.
 24. Zhu X, Ghahramani Z, Lafferty JD. Semi-supervised learning using gaussian fields and harmonic functions. In: *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*. ; 2003:912-919.
 25. Miller DJ, Uyar H. A mixture of experts classifier with learning based on both labelled and unlabelled data. *Adv Neural Inf Process Syst.* 1996;9.
 26. Salimans T, Goodfellow I, Zaremba W, Cheung V, Radford A, Chen X. Improved techniques for training gans. *Adv Neural Inf Process Syst.* 2016;29.
 27. Fatehi M, Asadi HH. Data integration modeling applied to drill hole

- planning through semi-supervised learning: A case study from the Dalli Cu-Au porphyry deposit in the central Iran. *J African Earth Sci.* 2017;128:147-160.
28. Wang J, Zuo R, Xiong Y. Mapping mineral prospectivity via semi-supervised random forest. *Nat Resour Res.* 2020;29:189-202.
 29. Li Q, Chen G, Wang D. Mineral prospectivity mapping using semi-supervised machine learning. *Math Geosci.* 2025;57(2):275-305.
 30. Chen G, Kusky T, Luo L, Li Q, Cheng Q. Hadean tectonics: Insights from machine learning. *Geology.* 2023;51(8):718-722.
 31. Tao J, Yuan F, Zhang N, Chang J. Three-dimensional prospectivity modeling of Honghai volcanogenic massive sulfide Cu-Zn deposit, Eastern Tianshan, Northwestern China using weights of evidence and fuzzy logic. *Math Geosci.* 2021;53:131-162.
 32. Chen M, Xiao F. Projection pursuit random forest for mineral prospectivity mapping. *Math Geosci.* 2023;55(7):963-987.
 33. Chen G, Huang N, Wu G, Luo L, Wang D, Cheng Q. Mineral prospectivity mapping based on wavelet neural network and Monte Carlo simulations in the Nanling W-Sn metallogenic province. *Ore Geol Rev.* 2022;143:104765.
 34. Bigdeli A, Maghsoudi A, Ghezelbash R. Application of self-organizing map (SOM) and K-means clustering algorithms for portraying geochemical anomaly patterns in Moalleman district, NE Iran. *J Geochemical Explor.* 2022;233:106923.
 35. Zuo R, Xiong Y, Wang J, Carranza EJM. Deep learning and its application in geochemical mapping. *Earth-science Rev.* 2019;192:1-14.
 36. Zuo RG, Peng Y, Li T, Xiong Y. Challenges of geological prospecting big data mining and integration using deep learning algorithms. *Earth Sci.* 2021;46(1):350-358.
 37. McMillan M, Haber E, Peters B, Fohring J. Mineral prospectivity mapping using a VNet convolutional neural network. *Lead Edge.* 2021;40(2):99-105.
 38. Yang N, Zhang Z, Yang J, Hong Z. Applications of data augmentation in mineral prospectivity prediction based on convolutional neural networks. *Comput Geosci.* 2022;161:105075.
 39. Yin B, Zuo R, Sun S. Mineral prospectivity mapping using deep self-attention model. *Nat Resour Res.* 2023;32(1):37-56.
 40. Luo Z, Farahbakhsh E, Müller RD, Zuo R. Multivariate statistical analysis and bespoke deviation network modeling for geochemical anomaly detection of rare earth elements. *Appl Geochemistry.* Published online 2024:106146.
 41. Hosseini-Dinani H, Yazdi M. Multi-dataset analysis to assess

- mineral potential of MVT-type zinc-lead deposits in Malayer-Isfahan metallogenic belt, Iran. *Arab J Geosci.* 2021;14:1-23.
42. Hajihosseini M, Maghsoudi A, Ghezlbash R. Intelligent mapping of geochemical anomalies: Adaptation of DBSCAN and mean-shift clustering approaches. *J Geochemical Explor.* 2024;258:107393.
 43. Yousefi M, Hronsky JMA. Translation of the function of hydrothermal mineralization-related focused fluid flux into a mappable exploration criterion for mineral exploration targeting. *Appl Geochemistry.* 2023;149:105561.
 44. Khalifani FM, Bahroudi A, Aliyari F, Abedi M, Yousefi M, Mohammadpour M. Generation of an efficient structural evidence layer for mineral exploration targeting. *J African Earth Sci.* 2019;160:103609.
 45. Kreuzer M, Fenske N, Schnelzer M, Walsh L. Lung cancer risk at low radon exposure rates in German uranium miners. *Br J Cancer.* 2015;113(9):1367-1369.
 46. Yousefi M, Nykänen V. Data-driven logistic-based weighting of geochemical and geological evidence layers in mineral prospectivity mapping. *J Geochemical Explor.* 2016;164:94-106.
 47. Amirajlo P, Hassani H, Beiranvand Pour A, Habibkhah N. Detection of multivariate geochemical anomalies using machine learning (ML) algorithms in Dehaq Pb-Zn mineralization, Sanandaj-Sirjan zone, Isfahan, Iran. *Earth Sci Informatics.* 2025;18(1):124.