Improving Porphyry Copper Prospectivity Mapping with Stacked Ensemble Machine Learning and Feature Contribution Analysis: A Case Study of the Eastern Sirjan 1:250,000 quadrangle map, Iran

Mohammad Reza Samiei [1], Ali Moradzadeh [2,*], Parham Pahlavani [3], Reza Ghezelbash [2]

1- PhD student, School of Mining, College of Engineering., University of Tehran, Tehran, Iran.
reza.samiei1995@ut.ac.ir
2- School of Mining, College of Engineering., University of Tehran, Tehran, Iran.
a_moradzadeh@ut.ac.ir, r.ghezelbash@ut.ac.ir
3- School of Surveying and Geospatial Engineering, College of Engineering., University of Tehran, Tehran, Iran.
pahlavani@ut.ac.ir

Keywords: Mineral prospectivity mapping, Support Vector Regression, Gaussian Process Regression, Artificial Neural Networks, Stacked Ensemble Learning.

Abstract

Mineral prospectivity mapping (MPM) is an essential tool for reducing exploration costs and risks by identifying areas with high mineralization potential. This study employed Support Vector Regression (SVR) and a multilayer perceptron-based Artificial Neural Network (MLP-ANN) to predict porphyry copper prospectivity. To improve accuracy and minimize predictive uncertainty, a stacked ensemble learning strategy was applied, in which the predictions of the two base models served as inputs to a meta-learner. Gaussian Process Regression (GPR) was selected as the meta-learner due to its nonparametric nature and ability in quantifying uncertainty. In an extended version, the original input features were also incorporated together with the SVR and ANN outputs, thereby enhancing the diversity and richness of meta-model's input space. The study area covers the eastern part of the Sirjan 1:250,000 quadrangle map within the Kerman Cenozoic Magmatic Arc (KCMA). Model performance was evaluated using correlation metrics (R and R^2), error indicators (MAE, MSE and RMSE), and prediction–area (P–A) analysis. Results showed that the stacked ensemble improved predictive performance compared to individual models. Feature contribution analysis further revealed that GPR could identify negatively contributing features. Removing these features led to a refined GPR model under feature filtering (so-called FF-GPR), which increased the R value to 0.98 and the R^2 value to 0.96. This refinement also improved the accuracy of P–A plot and reduced uncertainty. These findings highlight the effectiveness of feature-augmented stacked ensemble learning, particularly when combined with GPR, in advancing MPM and supporting future exploration.

1. Introduction

Mineral prospectivity mapping (MPM) plays a crucial role in contemporary mineral exploration by identifying areas with a high likelihood of hosting mineral deposits. This significantly reduces exploration costs and its associated risks. Despite its importance, one of the primary challenges in MPM lies in the uncertainty of predictive models, which can impact their reliability in supporting informed decision-making (Daviran et al., 2022). To address this issue, machine learning (ML) algorithms have gained prominence due to their ability in handling complex, nonlinear relationships among geological, geochemical, geophysical, and remote sensing data. Both supervised and unsupervised learning methods are widely used in the context of MPM. Unsupervised approaches, such as clustering analysis, focus on uncovering spatial patterns within datasets without requiring prior knowledge of known mineral occurrences (Parsa et al., 2022). In contrast, supervised methods rely on existing mineral deposit data to train predictive models (Cheng et al., 2023). Several machine learning algorithms, including Support Vector Machines (SVM), Random Forest (RF), Multilayer Perceptron (MLP), General Regression Neural Networks (GRNN), and geometric average (GA) models, have been successfully applied in mineral exploration studies, particularly in the Kerman Cenozoic Magmatic Arc (KCMA) of Iran, which is one of the most important metallogenic provinces for porphyry copper deposits (Daviran et al., 2025; Ghezelbash et al., 2025; Ghasemzadeh et al., 2023; Shabankareh and Hezarkhani, 2017).

Despite notable advancements, much of the existing researches in the KCMA has primarily focused on classification techniques aimed at distinguishing prospective areas from non-prospective ones. Conversely, regression-based models have been relatively underutilized despite their ability to produce continuous prospectivity values and capture nuanced variations in mineralization potential. Additionally, tree-based algorithms, such as RF, despite their popularity, often produce broad and generalized anomalies when applied to the available datasets, which compromises the spatial resolution of the prospectivity maps. To overcome these limitations and better capture the continuous nature of mineralization potential, this study employed two regression-based algorithms—Support Vector Regression (SVR) and Artificial Neural Networks (ANN, specifically MLP)—to predict porphyry copper prospectivity. To further enhance the prediction accuracy and improve the quality of the generated prospectivity maps, a stacked ensemble learning framework was developed. In this framework, Gaussian Process Regression (GPR) was utilized as a meta-learner owing to its nonparametric nature and ability to model uncertainty.

* Corresponding author

The ensemble strategy was implemented by feeding the outputs of the two SVR and MLP base models, along with the original input features, into the Stacking Ensemble Gaussian Process Regression (SE-GPR) meta-model. This framework allowed the meta-learner to benefit from the predictive insights of the base models and directly access the raw feature space, potentially capturing interactions and patterns that may not be fully represented in the base outputs alone. Moreover, GPR demonstrated the capability to identify features with negative contributions to the prediction. By filtering out these features, a refined version of the meta-model, termed the Feature-Filtered GPR (FF-GPR), was developed. This enhanced model achieved superior performance and produced spatially coherent prospectivity maps compared to the baseline models. The improved model was also reflected in the prediction–area (P–A) analysis, wherein the ensemble model exhibited improved curve behaviour. Additionally, the reduction in the standard deviation and increase in the mean prediction values indicated a notable decrease in uncertainty.

2. Geological setting

The study area, measuring approximately 25 × 25 km, is located in the southeastern part of the Urmia–Dokhtar Magmatic Arc (UDMA), within the 1:250,000 Sirjan quadrangle map (Hosseini et al., 1995) (Figure 1). This region, widely known as the KCMA or Kerman Porphyry Copper Belt (KPCB), is one of the most important metallogenic provinces in Iran. The KCMA region, dominated by Eocene–Oligocene volcanic–plutonic complexes such as Bahr-e-Asman, Razak, and Hezar, emerged through nearly 60 million years of Cenozoic magmatic activity (Shafiei et al., 2009). Geological and geochemical assessments have categorized the granitoids as the main host rocks for porphyry copper deposits within the KCMA into two distinct groups based on their metallogenic traits. The northwestern granitoids, classified as the Kuh Panj type, were emplaced during the Middle–Late Miocene and exhibit high fertility, closely linked to porphyry copper mineralization. Conversely, the southeastern granitoids, predominantly the Jebal Barez type that formed during the Eocene–Oligocene, are generally infertile with minimal economic mineralization potential (Taghipour et al., 2008; Shafiei et al., 2009). The study area is situated in the northwestern part of the KCMA, where Kuh Panj-type granitoids represent the primary host rocks for copper mineralization. Accordingly, these fertile granitoids are considered the main mineralization-related units in this study.

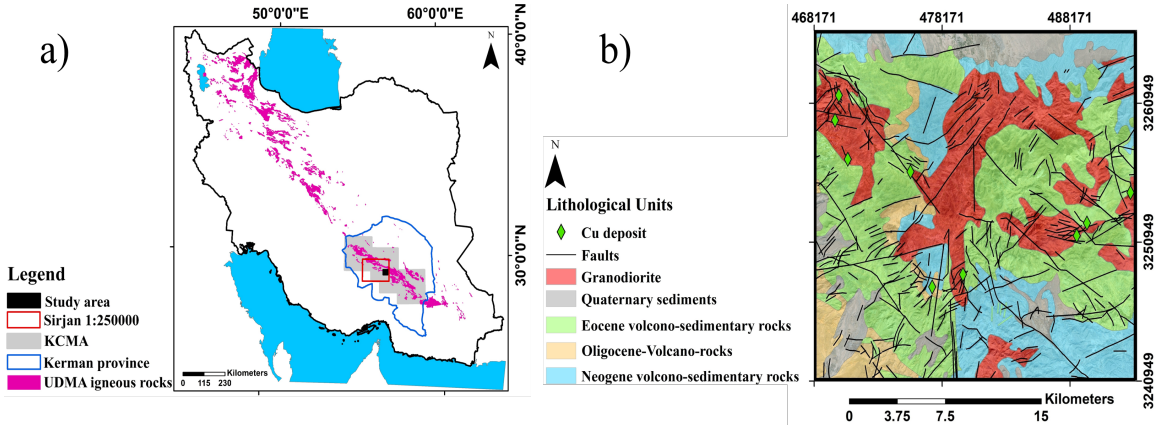

Figure 1. a) The position of study area on KCMA in SE part of Iran, and b) Simplified geological map of the study area, within the Sirjan 1:250,000 quadrangle (Hosseini et al., 1995).

3. Material and methods

In this study, ten exploration evidence layers were prepared to represent the geological, geochemical, geophysical, and remote sensing criteria associated with porphyry copper mineralization. A summary of these layers is presented in Table 1. To generate the geological and structural layers, distance to and density maps of faults and intrusive rocks were derived from the 1:250,000 geological map of Sirjan and supplementary remote sensing data. In addition, copper geochemical anomalies were extracted from stream sediment geochemistry, and magnetic data (including Reduction to the Pole (RTP) and Analytic Signal (AS)) and potassium radiometry were used to enhance the geophysical representation. Various remote sensing data (ASTER and OLI) were utilized to identify and extract hydrothermal alterations, focusing primarily on phyllic, argillic, and iron oxide zones, which are considered critical exploration indicators of porphyry copper deposits in the KCMA.

3.1 Support Vector Regression and Artificial Neural Networks

SVR and MLP-ANN are widely used machine learning algorithms in MPM due to their ability to capture nonlinear relationships between evidential features and mineralization targets (Jung and Choi, 2021; Dumakor-Dupey and Arya, 2021). A detailed description of these algorithms is beyond the scope of this paper, and the reader is referred to the cited works for further information.

3.2 Gaussian Process Regression

Gaussian Process Regression (GPR) is a nonparametric, Bayesian-based regression method particularly suitable for small datasets and complex nonlinear problems (Williams and Rasmussen, 2006). A Gaussian process describes any finite set of random variables that follows a joint Gaussian distribution and is characterized by its mean m(x) and variance functions k (x, x') as follow:

$$f(x) \sim GP(m(x), k(x, x')) \quad (1) \qquad m(x) = E[f(x)],\ k(x, x') = E[(f(x) - m(x))(f(x') - m(x'))] \quad (2)$$

Dataset	Source	Predictor maps
Geology	Geological Survey of Iran (Hosseini et al., 1995)	Three maps included: Measured distance from intrusive rocks (Granodiorites), Fault density and Fault NW- trending
Geophysics	Airborne magnetic & radiometry data (collected by AEOI)	Three maps included: RTP, analytical signal and potassium alteration
Geochemistry	Geological Survey of Iran (Hosseini et al., 1995)	A map includes: Cu geochemical signature
Remote sensing	OLI and ASTER satellites	Three maps included: Phyllic, Argillic and Iron-oxide alteration

Table 1. Summary of the targeting criteria applied in the creation of exploration layers.

In most practical applications, the mean function is assumed to be zero, and the covariance is defined by a kernel such as the Radial Basis Function (RBF):

$$K(x, x') = \sigma_0^2 * \exp(-(x - x')^2 / (2\ell^2)) \quad (3)$$

Given training data X and targets y, the predictive distribution at a new input x* is Gaussian with following mean and variance:

$$\mu(x^*) = k(x^*, X)\,[K(X, X) + \sigma_n^2 I]^{-1}\, y \quad (4)$$

$$\sigma^2(x^*) = k(x^*, x^*) - k(x^*, X)\,[K(X, X) + \sigma_n^2 I]^{-1} k(X, x^*) \quad (5)$$

This framework enables not only accurate predictions but also estimation of prediction uncertainty, which is crucial in exploration studies where data are often sparse and noisy. The GPR model was implemented in the R environment, whereas the SVR and ANN models were developed in Python using standard ML libraries.

3.3 Stacked Ensemble Machine Learning

Ensemble learning integrates multiple predictive models into a single framework to improve robustness and generalization (Sun and Trevor, 2018). Among the main ensemble approaches—bagging, boosting, and stacking—the stacking method has gained particular attention because it combines the strengths of diverse base learners using a layered architecture. In traditional stacking, the outputs of the base models (Level 0) are passed to a meta-learner (Level 1), which uses them as new inputs to produce the final predictions. This process often replaces or reduces the original feature space by focusing solely on model outputs. However, in this study, a modified stacking strategy was applied. Instead of discarding the original features, the predictions of the base learners were incorporated as additional attributes with the geological, geochemical, remote sensing and geophysical variables. Thus, the feature space was enriched rather than reduced, allowing the meta-learner to simultaneously benefit from both the original data and the complementary information provided by the base models. This approach increases the diversity and reliability of the ensemble while mitigating the risk of bias from individual learners.

4. Results and discussion

Two types of labelled training samples were used: (1) mineralized locations (label = 1), provided by the Geological Survey of Iran, and (2) non-mineralized locations (label = 0). The selection of non-mineralized sites followed a point-pattern analysis approach (Carranza et al., 2008b). Based on this method, a buffer distance of 2850 m from the mineralized points was determined as the optimal threshold for distinguishing non-deposit areas. Furthermore, to ensure both randomness and spatial coverage, a 2,500 m grid was superimposed on the study area, and nine candidate points were randomly selected by the computer. The final number of non-deposit points was matched with that of deposit points to maintain a balanced dataset. The study area was geographically divided into training and validation zones, allocating approximately 70% of the data for model training and 30% for independent validation, ensuring robust performance assessment of the machine learning models (Figure 2).

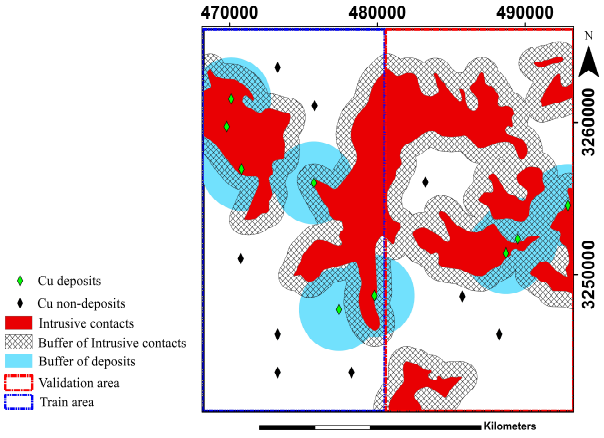

Figure 2. Locations of deposit and non-deposit samples within the study area. The blue region on the left includes 70% of the data used for training, while region on the right represents 30% of the data used for validation.

To achieve the goal, all ten evidential layers representing geological, geochemical, geophysical, and remote sensing parameters were constructed with a cell size of 100m. Accordingly, each prospectivity map consisted of 62,500 pixels. By applying a 500-meter buffer around both deposit and non-deposit points, 1,388 training pixels were extracted. Considering the 10 exploration layers used as predictors, this resulted in a training matrix with dimensions of 10 × 1388, corresponding to the known mineral occurrences and background locations.

Hyperparameter tuning for all models was performed using a grid search combined with k-fold cross-validation to avoid overfitting and ensure robust performance estimation. Following model optimization, the SVR and ANN algorithms were trained and used to generate porphyry copper prospectivity maps (Figure 3a and 3b).

Based on the generated prospectivity maps and model evaluation results, the SVR regression model outperformed the ANN model. According to the P–A curves, SVR detected a similar number of mineral occurrences within a smaller area, indicating higher spatial efficiency (Table 3). Additionally, SVR achieved higher mean scores and lower standard deviations, along with greater accuracy and reduced error rates, confirming its robustness for porphyry copper prospectivity mapping (Tables 2 and 4).

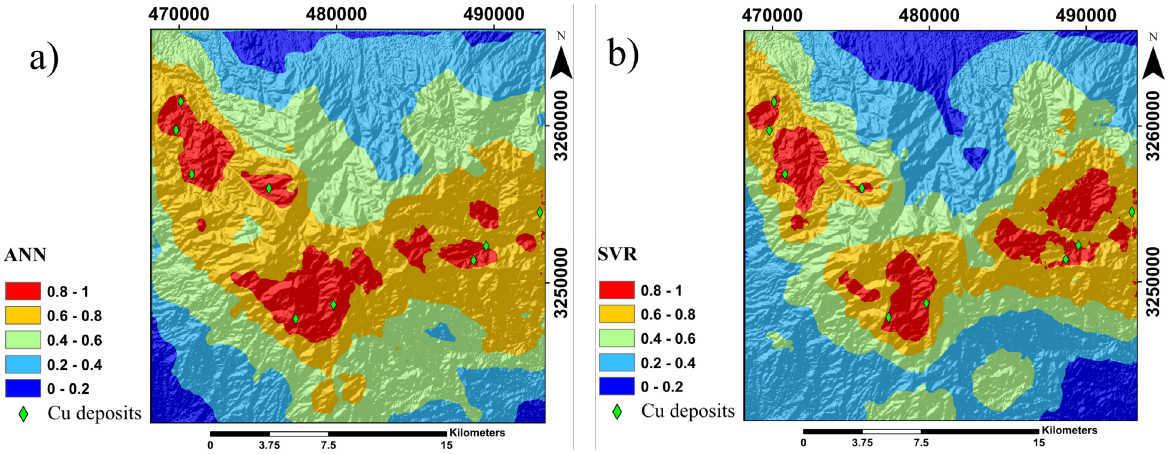

Figure 3. Porphyry copper prospectivity maps generated for the study area using: a) ANN and b) SVR algorithms.

Afterward, to further assess the effectiveness of the stacked ensemble learning in MPM, the GPR algorithm was employed as a meta-learner to improve the accuracy of the prospectivity predictions. In this context, which is named as SE-GPR model, the outputs of the two SVR and ANN base models, along with the original input features, were provided as inputs to the GPR meta-model. This integration led to the generation of an enhanced prospectivity map (Figure 4a). The evaluation metrics for the GPR meta-model showed improved performance compared to the individual base models (Table 2), with the correlation coefficient (R) increasing up to 0.90 and the coefficient of determination (R^2) reaching 0.82. Additionally, the P–A plot showed an improved curve for the ensemble-generated map (Table 3), where the SE-GPR model successfully identified 89% of the known mineral occurrences within only 11% of the study area. As expected, incorporating the original features alongside the base model outputs assists the meta-learner in increasing insights into the contribution of each input variable. As a result, GPR identified three features—RTP, Analytic Signal, and Potassium—as having negative impacts on prediction accuracy (Figure 4b). The permutation feature importance method was used to assess the feature importance.

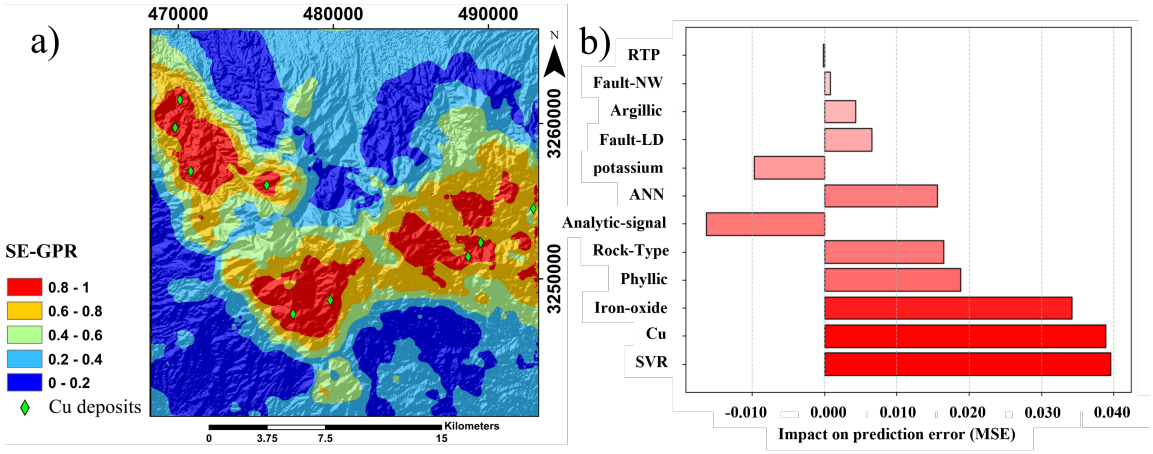

Figure 4. a) Porphyry copper prospectivity map generated by the SE-GPR model for the study area; b) Permutation feature importance chart illustrating the contribution of each input variable to the SE-GPR model.

Subsequently, these negatively contributing features were removed, and the meta-model was retrained using a feature-filtering strategy, resulting in a refined model referred to as FF-GPR. The FF-GPR model produced an optimized prospectivity map (Figure 5a) and demonstrated significant improvements in evaluation metrics, with the coefficient of determination R^2 increasing to 0.96. The model also enhanced the prediction–area (P–A) curve, successfully identifying 90% of the mineralization points within only 10% of the study area (Figure 5b).

Parameters	ANN	SVR	SE-GPR	FF-GPR
R	0.80	0.84	0.90	0.98
R^2	0.65	0.71	0.82	0.96
MAE	0.26	0.19	0.25	0.15
MSE	0.08	0.07	0.06	0.02
RMSE	0.29	0.26	0.20	0.13

Table 2. Evaluation results of the developed models for predicting porphyry copper prospectivity in the study area.

Lower standard deviation and higher mean values are widely recognized as indicators of reduced uncertainty in mineral prospectivity mapping processes (Daviran et al., 2022). In line with this, the results of the present study demonstrate that the standard deviation decreased to 0.008, whereas the mean value increased to 0.53 for the FF-GPR model, indicating a substantial reduction in prediction uncertainty and enhanced reliability. Among all the evaluated models, FF-GPR exhibited the lowest uncertainty, confirming its robustness and suitability for porphyry copper exploration within the study area (Table 4).

Targeting criteria	P_r	O_a	N_d
ANN	0.81	0.19	4.26
SVR	0.84	0.16	5.25
SE-GPR	0.89	0.11	8.09
FF-GPR	0.9	0.1	9

Table3. Evaluating the effectiveness of exploration targeting models using normalized density indices for the study area., P_r: prediction rate, O_a: occupied area, N_d: normalized density.

Parameters	ANN	SVR	SE-GPR	FF-GPR
Mean Standard	0.39	0.42	0.47	0.53
deviation	0.30	0.28	0.012	0.008

Table4. Table 4. Standard deviation and mean values of the prospectivity maps generated by each model.

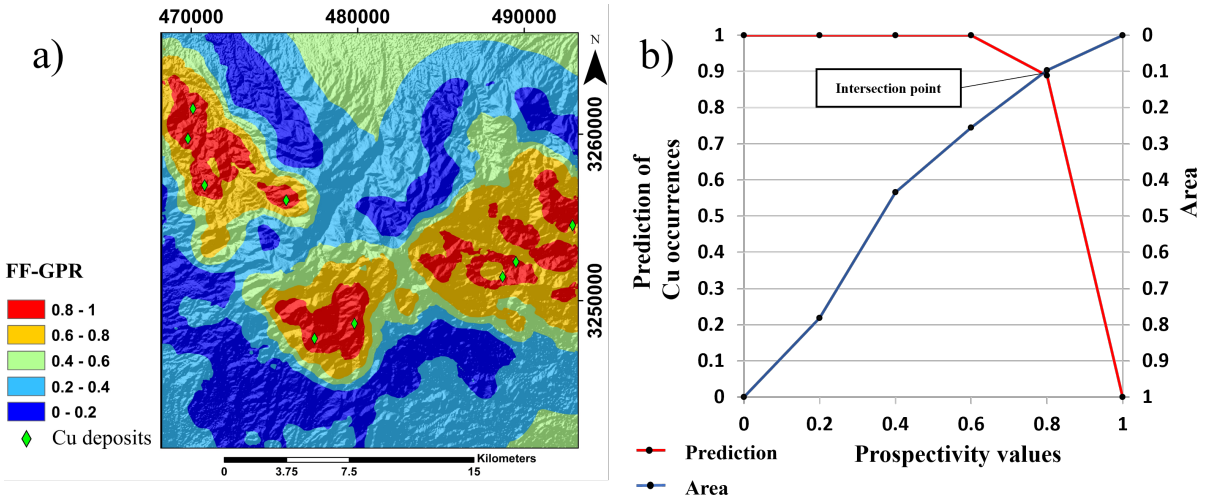

Figure 5. a) Prospectivity map generated by the FF-GPR model; b) P–A curve illustrating the spatial prediction performance of the FF-GPR model within the study area.

5. Conclusions

This study highlighted the effectiveness of regression-based machine learning techniques for porphyry copper prospectivity mapping in the KCMA. The SVR and ANN models served as baseline models, with their outputs combined with the original input features in a stacking ensemble outline. This agenda utilized the GPR model as a meta-learner (SE-GPR), which significantly enhanced both the predictive accuracy and spatial resolution of the generated prospectivity maps. An additional refinement step involved feature filtering, leading to the development of the FF-GPR model, which outperformed all other approaches in this study. The FF-GPR achieved a R^2 value of 0.96. It successfully detected 90% of the known mineralization points while covering only 10% of the study area, as demonstrated by the P-A analysis. Moreover, it reduced the standard deviation to 0.008 and increased the mean prediction value to 0.53. These outcomes indicate that FF-GPR not only improves predictive performance but also reduces uncertainty, establishing it as a competent and dependable tool for advancing porphyry copper exploration.

References

Carranza, E.J.M., Hale, M., Faassen, C., 2008: Selection of coherent deposit-type locations and their application in data-

driven mineral prospectivity mapping. Ore Geol. Rev., 33(3-4), 536-558. doi.org/10.1016/j.oregeorev.2007.07.001.

Cheng, H., Zheng, Y., Wu, S., Lin, Y., Gao, F., Lin, D., Wei, J., Wang, S., Shu, D., Wei, S., Chen, L., 2023: GIS-based mineral prospectivity mapping using machine learning methods: a case study from Zhuonuo ore district, Tibet. Ore Geol. Rev., 161, 105627. doi.org/10.1016/j.oregeorev.2023.105627.

Daviran, M., Parsa, M., Maghsoudi, A., Ghezelbash, R., 2022: Quantifying uncertainties linked to the diversity of mathematical frameworks in knowledge-driven mineral prospectivity mapping. Nat. Resour. Res., 31(5), 2271-2287. doi.org/10.1007/s11053-022-10089-w.

Daviran, M., Maghsoudi, A., Ghezelbash, R., 2025: Optimized AI-MPM: Application of PSO for tuning the hyperparameters of SVM and RF algorithms. Comput. Geosci., 195, 105785. doi.org/10.1016/j.cageo.2025.105785.

Dumakor-Dupey, N.K., Arya, S., 2021: Machine learning—a review of applications in mineral resource estimation. Energies, 14(14), 4079. doi.org/10.3390/en14144079.

Ghasemzadeh, S., Maghsoudi, A., Yousefi, M., Kreuzer, O., 2023: Spatially weighted singularity mapping in conjunction with random forest algorithm for mineral prospectivity modeling. Int. J. Min. Geo-Eng., 57(4), 455-460. doi.org/10.22059/ijmge.2023.366283.595102.

Ghezelbash, R., Daviran, M., Maghsoudi, A., 2025: Combination of Machine Learning and Fractal approaches for AI-MPM: Identifying Low-Risk Exploration Targets associated with Porphyry-Cu Deposits in the Kerman Belt, Iran. Remote Sens. Appl. Soc. Environ., 101596. doi.org/10.1016/j.rsase.2025.101596.

Hosseini, Z., Ghaemi, J., Mohabbi, A.R., 1995: Geological Map of Sirjan. Geological Survey of Iran, Tehran, Iran (1:250,000).

Hou, S., Liu, Y., Yang, Q., 2022: Real-time prediction of rock mass classification based on TBM operation big data and stacking technique of ensemble learning. J. Rock Mech. Geotech. Eng., 14(1), 123-143. doi.org/10.1016/j.jrmge.2021.05.004.

Jung, D., Choi, Y., 2021: Systematic review of machine learning applications in mining: Exploration, exploitation, and reclamation. Minerals, 11(2), 148. doi.org/10.3390/min11020148.

Parsa, M., Carranza, E.J.M., Ahmadi, B., 2022: Deep GMDH neural networks for predictive mapping of mineral prospectivity in terrains hosting few but large mineral deposits. Nat. Resour. Res., 31(1), 37-50. doi.org/10.1007/s11053-021-09984-5.

Shabankareh, M., Hezarkhani, A., 2017: Application of support vector machines for copper potential mapping in Kerman region, Iran. J. Afr. Earth Sci., 128, 116-126. doi.org/10.1016/j.jafrearsci.2016.11.032.

Shafiei, B., Haschke, M., Shahabpour, J., 2009: Recycling of orogenic arc crust triggers porphyry Cu mineralization in Kerman Cenozoic arc rocks, southeastern Iran. Miner. Deposita, 44, 265-283. doi.org/10.1007/s00126-008-0216-0.

Sun, W., Trevor, B., 2018: A stacking ensemble learning framework for annual river ice breakup dates. J. Hydrol., 561, 636-650. doi.org/10.1016/j.jhydrol.2018.04.008.

Taghipour, N., Aftabi, A., Mathur, R., 2008: Geology and Re-Os Geochronology of Mineralization of the Miduk Porphyry Copper Deposit, Iran. Resour. Geol., 58(2), 143-160. doi.org/10.1111/j.1751-3928.2008.00054.x.

Williams, C.K., Rasmussen, C.E., 2006: Gaussian processes for machine learning. Cambridge, MA: MIT Press.