Journal of Mining and Environment (JME), Vol. 14, No. 4, 2023, 1321-1342

Shahrood University of Technology

Journal of Mining and Environment (JME)

Journal homepage: www.jme.shahroodut.ac.ir

Iranian Society of Mining Engineering (IRSME)

Machine Learning Predictive Approaches for Cu-Au Mineral prospectivity Map in Sonajil, NW of Iran: an Improvement by a Bayesian Semi-supervised Algorithm

Mohammadzadeh Mohammadjafar[1]*, Mahboubiaghdam Majid[1], Jahangiri Moharram[2], and Nasseri Aynur[3]

1. Mining Engineering Faculty, Sahand University of Technology, Tabriz, Iran
2. Faculty of Mining, Petroleum & Geophysics Engineering, Shahrood University of Technology, Shahrood, Iran
3. Department of Mining Engineering, Ahar Branch, Islamic Azad University, Ahar, Iran

Article Info

Received 30 June 2023
Received in Revised form 29 July 2023
Accepted 7 August 2023
Published online 7 August 2023

DOI: 10.22044/jme.2023.13313.2445

Keywords

Cu-Au mineral prospectivity map
Bayesian semi-supervised method
Sonajil
E-Azerbaijan

Abstract

Most machine learning-monitored algorithms used to create mineral potential prediction maps require noise-free data to achieve high performance and reliable results. Unsupervised clustering methods are highly effective for uncovering a dataset's hidden structures. Therefore, this study attempts a combination of supervised and unsupervised methods employing training and testing data to generate a highly accurate potential map of the Sonajil copper-gold deposit located in the NW of Iran. Here, a semi-supervised Bayesian algorithm is used to map the mineral landscape. Initially, ten raster layers of exploratory features are prepared. Then based on the copper concentration, 27 exploratory drilled boreholes are divided into four classes, C1 to C4, and from each class, two boreholes are selected, and 100-meter buffering is performed around these boreholes to extract 1113 training data based on the behavioral pattern of boreholes and surface samples. Subsequently, the existing data is clustered using the FCM method, and the total dataset and the clustering data are entered into the Bayesian algorithm to evaluate the accuracy of the Bayesian classifier method across five distinct clusters. The results show increased average accuracy when using clustered data instead of whole data for MPM mapping. Notably, the Bayesian semi-supervised algorithm achieved an impressive accuracy rate of 96% when cluster five data is excluded. To validate the Bayesian semi-supervised method, boreholes data that is not used in training were employed, which confirm the credibility of generated MPM. Overall results highlight the value of the Bayesian semi-supervised algorithm in improving the accuracy and reliability of mineral prospectivity mapping via the application of the FCM clustering method that efficiently organize the data, enabling the Bayesian algorithm to evaluate the accuracy of the Bayesian classifier method across different clusters and providing a successful optimal result in detecting blind ores in areas without exploratory boreholes and delineating more mineralization targets in the Sonajil and adjoining areas.

1. Introduction

Exploration for mineral deposits with the aim of delimiting the area for promising zones involves contemporaneous consideration of various parameters such as lithological, geochemical, geophysical, structural, alterations, and remote sensing datasets [1, 2]. Combining these datasets as a basis for establishing an appropriate mineral prospecting map (MPM) is the optimal task for the recent exploration programs [3-6]. MPM is a crucial procedure in mineral exploration programs, aiming to identify areas with high potential for undiscovered mineral deposits. In the recent years, there has been a growing interest in applying machine learning (ML) techniques to enhance the

Corresponding author: mohammadzadeh@sut.ac.ir (M. Mohammadjafar)

accuracy and efficiency of mineral prospectivity mapping [56-61]. In order to prepare the potential map, the ML techniques have been developed into four main groups: supervised, unsupervised, semi-supervised, and reinforcement [7]. Common methods of preparing the mineral potential map that belongs to supervised models include weights of evidence [8], fuzzy weights of evidence[9], boosted weights of evidence[10], support vector machine [11, 12], random forest [13-15], logistic regression and its variants [16, 17], neural networks [18-20], Bayesian networks [21], evidence belief functions [22], to name a few. The main techniques of unsupervised approach include Boolean logic [23], index overlay [24], Dempster-Shafer evidence theory [25], fuzzy logic [26], data envelopment analysis [27], self-organizing map (SOM), and K-means clustering [28]. In supervised learning, which is one of the most successful types of ML for detecting geochemical anomalies, the algorithm is built based on the desired input (training) data [29-31]. One example in mineral prospectivity mapping is the use of supervised mineral exploration targeting and the challenges with the selection of deposit and non-deposit sites presented by Rahimi *et al*. (2021) [32]. The paper revealed that the selection of non-deposit sites is a challenging issue affecting the application of supervised algorithms for modeling mineral exploration targets. The study found that exploration targeting models are affected by the ratio of non-deposit and deposit sites. Therefore, balancing between the number of deposit and non-deposit sites is an efficient way to produce more-reliable exploration targets when supervised algorithms are applied for modeling. The proposed procedure for selecting non-deposit sites can be used to improve the effectiveness of exploration targeting models. The study used an exploration dataset of porphyry Cu mineralization in the Chahargonbad area, SE Iran, and applied a sequence application of a self-organizing map and multi-layer perceptron neural network algorithm to better illustrate the changing effects of the number of non-deposit sites on the ensuing exploration targeting models [32]. Another study presented by Yusefi and Hornsky (2023) [33] is about the translation of the function of hydrothermal mineralization-related focused fluid flux into a mappable exploration criterion for mineral exploration targeting [33]. The paper proposes a simplified approach for mineral exploration targeting by focusing on a geological point feature that represents strongly concentrated ore fluid flux, which can be derived as a mappable criterion. The authors demonstrate that this point feature can be used as a more effective criterion for mineral exploration targeting than existing point and line features, such as intrusive contacts, fault density, fault intersections, and proximity to faults. The proposed approach is illustrated using geochemical and geological datasets of porphyry Cu deposits of Iran. The paper concludes that the proposed approach can help mineral exploration companies to identify areas with a higher potential for hydrothermal mineral deposits, which can lead to more efficient and cost-effective exploration programs. A study about the data analysis methods for prospectivity modeling as applied to mineral exploration targeting: State-of-the-Art and Outlook was proposed by Yusefi *et al*. (2021) [34]. The paper discusses the use of Geographic Information Systems (GIS) for mineral exploration targeting through mineral potential modeling or mineral prospectivity mapping (MPM). The authors review the fundamental aspects of MPM, identify significant deficiencies of MPM, and propose the use of an intelligence amplification system such as the exploration information system (EIS), for improving decision-making in mineral exploration targeting [34]. Considering that in mineral deposits we are dealing with the development of complex lithology and multiple stages of mineralization, therefore, it is better to homogenize the data first so that the training data can be optimally entered into the classification algorithm. In the recent years, unsupervised learning models have been utilized to map potential mineral areas in situations where exploratory data is dispersed. These models (e.g. C-means) aim to determine the underlying distribution of predictor variables and then use this information to guide MPM [32, 33]. Therefore, it is better to use semi-supervised methods, which are a combination of supervised and unsupervised methods. This method was successfully used by [34] to prepare a mineral potential map using semi-supervised random forest [35] and to identify anomalous areas related to copper-gold porphyry mineralization from the semi-supervised TSVM algorithm [36], as well also from the hybrid machine learning method (HML) based on the combination of nearest neighbour regression (KNNR) and random forest regression (RFR) for Pb and Zn prediction. In addition, K-nearest neighbors semi-supervised method has been employed by [37] to detect and reveal regional anomalies. Also, Du *et al*. (2021) [56] developed a hybrid genetic algorithm-support vector machine model for mapping prospective areas for gold deposits in Karamay, northwest China. The genetic

algorithm was used to optimize the parameters of the support vector machine model, resulting in improved performance in identifying prospective areas for gold deposits [56]. Despite the success of machine learning approaches in mineral prospectivity mapping, there are still challenges to overcome. One challenge is the limited availability of training data, as the formation of mineral resources involves multiple factors and rare labeled deposits [58]. Another challenge is the interpretability of machine learning models, as they often lack a proper combination with geoscience knowledge [58]. The researchers are actively working on addressing these challenges and developing new approaches to improve the accuracy and interpretability of mineral prospectivity mapping models [58]. Each of the methods has its advantages and drawbacks. This study demonstrates that the deficiencies can be remedied by implementing a robust method using a Bayesian semi-supervised algorithm. This is essential when dealing with complex lithological evolution involving multiple mineralization phases.

In most of the exploration programs, the algorithm is chosen based on ease of implementation, and the reliability of the available thematic spatial data sets. Therefore, the method should be transparent and practical [38-40]. But in using classification algorithms, the condition of optimal use of training data is raised. This means that if training data are used generally and without clustering, we will encounter the issue of data noise due to sampling or analysis or human error in network learning. For this purpose, clustering algorithms will be used to break the total space of the dataset into several homogeneous subspaces to improve the accuracy of the classifier during data training. In fact, information extracted in this manner without supervision can be used to classify samples into mineralized and non-mineralized regions.

Based on the above background, a semi-supervised Bayesian algorithm was used in the present study to generate a potential MPM map for the Sonajil region in NW Iran. Specifically, it employed the FCM clustering method to cluster training data, followed by the Bayesian algorithm to classify supervised learning, which can assist in improving MPM performance. Ultimately, the goal is to develop tools that can augment and assist geologists, not replace them. The best outcomes will come from combining machine intelligence with human intuition, knowledge, and expertise.

2. Geology and Metallogeny of Area

The Sonajil Cu-Au deposit is a part of Alp Himalayan metallogenic belt, entering to NW of Iran. The zone belongs to the Alborz Azerbaijan structural trend known as Arasbaran mineralization zone (AMZ), located at about 100 km NE of Tabriz City (Figure 1). The most significant types of mineralization in the (AMZ) are associated with Cenozoic intrusions, consisting of porphyry, skarn and epithermal deposits of 250 km long trending NW-SE direction. This zone is comparable to the “And” type porphyry metallogenic belt of South America. The distribution of both Cu-Mo and Cu-Au types of porphyry mineralization are present in the (AMZ). The most important Cu-Mo porphyry deposits of this zone formed in the northern part of the area such as Sungun and Baluja deposits whereas the southern part contains Cu-Au deposits of MirKuh Ali Mirza, Masjid Daghi and Shalloo.

The Sonajil area is located in the central part of the Ahar province of the AMZ zone, which is predominantly covered by semi-deep Eocene-Oligocene volcanic rocks. Structurally, the mineralized faults around the Sonajil intrusive mass have been observed trending NW-SE direction. Based on field observations, the outcrops in the area mainly include Eocene sedimentary volcanics, Oligo-Miocene intrusive masses, Miocene sedimentary rocks, as well as Quaternary basalts.

Eocene deposits composed of thick sequences of andesitic, basaltic, pyroclastic, and sedimentary rocks that are mainly distributed in the southern parts of the area. These rocks have undergone mild metamorphism in the vicinity of the intrusive masses and at the periphery of the Sonajil porphyry intrusive and subjected to various alterations.

The oldest intrusive masses in the region are porphyry monzo-syenite to micro monzo-syenite that are exposed in the center of the exploration area. The intrusion of the Sonajil porphyry to these rocks especially into micro monzo-syenite has caused a vast alteration system as well as the emplacement of part of the sulphides mineralization within these masses. The Sonajil massif with quartz monzodiorite to granodiorite composition is the cause of porphyry mineralization in the region. Various alteration systems in the region include potassic, phyllic, argillic, and propylitic zones. No evidence of a dominant potassic zone in the field and no K-feldspars are observed in the hand specimen. This may indicate lower level of erosion in the studied area and preservation of ore at depth. The argillic

alteration is mostly seen at the surface with the white kaolinite formation. Remarkably, the sulfide mineralization and mineralized quartz zone in the form of vein-veinlets are associated with the potassic zone at depth and partly with phyllic alteration. Later the area was subjected to another intrusion known as (incheh) with diorite to quartz-monzodiorite composition, exposed in the northern part of the area without significant alteration and mineralization.

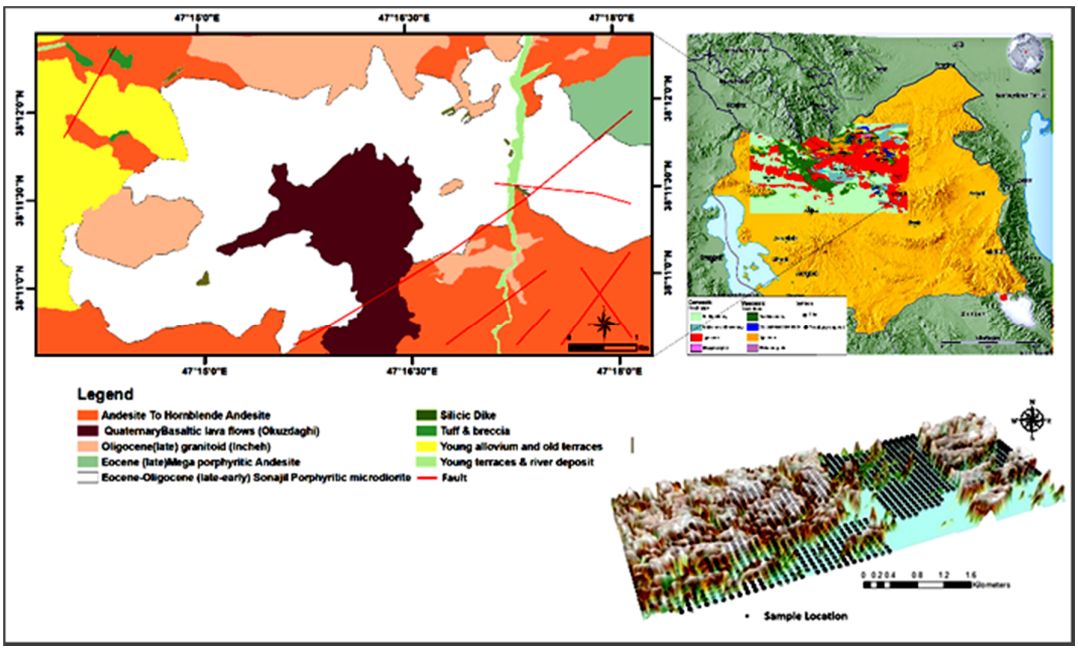

Figure 1. Geological map of the Sonajil area and the lithogeochemical sampling points.

From an ore point of view, [41] and [42] have reported (< 1-7%) pyrite within the veinlets, Chalcopyrite (0-3%) as the major sulfide mineral within the Sonajil porphyry intrusion, and have stated the depletion of Molybdenite in Sonajil compared to similar mineralization in the NW of Iran. Molybdenite can rarely be seen in the marginal part of the main intrusion in association with quartz veinlets in the jungle (forest) valley.

3. Methods

3.1. Naïve Bayes

One of the methods widely used to classify high-dimensional data is Naïve Bayes (NB). With Bayesian networks, features are weighed concerning one another [43, 44]. The features' independence in terms of class-condition is usually an optimistic condition for many classification jobs [45]. When such a condition is met, the NB classifier provides more efficient classification parameters and requires less training data than other classifiers. Nonetheless, this classifier may function correctly in practice without assuming independence. Regarding the data specifications utilized in this study, kernel and Gaussian distributions were taken into account to evaluate the efficiency of preparing potential maps. The Gaussian distribution is superior when the features are normally distributed across all classes. In contrast, the kernel distribution is superior when the feature distribution is skewed because it has multiple modes.

The Naïve Bayes classifier function is explained by [46]: assuming an instance, X, the NB classifier predicts that X belongs to the class for which is maximized. The class C_i represents the hypothesis of maximum posteriori based on Bayes's theorem as: Equation (1).

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \quad (1)$$

To minimize the calculation load in estimating $P(X|C_i)$ for high-dimensional datasets, class conditional independence (the Naïve assumption) is taken into account. Thus:

$$P(X|C_i) = \prod_{k=1}^{n} P(x_k|C_i) = P(x_1|C_i) * P(x_2|C_i) * P(x_3|C_i) * \dots\dots * P(x_n|C_i) \quad (2)$$

where x_k represents the attribute A_k such as X. Training data can yield more than the simple probabilities in Equation 2. Typically, the continuous variable has a Gaussian distribution with a mean of μ and a standard deviation (SD) of σ, as represented by Equations 3 and 4 [46].

$$g(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(x-\mu)^2}{2\sigma^2}} \quad (3)$$

$$P(x_k|C_i) = g(x_k, \mu_{c_i}, \sigma_{c_i}) \quad (4)$$

where μ_{c_i} and σ_{c_i} represent the mean and SD of the attribute A_k for training class C_i. To determine the class label of X, $P(X|C_i)P(C_i)$ is obtained for every C_i class. The classifier indicates that the class label of instance X is C_i if and only if Equation 5 is satisfied [46].

$$P(X|C_i)P(C_i) > P(X|C_j)P(C_j) \quad \text{for } 1 \le j \le m, j \ne i \quad (5)$$

3.2 Fuzzy C-means (FCM) clustering

There are many fields such as pattern recognition and machine vision that use fuzzy clustering to solve problems. Several fuzzy clustering algorithms employ distance criteria including FCM, which utilizes reverse distance for fuzzy membership. In this case, feature vectors are part of all clusters with a zero to one coefficient [47, 48].

In addition, each data point in algorithms (feature vector) is labeled according to the data point with the highest coefficient in each cluster. The cluster centers and fuzzy membership matrix are determined through minimizing the formula below [48]:

$$J_f(C, m) = \sum_{i=1}^{c} \sum_{k=1}^{n} (u_{i,k})^m d_{i,k} \quad (6)$$

$$\sum_{i=1}^{c} u_{i,k} = 1$$

where C represents the number of clusters, $d_{i,k}$ refers to the Euclidean distance of the cluster center and the data point, n is the number of data, $u_{i,k}$ is the fuzzy membership of the kth data point to the ith cluster, $m \in (1, \infty)$ is a fuzzy weighting factor that determined the fuzziness level of the outcomes.

With increasing m, data class degrades in terms of discriminating fuzzier. Generally, m = 2 is selected and it is notable that this m produces no optimal solutions for every problem. The constraints in (6) means that every point needs to totally distribute the membership between the clusters [48]. The centers of cluster represent the fuzzy weighted center of gravity of the data X.

$$V_i = \frac{\sum_{k=1}^{n} (u_{i,k})^m x_k}{\sum_{k=1}^{n} (u_{i,k})^m}, i = 1,2, \dots, c \quad (7)$$

Because $u_{i,k}$ influences the computation load of the cluster center v_i, the data that has a higher membership affects the location of the prototype more than points that have a low membership[48]. In the case of fuzzy C-means algorithm, $d_{i,k}$ is:

$$(d_{i,k})^2 = \|x_k - v_i\|^2 \quad (8)$$

The cluster center v_i indicates the usual values that a cluster can have; while the $u_{i,k}$ element of the membership matrix indicates the extent of data point x_k is identical to its prototype [48]. Through minimizing the partition functional in Equation 6, we have:

$$u_{i,k} = \frac{1}{\sum_{j=1}^{c} \left(\frac{d_{i,k}}{d_{j,k}}\right)^{1/(m-1)}} \quad (9)$$

Equation (9) is determined in an iterative way since the distance $d_{i,k}$ depends on membership $u_{i,k}$.

Figure 2 shows the flowchart of the applied FCM method in the present study.

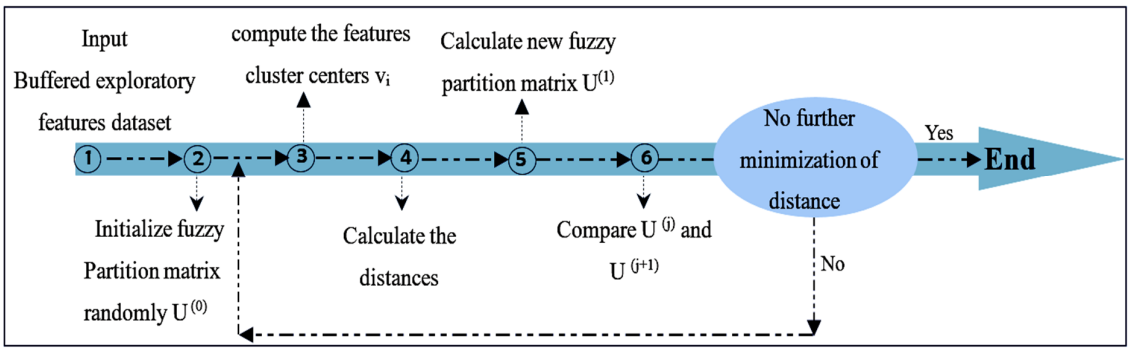

Figure 2. Flowchart of FCM method applied in the studied area.

3.3. Validation indices of fuzzy clustering algorithms

The process of determining the optimal number of clusters should be as interpretable as possible. To evaluate clustering results, validation criteria are used to determine the compaction or density (the members of each cluster should be as close to one another as possible) and separation (the clusters themselves should be widely separated) of clustering results. The first group (compression) only utilizes clustering members, whereas the second group (separation) uses data-related members. Partition coefficient [49] and classification entropy are frequently employed in the first group. In the second group, the Xie-Beni index [50], the partition index [51], and the separation index [51] are frequently validated. The following are the parameters of the validation indicators used in this study:

PC index

The PC index measures "overlap" between clusters, which is defined in accordance with the relationship 10 [49]:

$$PC = \frac{1}{N}\sum_{i=1}^{m}\sum_{k=1}^{N}\mu_{ik}^{2} \quad (10)$$

where N denotes the number of data and μ_{ik}^{2} is the membership amount of the k data point in the ith cluster.

CE index

The CE index measures the fuzzy value of the cluster segmentation, which is defined according to Equation 11 [49]:

$$CE = \frac{1}{N}\sum_{i=1}^{m}\sum_{k=1}^{N}\mu_{ik}\log_a \mu_{ik} \quad (11)$$

When different clusters are evaluated, the closer the PC index to one and the CE index to zero, the better the cluster.

Xie-Beni (XB) index

The Xie-Beni index is calculated based on Equation 12 to determine the ratio between the total amount of changes within clusters and the separation between clusters:

$$XB = \frac{\sum_{i=1}^{m}\sum_{k=1}^{N}\mu_{ik}^{\beta}\|X_k - u_i\|^2}{N.\min_{i.k}\|X_k - u_i\|^2} \quad (12)$$

X_k represents the sample under consideration and U_i represents the center of the cluster. The Xie-Beni index focuses on compaction and separation properties. The more the clusters are separated from each other, the lower the value of Xie-Beni.

SC index

The SC index is the ratio of the total density to the separation of clusters from each other and is defined according to Equation 13 [49]:

$$SC = \sum_{i=1}^{m}\frac{\sum_{j=1}^{N}\mu_{ij}^{\beta}\|X_j - u_i\|^2}{N_i.\sum_{k=1}^{m}\|u_k - u_i\|^2} \quad (13)$$

The lower the SC index, the better the clustering output.

Separation index S

The separation index S, defined in Equation 14, uses a minimum separation distance value for segmentation validity [51].

$$S = \frac{\sum_{i=1}^{m}\sum_{k=1}^{N} u_{ik}^{2}\|x_k - u_i\|^2}{N.\min_{I,K}\|u_i - u_k\|^2} \quad (14)$$

The separation index S indicates a valid optimal classification; therefore, validation indicators will be useful when predicting the number of clusters is unknown. Higher values of the S index indicate better separation of clusters from each other.

4. Materials and Data Processing

4.1. Datasets and exploratory layers

Since the evolutionary conditions of copper porphyry formation are complex, it is difficult to select possible areas with high potential. To partially overcome the problem, we developed new machine learning methods (combining supervised and unsupervised methods) to create a highly accurate MPM of the Sonajil. By taking this approach, we're able to more effectively analyze and interpret complex geological data, which ultimately allows us to better identify areas that are most likely to contain copper porphyry mineralization.

Note that the initial and most crucial stage of Cu-Au MPM involves the careful preparation of input data, which includes both predictors and targets. Essentially, these predictor variables serve as the key decision-making criteria for determining promising zones. Also for selecting the most reliable variables, the unique Cu-Au mineralization system in the area was taken into account. Accordingly, it was attempted to employ 10 different evidence maps, each of them provides valuable insights into various geological, geochemical, and alteration factors and Cu-Au pathfinders in the area. By integrating these multi-source predictor variables, the input data were generated through ArcGIS**.** For this purpose, a set of exploratory features (information layers) was prepared. These raster layers are in size of 15 x 15 meters' pixels, containing 83408 data (pixels) that was used in preparing the final potential map. Figure 3 depicts the exploratory layers, which include 10 layers of information.

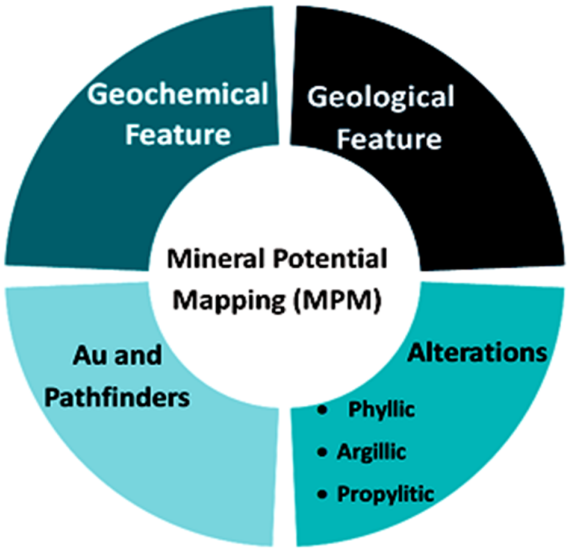

Figure 3. Exploration features in the Sonajil area for providing a potential map.

4.1.1. Multi-element geochemical layer

In order to achieve optimal results, 1248 well-organized lithogeochemical samples were collected in a grid pattern at Sonajil (Figure 1). They were analyzed by ICP(MS) for Au, Ag, As, Sb, Cu, Pb, Zn, Cr, Ni, Co, Cd, S, Sr, Hg, Ba, Be, Bi, Li, Mo, Re, Sn, W, B, Zr, Rb, Nb, U, La, Th, Te, Sc Y, Cs, and V to prepare geochemical map along with related pathfinders. Additionally, the major elements were also analyzed to detect alterations-related layers. Factor Analysis was utilized to reduce the data dimensions to the main non-dependent components using the correlation matrix to overcome the abundance of analytical data. Then the noisy elements were eliminated and purified from different factors to get clean mineralization-related pathfinders. By each step of the iteration, more cleaned mineralization FAs were obtained and non-significant FAs were eliminated (Table 1). At first, 12 components were justified with the most variability of the area where FA1 and FA2 components had higher variability and the others with low variability were negligible.

Ultimately, based on the obtained values from the rotational correlation matrix, 12 pertinent elements were extracted including Au, As, Sb, Ag, Bi, Cu, Mo, Pb, S, Zn, Hg, and Cd. Among them, the FA_1 shows geochemical factors characterizing the positive spatial relationship of the paragenetic elements such as sulphides associated Au (Table1). Therefore, FA_1 is ultimately selected as the most effective FA in preparation of the final multi-element geochemical map, illustrated in (Figure 6-a) and regarded as the final geochemical layer. Here we implemented another complementary method as a cluster analysis program to the Factor Analysis results in order to detect paragenetic indicators more precisely and fulfill the demands of geochemical mapping aspects. Fig-4 illustrates the discrimination of Cu-Au mineralization-related clusters.

Table 1. Shows the FA and Eigen values of the 12 extracted elements, the variance and the cumulative variance corresponding to each of the FAs. Here the FA_1 considered as significant geochemical layer in the final MPM.

Elements	As	Sb	Au	Ag	Bi	Cu	Mo	Pb	S	Zn	Hg	Cd	Eigen value	Variance	Cum.Var
FA1	0.319	0.82	0.81	0.51	0.027	0.65	0.05	0.034	-0.03	0.079	-0.12	0.52	2.38	23.6	23.6
FA2	0.59	0.177	-0.1	-0.14	0.51	0.058	0.75	0.77	0.197	0.097	0.126	0.242	1.93	16.1	39.7
FA3	-0.44	-0.05	-0.09	0.062	-0.28	0.149	0.07	0.121	-0.62	0.782	-0.06	0.46	1.53	12.8	52.5
FA4	-0.07	-0.05	0.055	0.668	-0.13	0.051	0.134	0.016	0.026	-0.02	0.85	0.175	1.24	10.5	63

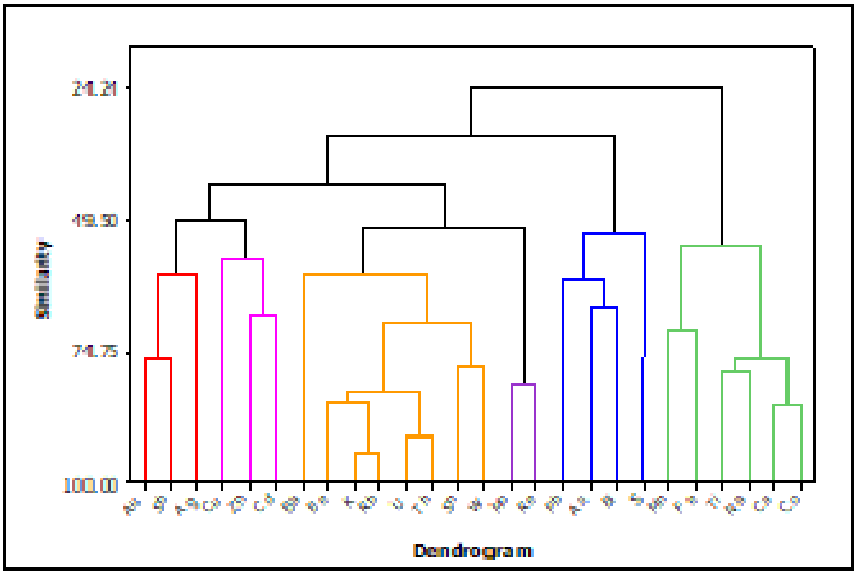

Figure 4. Clustering diagram showing four distinct sub-related paragenetic associations (Ore-related elements are in accordance with the factor analysis trend.)

Cluster-1 is a confirmation of Cu-Au mineralization along with important indicators such as Sb, Ag, Zn, and Cd. This shows that at least two phases of mineralization occurred in the area. Cluster-2 includes Be, K, Rb, U, Th, Sn, and W, which indicate the intrusion of acidic solutions along with relevant elements.

Cluster-3 shows the entry of sulphides solution to the area, along with some volatile elements such as As and related paragenesis such as Pb and Bi. Two other isomorphic elements, Mo and Re, are also accompanying the 3rd cluster, but their effects are less due to their depletion in the Sonajil compared with the Sungun deposit. The 4th cluster includes syngenetic elements of rock forming minerals such as Fe, Mn, Ti, Na, Ca, and Co.

Combining the results of factor analysis and clustering diagrams (Table 2 and Figure 4), the Cu-Au paragenesis were identified as (Au, Sb, Cu, Ag, and Cd), which were used as the exploratory layer to improve MPM results. Figure 6 (f-g-h-k-m) shows the raster layers prepared for the mentioned elements.

4.1.2. Hydrothermal alteration layer

Hydrothermal alteration as a wall rock metasomatism is a well-known epigenetic process that encompasses the deposit, formed due to the

reactions between highly heated aqueous fluids and the country rocks as it flows through the suitable conduits expanding in the form of aconcentric cone from ore roots at depth, and responsible for the formation of ore shoots in a structurally controlled susceptible area. It is characterized by geochemical signatures, and pronounced anomalous zones are consistently associated with enhanced alteration halos.

This paper reviews the general aspects of hydrothermal alteration processes and their detection in intrusion-related porphyry of the Sonajil, using the objectives of remote sensing methods since they contain high concentrations of base metals. The method is useful for identifying hydrothermal channels, which in turn may lead to the identification of similar mineralized zones in continuation of the same path [52, 53]. In this regard, Aster images were successfully employed to detect hydrothermal alteration. Accordingly, a set of Aster data in the ENVI software environment was utilized to prepare phyllic, argillic and propylitic alternation maps based on the band ratio method. Initially, geometric and radiometric corrections were performed and the vegetation canopy of the area was removed. Using the standard reflection spectrum, the index minerals of three alterations were detected by band ratio of 7/5 [54] for argillic, (7+5)/6 [54] for phyllic alteration, and (9 + 7)/8 for propylitic alteration. Finally, three relevant continuous maps of argillic, phyllic and propylitic alterations were prepared (Figures 6-c, d, and e). The alteration assemblages of these zones are dominated mainly by kaolinite corresponding to relatively weakly altered parts of the hydrothermal alteration system denoting argillic zone near the surface as a blanket overlying Cu–Au mineralization in the area, whereas sericite and quartz signify phyllic zones preferentially toward the core and the most widespread altered parts of the mineralization. Propylitic alteration as a peripheral zone exists with prevalent chloritization on the surface. Therefore, this map was potentially used as an important layer of vector to Cu–Au mineralization in the process.

4.1.3. Geological layer

Many porphyry copper deposits are genetically associated with intermediate calc-alkaline to felsic rocks above active subduction zones of the volcanic arc in the northwest of Iran. Their lithological units consist of porphyritic andesite, mega porphyritic andesite, granitoid, andesite to Hornblende andesite [55]. Similar lithologies prevail in the Sonajil where porphyritic andesite is pronounced with phyllic alteration. In this study, using the geological map of the area, the integration of several lithologies was performed and the importance of each lithology was determined and scored based on the development trend of Cu mineralization as well as expert opinions about the geological units of the region. As a result, 6 classes were defined for the geological layers and given in Figure 5 with their relevant scores in (Figure 6-b).

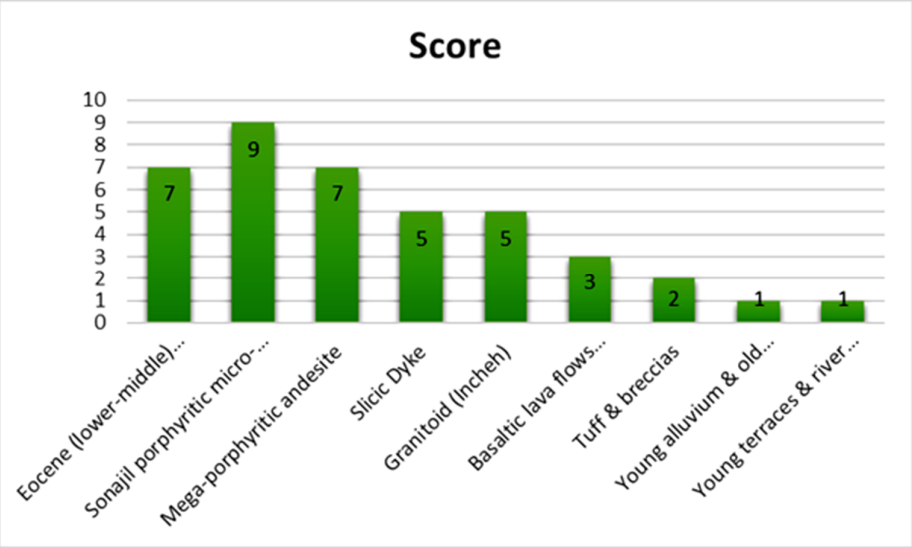

Figure 5. Sonajil lithologies and their relevant scores.

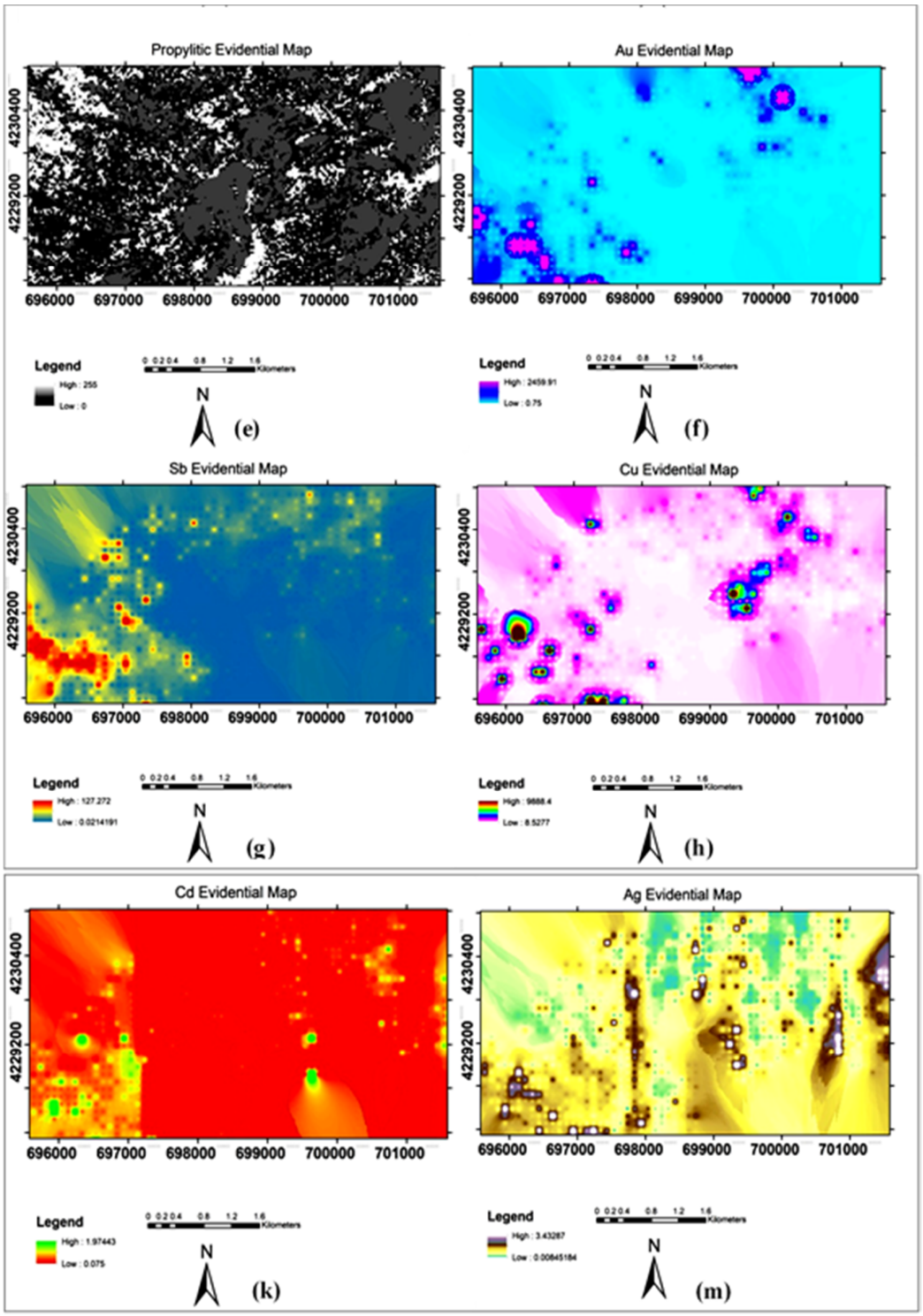

Figure 6. Continuous raster maps for exploratory evidence layers: a) geochemistry, b) geological raster map, c) argillic alteration, d) phyllic alteration, e) propylitic alteration, f) Au raster map, g) Sb raster map, h) Cu raster map, k) Cd raster map, m) Ag raster map.

4.2. Selection of training data

After obtaining general data from 10 components (exploratory feature), the information of boreholes drilled in the area (Figure 7) was used to select training data. The number of boreholes drilled in the area is a total of 27 boreholes that have been drilled in three stages, listed in Table 2, along with their acronyms. In order to select training data for

the Bayesian classifier, the boreholes were divided into mineralized and non- mineralized categories (Table 3). Subsequent to classifying eight boreholes into the four classes C1 to C4, two from each class were selected, and a 100-meter buffer was performed around them to determine the training data. Considering that intelligent methods have the ability to identify behavioral patterns among the elements of the deposit under study, therefore, the simultaneous study of behavioral patterns of borehole and surface samples can help to identify behavioral patterns between elements. By identifying the pattern that governs the area using surface data, it helps to identify the presence or absence of blind mineralization in areas that lack exploratory boreholes.

Figure 8 depicts the concentration diagrams for selected boreholes from classes C1 to C4 and Figure 9 illustrates the location of these boreholes on the geological map.

Table 2. Number of boreholes and their abbreviations, drilled in the area.

Number	Boreholes
11	Old boreholes (SOL)
13	Porphyry-oxide boreholes (PH-OX)
3	Epithermal boreholes (EP)

Table 3. Borehole classification based on variation in copper concentration graph.

Abbreviation of boreholes for each class	Number of boreholes	Assay (ppm)	Borehole class
SOL6 -SOL9-SOL10 -SOL15-SOL16-SOL17	6	C < 1000	C1
SOL12 -OX6-OX7-SOL8-OX5	5	1000 < C < 2000	C2
SOL7-SOL14-OX1-OX8-Ph2-EP1-EP4-OX4-Ph1	9	2000 < C < 5000	C3
OX3 -Ph5-Ph6-SOL11-EP5-OX2-Ph3	7	5000 < C	C4

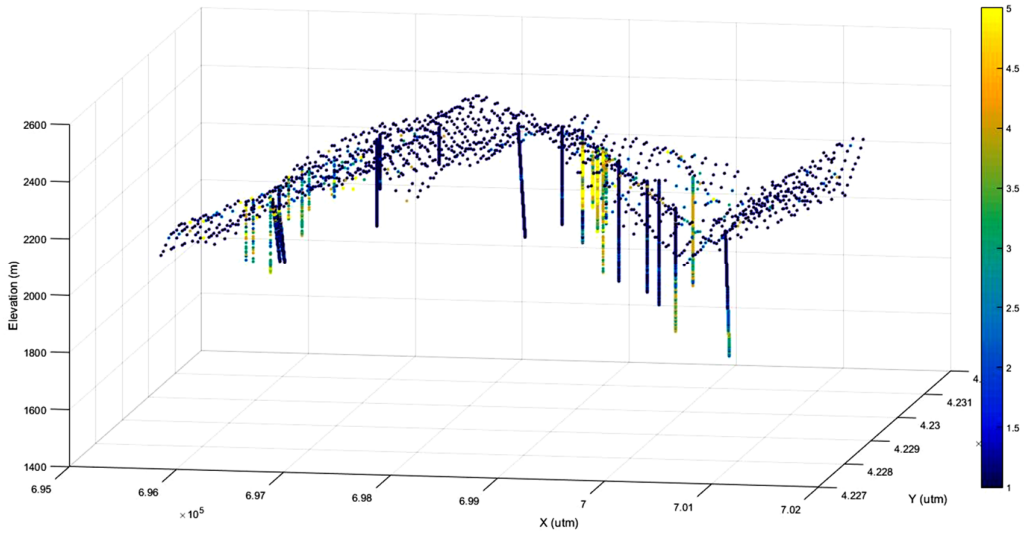

Figure 7. Location of all boreholes in 3D view with respect to surface sampling data.

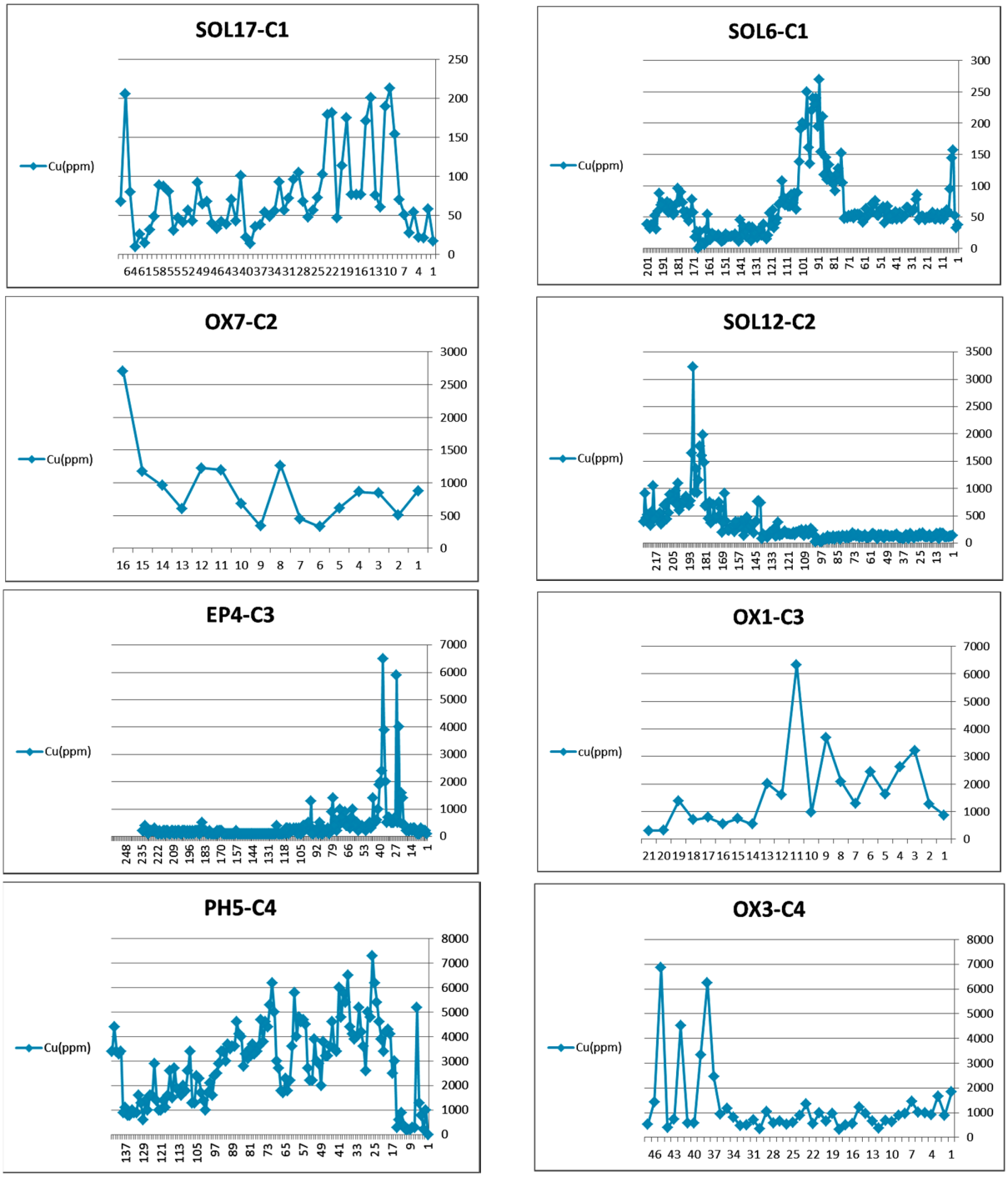

Figure 8. Concentration diagrams of SOL6 and SOL17 boreholes selected from class C1 boreholes - Concentration diagrams of SOL12 and OX7 boreholes selected from class C2 boreholes - Concentration diagrams of OX1 and EP4 boreholes selected from class C3 boreholes - Concentration diagrams of OX3 and PH5 boreholes selected from class C4 boreholes.

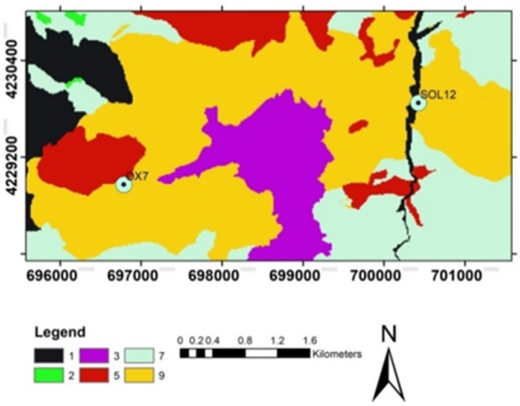

Figure 9. a) SOL12 and OX7 borehole position with a 100 m radius bore among C2 boreholes

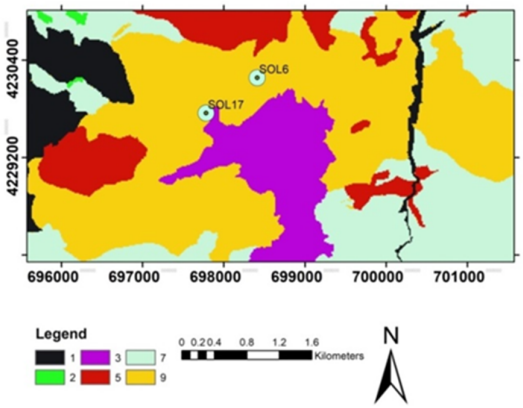

Figure 9. b) Position of SOL6 and SOL17 boreholes with a 100 m radius buffering among class C1 boreholes.

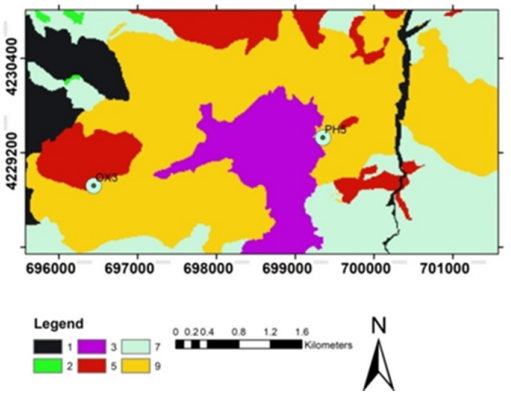

Figure 9. c) Position of PH5 and OX3 boreholes with a buffering radius of 100 meters among Class C4 boreholes.

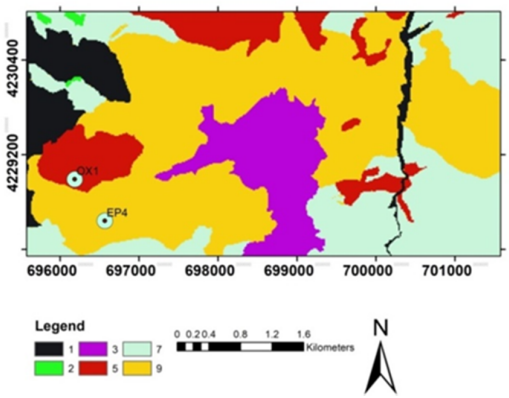

Figure 9. d) EP4 and OX1 borehole position with a radius borehole of 100 meters among Class C3 boreholes.

5. Results and Discussion

5.1. Model prediction and mineral mapping

In porphyry copper deposits, due to the nature of the hydrothermal fluid and the physicochemical conditions governing the region, various zones with different concentrations are created where each zone is related to a specific part of the deposit. Therefore, by clustering the samples, data that have common characteristics are collected in a cluster (zone). This leads to a better interpretation and more accurate modeling of the potential map. In exploratory feature datasets, there is likely to be a series of data that confuses the performance of the entire set in exploratory analyzes. By dividing the data into separate clusters, the best clusters can be identified and used in the analysis. Before clustering, the number of optimal clusters for which validation indicators were used should be determined. Then the FCM clustering algorithm was applied and the members of each cluster were identified. Thereupon the data of each cluster was divided into 70% of training and 30% of the testing, and then the Bayesian classifier was designed for preparing a potential map based on the accuracy and error of the

Classifier. Figure 10 shows the validation indicators for training consisting of 1113 data, depicted in Table 4 along with the data of each cluster. According to the determinant indices for the number of optimal clusters, in all indices after 5^{th} one, the values of the indices tend to be relatively constant. Therefore, the optimal number of clusters for the FCM algorithm was considered to be five clusters. At this stage, all the training and then the clustered data are considered as the input of the Bayesian classifier to generate the potential map. Figure 11 depicts the stepwise performance of a single Bayesian and hybrid FCM-NB classifiers to prepare an optimal MPM using all geo-data sets of the Sonajil area.

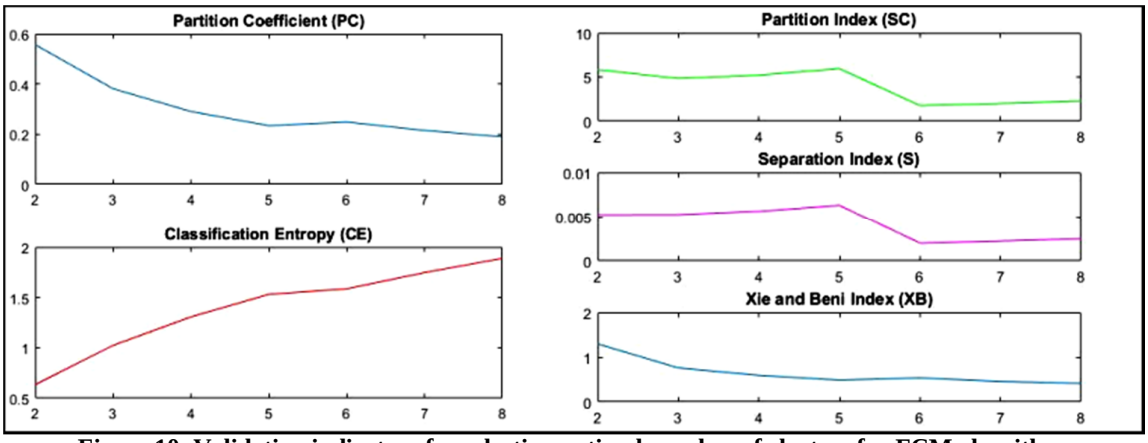

Figure 10. Validation indicators for selecting optimal number of clusters for FCM algorithm.

Table 4. Data clustering results using the FCM clustering algorithm.

Clustering algorithm	Total data	Data number in the first cluster	Data number in the second cluster	Data number in the third cluster	Data number in the fourth cluster	Data number in the fifth cluster
FCM	1113	386	298	122	37	270

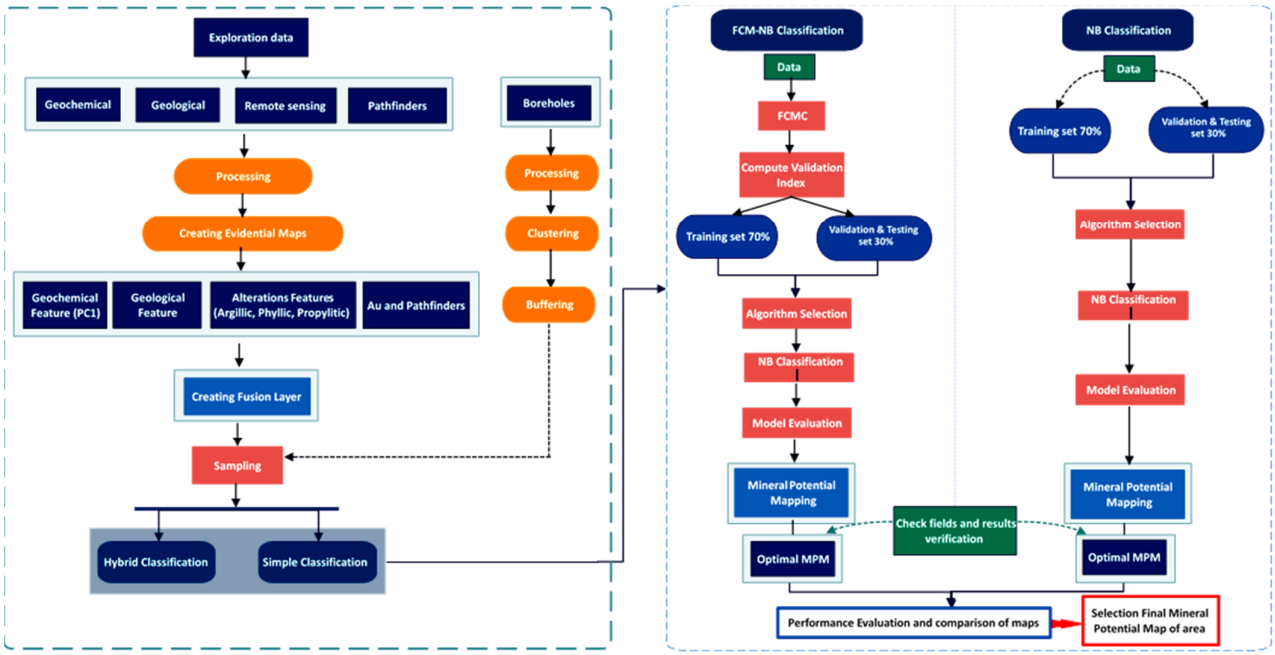

Figure 11. Stepwise preparation of optimal MPM based on FCM-NB in the Sonajil area.

In order to produce a robust prediction method for the exploration of porphyry copper in the region, we used a combination of clustering and Bayesian methods. Thus, first, the information layers including 10 exploratory features were obtained. In preparing the exploration features, the pixel size was 15 × 15 m therefore; the number of pixels for preparing the decision matrix and continuing the work was 83,408 pixels. To enhance the performance of the NB classifier method, training data were clustered with the FCM clustering algorithm and compared to the results obtained with the entire data set. Since the clustering algorithm can divide the data into more internally correlated clusters thereby the results are expected to be more accurate than in the case of using the whole dataset.

The classifier results were evaluated in different clustering modes, in the first case cluster 1 was excluded and the results were evaluated with four clusters of 2, 3, 4, and 5. In the second instance, cluster 2 was excluded and the results were

evaluated with four clusters 1, 3, 4, and 5. In the third case, cluster 3 was eliminated and the results were evaluated with four clusters 1, 2, 4, and 5. In the fourth case, cluster 4 was discarded and the results were evaluated using clusters 1, 2, 3, and 5. Finally, in the fifth case, cluster 5 was omitted and the results were evaluated with 4 clusters 1, 2, 3, and 4. In fact the confusion matrix shows the lowest error of (3.9) in the mode of eliminating the fifth cluster and using the other four with an increase in accuracy to 96.1, indeed about 96% of all samples are correctly classified into classes. Overall results of the clustering indicate the accurate and better performance of this model in determining the promising zones of mineralization compared to the mode of using whole data. Also considering that the training data used in the semi-supervised Bayesian algorithm is obtained based on the relationship between surface data and boreholes, it is possible to identify the pattern governing the region using surface data to determine the presence or absence of hidden deposits in areas that lack exploratory boreholes.

5.1. (FCM-NB) model assessments

Figures 12 to 17 show the potential map of the Sonajil in different clustering modes along with matrix confusion. In the selection of training data, boreholes were classified into four classes, C1 (low probability mineralization) and C4 (high probability mineralization), which were used according to the Cu concentration; therefore, in the process of preparing the potential map based on the Bayesian algorithm, class C4 represents the high probability mineralization class that confirms each other in different clusters. This paper presents a promising idea for potential mapping in unknown regions by addressing the issue of a semi-supervised approach. Table 3 outlines the classification of the boreholes used for cross-validation, which were not included in the algorithm training. As shown in Figure 18, most of the C4 class assumptions are located in the high probability mineralization class, indicating an acceptable accuracy of the hybrid supervised and unsupervised FCM method.

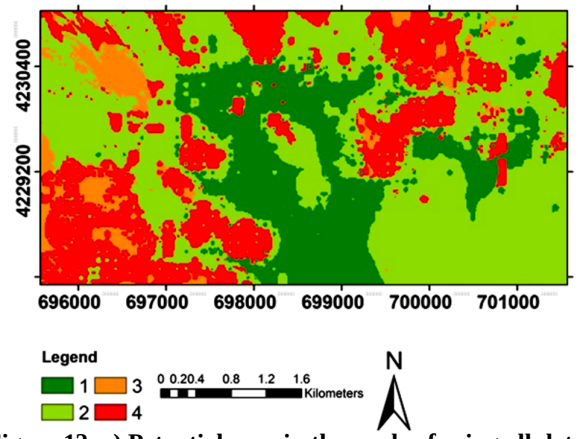

Figure 12. a) Potential map in the mode of using all data.

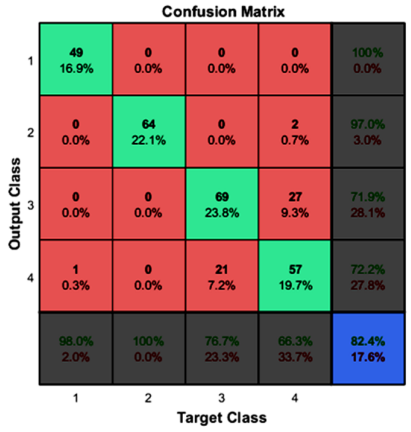

Figure 12. b) Confusion matrix with 82% accuracy.

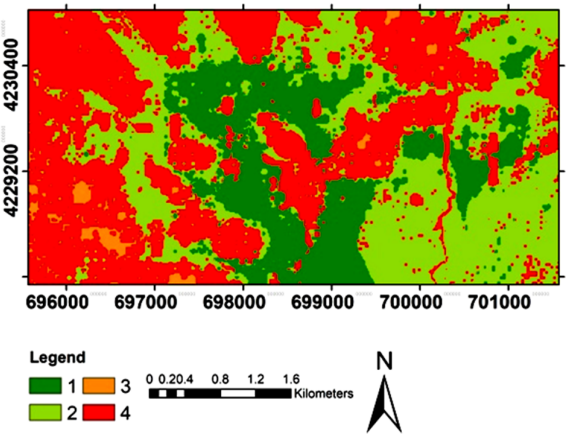

Figure 13. a) Potential map in the first cluster deletion mode.

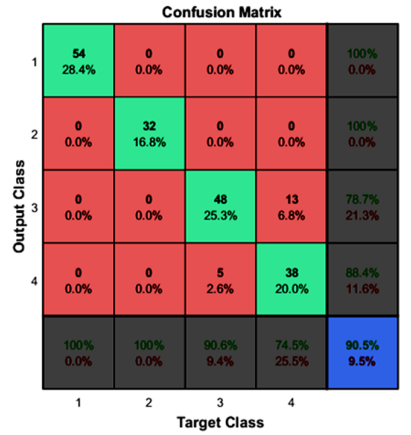

Figure 13. b) Confusion matrix with 90% accuracy.

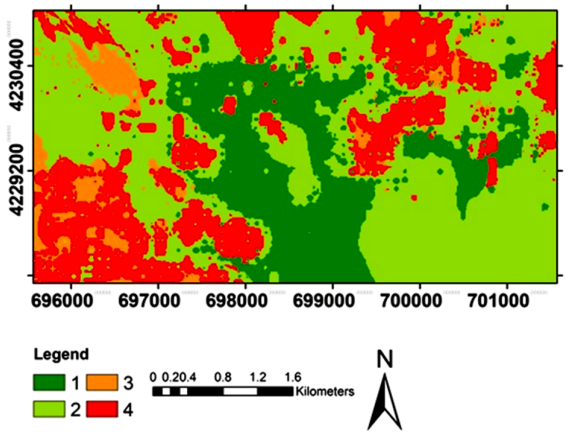

Figure 14. a) Potential map in the second cluster deletion mode.

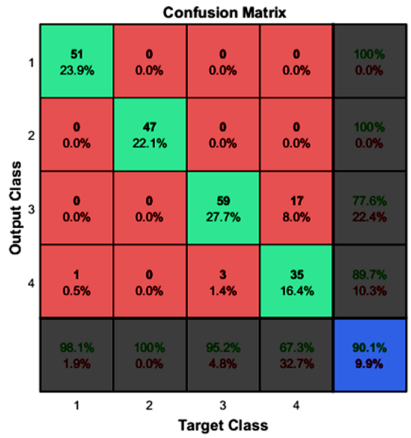

Figure 14. b) Confusion matrix with 90% accuracy.

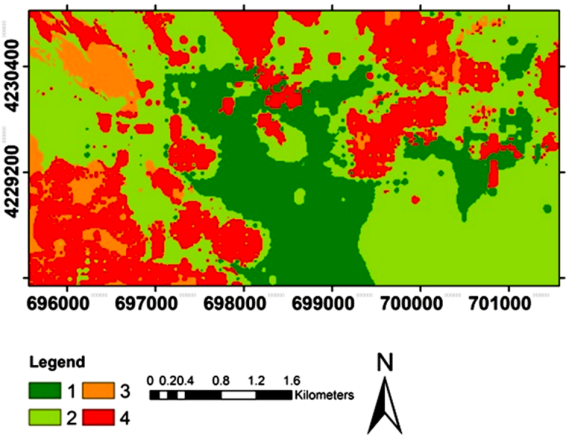

Figure 15. a) Potential map in the third cluster deletion mode.

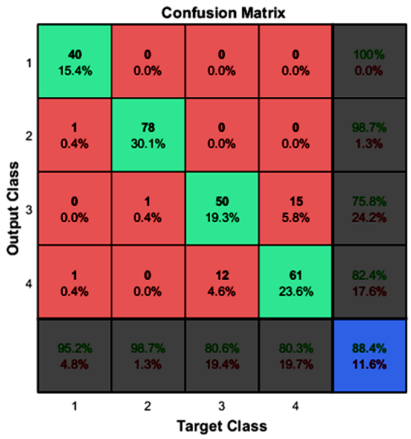

Figure 15. b) Confusion matrix with 88% accuracy.

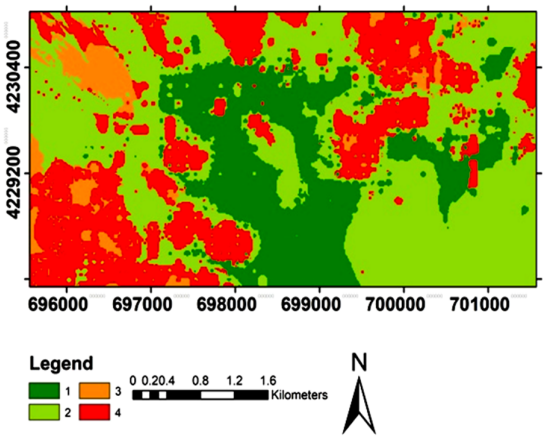

Figure 16. a) Potential map in the fourth cluster deletion mode.

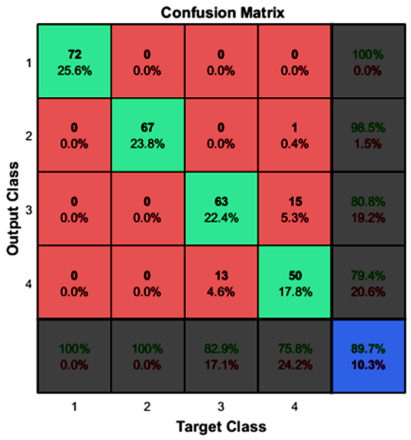

Figure 16. b) Confusion matrix with 89% accuracy.

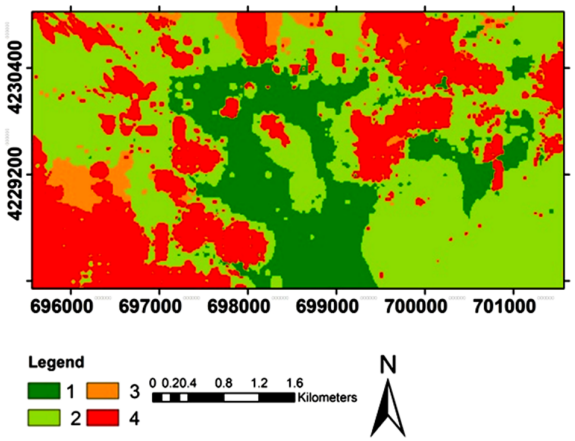

Figure 17. a) Potential map in the fifth cluster deletion mode.

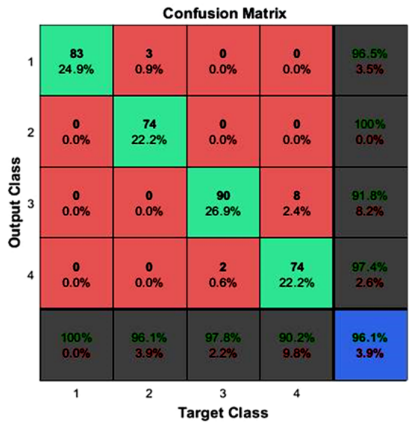

Figure 17. b) Confusion matrix with 96% accuracy.

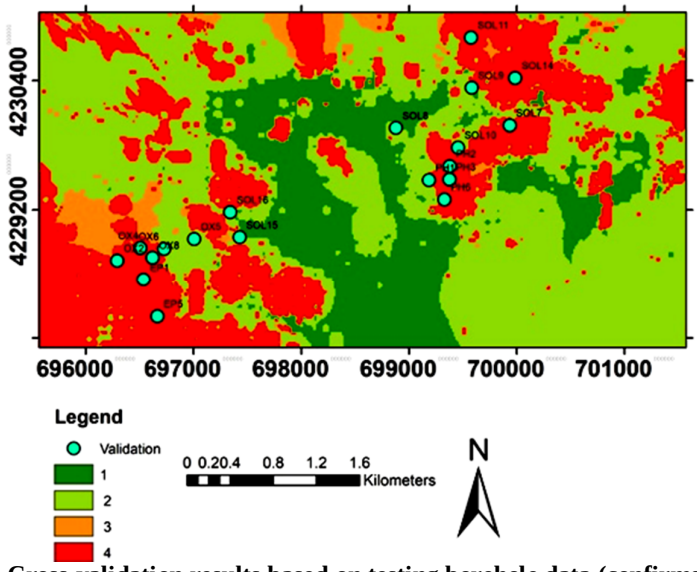

Figure 18. Cross-validation results based on testing borehole data (confirms Table 4).

6. Conclusions

Traditional exploration methods are labor intensive and have low success rates. Machine learning models can potentially improve the discovery process. This research work demonstrates that the application of machine learning techniques to mineral prospectivity mapping holds tremendous promise for resource exploration, seeking to identify new ore deposits in an efficient and data-driven manner. By analyzing and identifying patterns in large geoscientific datasets, algorithms can help discover subtle indicators that may indicate mineralization potential in the study area. The Bayesian semi-supervised approach presented in this study demonstrates how incorporating both labeled and unlabeled data can improve the prediction performance compared to traditional supervised learning models.

The preliminary results derived from similarity clustering exhibit 4 distinct sub-related paragenetic associations where ore-related elements are in accordance with the Factor Analysis trend. As a result, the Cu-Au paragenesis were identified as (Au, Sb, Cu, Ag, and Cd), which were used as the exploratory layer to improve MPM results. Depletion of Mo in comparison with the Sungun deposit and its non-accompanying with base metals in the area is evidenced. In general, the distribution of Pb, As and S is affected by the Fe pattern rather than by Cu and Zn. The prominent pathfinder for gold is Sb followed by Ag, Cu, Zn, and Cd.

The utilization of the FCM clustering method further enhanced the efficiency of the algorithm by organizing the data into distinct clusters. This clustering approach, coupled with the Bayesian classifier method, demonstrated a remarkable increase in the average accuracy of the mineral prospectivity map. Particularly, the exclusion of

cluster 5 data yielded an impressive accuracy rate of 96%. The validation of the Bayesian semi-supervised method using testing boreholes data adds credibility to the findings, highlighting its robustness and effectiveness in identifying potential mineralization targets.

This research marks an important step forward in the quest to harness the power of machine learning to boost copper and gold discovery. As datasets grow ever larger and computational capabilities increase, these techniques will only become more refined and capable of providing valuable insights to guide future exploration efforts. With better prospectivity maps identifying high priority targets, experts can focus their on-ground exploration where it matters most. The results illustrate the value of an interdisciplinary fusion of geoscience and machine learning, opening up possibilities for further optimization and novel applications that push the boundaries of our knowledge about mineralization patterns in region.

Finally, the research work concluded that:

- The Bayesian semi-supervised approach leverages both labeled data (known deposit locations) and unlabeled data (the wider geological dataset). This helps train the model more comprehensively compared to using labeled data alone. The utilization of the Bayesian semi-supervised method has an optimal outcome in the preparation of potential map in the region with its simplicity in data handling and accuracy in processing, which ultimately enable the detection of hidden ores in areas without exploratory boreholes.
- The improved performance shows the value of incorporating all available data, not just known deposits, to build a more accurate prospectivity map. This indicates there is likely useful information in the wider dataset that traditional approaches miss.
- As machine learning and computational power advance, these techniques will continue to improve and allow for more complex models that can uncover ever subtler indicators of mineralization.

Acknowledgments

The authors would like to thank the Azerbaijan Gold-Copper Processing Company for providing exploratory information about the Sonajil deposit. The immense cooperation received in the field from chief engineers Gholami and Habibi is highly appreciated.

References

[1]. Ghezelbash, R., Maghsoudi, A., & Carranza, E. J. M. (2019). An improved data-driven multiple criteria decision-making procedure for spatial modeling of mineral prospectivity: adaption of prediction–area plot and logistic functions. *Natural Resources Research*, *28*, 1299-1316.R.

[2]. Sun, G., Zeng, Q., & Zhou, J. X. (2022). Machine learning coupled with mineral geochemistry reveals the origin of ore deposits. *Ore Geology Reviews*, *142*, 104753.

[3]. Liu, Y., Cheng, Q., & Zhou, K. (2019). New insights into element distribution patterns in geochemistry: a perspective from fractal density. *Natural Resources Research*, *28*, 5-29.

[4]. Jahangiri, M., Ghavami Riabi, S. R., & Tokhmechi, B. (2018). Estimation of geochemical elements using a hybrid neural network-Gustafson-Kessel algorithm. *Journal of Mining and Environment*, *9*(2), 499-511.

[5]. Zekri, H., Cohen, D. R., Mokhtari, A. R., & Esmaeili, A. (2019). Geochemical prospectivity mapping through a feature extraction–selection classification scheme. *Natural Resources Research*, *28*, 849-865.

[6]. Zuo, R., Kreuzer, O. P., Wang, J., Xiong, Y., Zhang, Z., & Wang, Z. (2021). Uncertainties in GIS-based mineral prospectivity mapping: Key types, potential impacts and possible solutions. *Natural Resources Research*, *30*, 3059-3079.

[7]. Vemuri, V. K. (2020). The Hundred-Page Machine Learning Book: by Andriy Burkov, Quebec City, Canada, 2019, 160 pp., 49.99(Hardcover); 29.00 (paperback); 25.43(KindleEdition),(Alternatively,canpurchaseatleanpub.comataminimumpriceof 20.00), ISBN 978-1999579517.

[8]. Nykänen, V., Groves, D. I., Ojala, V. J., & Gardoll, S. J. (2008). Combined conceptual/empirical prospectivity mapping for orogenic gold in the northern Fennoscandian Shield, Finland. *Australian Journal of Earth Sciences*, *55*(1), 39-59.

[9]. Zhang, Z., Zuo, R., & Xiong, Y. (2016). A comparative study of fuzzy weights of evidence and random forests for mapping mineral prospectivity for skarn-type Fe deposits in the southwestern Fujian metallogenic belt, China. *Science China Earth Sciences*, *59*, 556-572.

[10]. Cheng, Q. (2015). BoostWofE: A new sequential weights of evidence model reducing the effect of conditional dependency. *Mathematical Geosciences*, *47*(5), 591-621.

[11]. Ghezelbash, R., Maghsoudi, A., Bigdeli, A., & Carranza, E. J. M. (2021). Regional-scale mineral prospectivity mapping: Support vector machines and an

improved data-driven multi-criteria decision-making technique. *Natural Resources Research*, *30*, 1977-2005.

[12]. Tao, J., Zhang, N., Chang, J., Chen, L., Zhang, H., & Chi, Y. (2022). Unlabeled sample selection for mineral prospectivity mapping by semi-supervised support vector machine. *Natural Resources Research*, *31*(5), 2247-2269.

[13]. Carranza, E. J. M., & Laborte, A. G. (2016). Data-driven predictive modeling of mineral prospectivity using random forests: A case study in Catanduanes Island (Philippines). *Natural Resources Research*, *25*, 35-50.

[14]. Gao, Y., Zhang, Z., Xiong, Y., & Zuo, R. (2016). Mapping mineral prospectivity for Cu polymetallic mineralization in southwest Fujian Province, China. *Ore Geology Reviews*, *75*, 16-28.

[15]. McKay, G., & Harris, J. R. (2016). Comparison of the data-driven random forests model and a knowledge-driven method for mineral prospectivity mapping: A case study for gold deposits around the Huritz Group and Nueltin Suite, Nunavut, Canada. *Natural Resources Research*, *25*(2), 125-143.

[16]. Zhang, D., Ren, N., & Hou, X. (2018). An improved logistic regression model based on a spatially weighted technique (ILRBSWT v1. 0) and its application to mineral prospectivity mapping. *Geoscientific Model Development*, *11*(6), 2525-2539.

[17]. Zhang, Z. J., Cheng, Q. M., Yang, J., & Hu, X. L. (2018). Characterization and origin of granites from the Luoyang Fe deposit, southwestern Fujian Province, South China. *Journal of Geochemical Exploration*, *184*, 119-135.

[18]. Mohammadzadeh, M., & Nasseri, A. (2018). Geochemical modeling of orogenic gold deposit using PCANN hybrid method in the Alut, Kurdistan province, Iran. *Journal of African Earth Sciences*, *139*, 173-183.

[19]. Chen, G., Huang, N., Wu, G., Luo, L., Wang, D., & Cheng, Q. (2022). Mineral prospectivity mapping based on wavelet neural network and Monte Carlo simulations in the Nanling W-Sn metallogenic province. *Ore Geology Reviews*, *143*, 104765.

[20]. Li, S., Chen, J., Liu, C., & Wang, Y. (2021). Mineral prospectivity prediction via convolutional neural networks based on geological big data. *Journal of Earth Science*, *32*, 327-347.

[21]. Porwal, A., Carranza, E. J. M., & Hale, M. (2006). Bayesian network classifiers for mineral potential mapping. *Computers & Geosciences*, *32*(1), 1-16.

[22]. Liu, Y., Cheng, Q., Xia, Q., & Wang, X. (2015). The use of evidential belief functions for mineral potential mapping in the Nanling belt, South China. *Frontiers of Earth Science*, *9*, 342-354.

[23]. Zaidi, F. K., Nazzal, Y., Ahmed, I., Naeem, M., & Jafri, M. K. (2015). Identification of potential artificial groundwater recharge zones in Northwestern Saudi Arabia using GIS and Boolean logic. *Journal of African Earth Sciences*, *111*, 156-169.

[24]. Aryafar, A., & Roshanravan, B. (2020). Improved index overlay mineral potential modeling in brown-and green-fields exploration using geochemical, geological and remote sensing data. *Earth Science Informatics*, *13*, 1275-1291.

[25]. Carranza, E. J. M., Van Ruitenbeek, F. J. A., Hecker, C., van der Meijde, M., & van der Meer, F. D. (2008). Knowledge-guided data-driven evidential belief modeling of mineral prospectivity in Cabo de Gata, SE Spain. *International Journal of Applied Earth Observation and Geoinformation*, *10*(3), 374-387.

[26]. Abedi, M., Norouzi, G. H., & Fathianpour, N. (2013). Fuzzy outranking approach: a knowledge-driven method for mineral prospectivity mapping. *International Journal of Applied Earth Observation and Geoinformation*, *21*, 556-567.

[27]. Hosseini, S. A., & Abedi, M. (2015). Data envelopment analysis: a knowledge-driven method for mineral prospectivity mapping. *Computers & Geosciences*, *82*, 111-119.

[28]. Zuo, R. (2017). Machine learning of mineralization-related geochemical anomalies: A review of potential methods. *Natural Resources Research*, *26*, 457-464.

[29]. Müller, A. C., & Guido, S. (2016). *Introduction to machine learning with Python: a guide for data scientists*. " O'Reilly Media, Inc.".

[30]. Farhadi, S., Afzal, P., Boveiri Konari, M., Daneshvar Saein, L., & Sadeghi, B. (2022). Combination of Machine Learning Algorithms with Concentration-Area Fractal Method for Soil Geochemical Anomaly Detection in Sediment-Hosted Irankuh Pb-Zn Deposit, Central Iran. *Minerals*, *12*(6), 689.

[31]. Zhou, Q., Yin, J. Y., Liang, W. Y., Chen, D. M., Yuan, Q., Feng, B. L., ... & Wang, Y. T. (2021). Various machine learning approaches coupled with molecule simulation in the screening of natural compounds with xanthine oxidase inhibitory activity. *Food & function*, *12*(4), 1580-1589.

[32]. Rahimi, H., Abedi, M., Yousefi, M., Bahroudi, A., & Elyasi, G. R. (2021). Supervised mineral exploration targeting and the challenges with the selection of deposit and non-deposit sites thereof. *Applied Geochemistry*, *128*, 104940.

[33]. Yousefi, M., & Hronsky, J. M. (2023). Translation of the function of hydrothermal mineralization-related focused fluid flux into a mappable exploration criterion for mineral exploration targeting. *Applied Geochemistry*, *149*, 105561.

[34]. Yousefi, M., Carranza, E. J. M., Kreuzer, O. P., Nykänen, V., Hronsky, J. M., & Mihalasky, M. J. (2021). Data analysis methods for prospectivity modelling as applied to mineral exploration targeting: State-of-the-art and outlook. *Journal of Geochemical Exploration*, *229*, 106839.

[35]. Keykhay-Hosseinpoor, M., Kouhsari, A. H., Hossein Morshedy, A., & Porwal, A. (2021). Porphyry Cu-Au prospectivity modelling using semi-supervised learning algorithm in Dehsalm district, eastern Iran. *Journal of Economic Geology*, *13*(1), 193-213.

[36]. Afzal, P., Farhadi, S., Boveiri Konari, M., Shamseddin Meigooni, M., & Daneshvar Saein, L. (2022). Geochemical anomaly detection in the Irankuh District using Hybrid Machine learning technique and fractal modeling. *Geopersia*, *12*(1), 191-199.

[37]. Chen, Y., & Lu, L. (2023). The Anomaly Detector, Semi-supervised Classifier, and Supervised Classifier Based on K-Nearest Neighbors in Geochemical Anomaly Detection: A Comparative Study. *Mathematical Geosciences*, 1-23.

[38]. Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. J. O. G. R. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geology Reviews*, *71*, 804-818.

[39]. Wang, Z., & Zuo, R. (2022). Mineral prospectivity mapping using a joint singularity-based weighting method and long short-term memory network. *Computers & Geosciences*, *158*, 104974.

[40]. Seyedrahimi-Niaraq, M., Mahdiyanfar, H., & Mokhtari, A. R. (2022). Integrating principal component analysis and U-statistics for mapping polluted areas in mining districts. *Journal of Geochemical Exploration*, *234*, 106924.

[41]. Hosseinzadeh, G., Calagari, A. A., Moayyed, M., Hadj-Alilu, B., & Moazzen, M. (2010). Study of Hypogen Alteration and Copper Mineralization in Sonajil Area (East of Herris, East Azarbaidjan). *Scientific Quarterly Journal of Geosciences*, *19*(74), 3-12.

[42]. Moshefi, P., Hosseinzadeh, M. R., Moayyed, M., & Lentz, D. R. (2018). Comparative study of mineral chemistry of four biotite types as geochemical indicators of mineralized and barren intrusions in the Sungun Porphyry Cu-Mo deposit, northwestern Iran. *Ore Geology Reviews*, *97*, 1-20.

[43]. Hernández-González, J., Inza, I., & Lozano, J. A. (2013). Learning Bayesian network classifiers from label proportions. *Pattern Recognition*, *46*(12), 3425-3440.

[44]. Wu, J., Pan, S., Zhu, X., Cai, Z., Zhang, P., & Zhang, C. (2015). Self-adaptive attribute weighting for Naive Bayes classification. *Expert Systems with Applications*, *42*(3), 1487-1502.

[45]. Webb, G. I., Boughton, J. R., Zheng, F., Ting, K. M., & Salem, H. (2012). Learning by extrapolation from marginal to full-multivariate probability distributions: decreasingly naive Bayesian classification. *Machine learning*, *86*, 233-272.

[46]. Taalab, K., Corstanje, R., Zawadzka, J., Mayr, T., Whelan, M. J., Hannam, J. A., & Creamer, R. (2015). On the application of Bayesian networks in digital soil mapping. *Geoderma*, *259*, 134-148.

[47]. Webb, A. R., & Copsey, K. D. (2002). Statistical Pattern Recognition. John Wiley & Sons. *New York, USA*.

[48]. Zabihi, S. M., & Akbarzadeh-T, M. R. (2012). Generalized fuzzy C-means clustering with improved fuzzy partitions and shadowed sets. *International Scholarly Research Notices*, *2012*.

[49]. Bezdek, J. C. (2013). *Pattern recognition with fuzzy objective function algorithms*. Springer Science & Business Media.

[50]. Xie, X. L., & Beni, G. (1991, August). A new fuzzy clustering validity criterion and its application to color image segmentation. In *Proceedings of the 1991 IEEE International Symposium on Intelligent Control* (pp. 463-468). IEEE.

[51]. Bensaid, A. M., Hall, L. O., Bezdek, J. C., Clarke, L. P., Silbiger, M. L., Arrington, J. A., & Murtagh, R. F. (1996). Validity-guided (re) clustering with applications to image segmentation. *IEEE Transactions on fuzzy systems*, *4*(2), 112-123.

[52]. Salehi, T., & Tangestani, M. H. (2020). Per-pixel analysis of ASTER data for porphyry copper hydrothermal alteration mapping: a case study of NE Isfahan, Iran. *Remote Sensing Applications: Society and Environment*, *20*, 100377.

[53]. Zhao, Z. F., Zhou, J. X., Lu, Y. X., Chen, Q., Cao, X. M., He, X. H., ... & Feng, W. J. (2021). Mapping alteration minerals in the Pulang porphyry copper ore district, SW China, using ASTER and WorldView-3 data: Implications for exploration targeting. *Ore Geology Reviews*, *134*, 104171.

[54]. Pour, A. B., & Hashim, M. (2012). The application of ASTER remote sensing data to porphyry copper and epithermal gold deposits. *Ore geology reviews*, *44*, 1-9.

[55]. Aghazadeh, M., Hou, Z., Badrzadeh, Z., & Zhou, L. (2015). Temporal–spatial distribution and tectonic setting of porphyry copper deposits in Iran: constraints from zircon U–Pb and molybdenite Re–Os geochronology. *Ore geology reviews*, *70*, 385-406.

[56]. Du, X., Zhou, K., Cui, Y., Wang, J., & Zhou, S. (2021). Mapping Mineral Prospectivity Using a Hybrid Genetic Algorithm–Support Vector Machine (GA–

SVM) Model. *ISPRS International Journal of Geo-Information*, *10*(11), 766.

[57]. Chen, Y., & Wu, W. (2017). Mapping mineral prospectivity by using one-class support vector machine to identify multivariate geological anomalies from digital geological survey data. *Australian Journal of Earth Sciences*, *64*(5), 639-651.

[58]. Ding, K., Xue, L., Ran, X., Wang, J., & Yan, Q. (2022). Siamese network based prospecting prediction method: A case study from the Au deposit in the Chongli mineral concentrate area in Zhangjiakou, Hebei Province, China. *Ore Geology Reviews*, 105024.

[59]. Oonk, S., & Spijker, J. (2015). A supervised machine-learning approach towards geochemical predictive modelling in archaeology. *Journal of archaeological science*, *59*, 80-88.

[60]. Sun, G., Zeng, Q., & Zhou, J. X. (2022). Machine learning coupled with mineral geochemistry reveals the origin of ore deposits. *Ore Geology Reviews*, *142*, 104753.

[61]. Zhang, N., Zhou, K., & Li, D. (2018). Back-propagation neural network and support vector machines for gold mineral prospectivity mapping in the Hatu region, Xinjiang, China. *Earth Science Informatics*, *11*, 553-566.

رویکردهای پیش‌بینی یادگیری ماشین برای ثبت پتانسیل کانی زایی Cu-Au در منطقه سوناجیل، شمال غرب ایران: بهبود الگوی اکتشافی با الگوریتم نیمه نظارتی بیزین

محمدجعفر محمدزاده[1*]، مجید محبوبی اقدم[1]، محرم جهانگیری[2] و آینور ناصری[3]

1. دانشکده مهندسی معدن، دانشگاه صنعتی سهند، تبریز، ایران
2. دانشکده مهندسی معدن، نفت و ژئوفیزیک، دانشگاه صنعتی شاهرود، شاهرود، ایران
3. گروه مهندسی معدن دانشگاه آزاد اسلامی واحد اهر، اهر، ایران

ارسال 2023/06/30، پذیرش 2023/08/07

* نویسنده مسئول مکاتبات mohammadzadeh@sut.ac.ir

چکیده

بیشتر الگوریتم‌های تحت نظارت یادگیری ماشین که در تهیه نقشه پیش بینی پتانسیل معدنی به کار میروند به داده‌های بدون نویز جهت دستیابی به نتایج با کارایی و قابلیت اطمینان بالا نیاز دارند. همچنین به منظور کشف ساختارهای پنهان یک مجموعه داده، روش‌های خوشه‌بندی بدون نظارت کارایی بالایی دارند. بنابراین مطالعه حاضر سعی دارد با استفاده از ترکیب روش‌های نظارت شده و بدون نظارت با بکارگیری داده‌های آموزشی و تست، نقشه پتانسیل دقیقی از کانسار مس- طلا سوناجیل واقع در شمال غرب ایران تولید و ارائه دهد. در این مطالعه الگوریتم نیمه نظارت بیزین برای ثبت بهینه پتانسیل معدنی در نهشته مس – طلای سوناجیل، به کار گرفته شد. در ابتدا لایه‌های رستری 10 فیچر اکتشافی تهیه شد. سپس 27 گمانه اکتشافی حفر شده در منطقه به 4 کلاس C1 تا C4 بر اساس تغییرات غلظت مس طبقه‌بندی شدند و از هر کلاس 2 گمانه انتخاب و بافر زنی 100 متری حول این گمانه‌ها زده شد، تا 1113 داده آموزشی بر اساس الگوی رفتاری داده‌های حاصل از نمونه‌های سطحی و گمانه‌ها استخراج شود. متعاقباً داده‌های موجود به روش FCM خوشه‌بندی شده و سپس کل داده‌ها و داده‌های خوشه‌بندی به منظور بررسی دقت روش کلاسیفایر بیزین در خوشه‌های مختلف وارد الگوریتم شدند. نتایج حاکی از افزایش دقت میانگین درحالت استفاده از داده‌های خوشه‌بندی نسبت به حالت استفاده از کل داده‌ها برای تهیه نقشه پتانسیل را دارد. قابل ذکر است، الگوریتم نیمه نظارت بیزین، در حالت حذف خوشه پنجم دارای دقت قابل توجه 96 درصد شد. برای اعتبارسنجی این روش نیمه نظارتی، از داده‌های گمانه‌ای که در آموزش کلاسیفایر بیزین استفاده نشده بودند، استفاده شد که صحت نقشه پتانسیل معدنی تولید شده را تأیید می‌کند. نتایج کلی نشان می‌دهد استفاده از روش خوشه بندی FCM که به طور کارآمد داده‌ها را سازماندهی می‌کند، نقش مهمی در بهبود دقت روش نیمه نظارتی بیزین و تهیه نقشه پتانسیل برای کشف کانی سازی پنهان در مناطق فاقد گمانه‌های اکتشافی دارد.

کلمات کلیدی: نقشه پتانسیل معدنی مس- طلا، روش نیمه نظارتی بیزین، سوناجیل، آذربایجان شرقی.