Shahrood University of Technology

Journal of Mining and Environment (JME)

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

Iranian Society of Mining Engineering (IRSME)

Application of Clustering Methods to Identify High-grade Zones a Case Study: Lar Porphyry Deposit, Sistan and Baluchistan Province, Southeastern Iran

Moslem Jahantigh and Hamidreza Ramazi*

Department of Mining Engineering, Faculty of Mine, AmirKabir University, Tehran, Iran

Article Info

Received 24 May 2024

Received in Revised form 14 July 2024

Accepted 13 October 2024

Published online 13 October 20242024

DOI: 10.22044/jme.2024.14582.2747

Keywords

Clustering

SOM

K-MEANS

Drilling

Silhouette

Abstract

Various methods have been used for clustering big data. Pattern recognition methods are suitable methods for clustering these data. Due to the large volume of samples taken in the drilling of mines and their analysis for various elements, this category of geochemical data can be considered big data. Examining and evaluating drilling data in the Lar copper mine in Sistan and Baluchistan province located in the southeast of Iran requires the use of these methods. Therefore, the main goal of the article is the clustering of the drilling data in the mentioned mine and its zoning of the geochemical data. To achieve this goal, 3500 samples taken from drilling cores have been used. Elemental analysis for six elements has been done using the ICP-Ms method. Pattern recognition methods including SOM and K-MEANS have been used to evaluate the relation between these elements. The silhouette method has been used to determine and evaluate the number of clusters. Using this method, 4 clusters have been considered for the mentioned data. According to this method, it was found that the accuracy of clustering is higher in the SOM method. By considering the 4 clusters, 4 zones were identified using clustering methods. By comparing the results of the two methods and using the graphical method, it was determined that the SOM method has a better performance for clustering geochemical data in the studied area. Based on that, zones 2 and 4 were recognized as high-grade zones in this area.

1. Introduction

Mineral reserve estimation is one of the most important steps in the economic evaluation of a deposit [1,2]. Mineral grade estimation directly affects the reserve estimation results. Various factors including the physical and structural characteristics of the deposit control grade distribution The properties of metal deposits and related systems are very variable, and various models have been investigated to describe the spatial patterns of their diverse and complex structures. The grades for the different metals vary considerably but generally average less than 1%. In porphyry Cu deposits, for example, Cu grades range from 0.2% to more than 1% Cu; In porphyry Mo deposits, Mo grades range from 0.07% to nearly 0.3% Mo. In porphyry Au and Cu-Au deposits, Au grades range from 0.2 to 2 g/t Au [3]. The data used in estimating the grade mostly have a complex and non-linear model, high variance, and high skewness. Common geometric and geostatistical methods available in practice are not able to identify these relationships well. On the other hand, these methods require an expert and assumptions that cannot be realized in practice. To improve their estimation performance and accuracy, as well as to find the appropriate variability pattern governing the data, it is necessary to evaluate the performance of more complex machine learning methods.

Since the 1970s, machine learning methods have been used to find hidden information in geosciences [4]. Also, Since the 1990s, various mathematical methods have been employed in identifying geochemical anomalies, including concentration-area fractal/multifractal modeling, spatial analysis/geoinformatics, and machine learning techniques like neural and deep learning algorithms [5,6]. Clustering is one of the popular and common

Corresponding author: Ramazi@aut.ac.ir (H.R. Ramazi)

methods in machine learning science which is one of the unsupervised machine learning methods [7]. The application of clustering is to divide data into different groups. In this division, the data in each group have the most similarity, and the data in different groups have the least similarity [8]. Clustering methods have had successful applications in geochemistry. The combination of clustering methods and geochemical data can identify exploration targets. Several researches have been done in this field to determine the distribution of elements and their division into different clusters. The choice of method is very important in this field. Among these methods, the neural network method can be mentioned [9, 10]. One of the important applications of neural networks is its use in clustering. Clustering is one of the important methods of machine learning in the field of big data analysis that contains many variables and features. Among other important algorithms in the field of clustering is the K-means method. The k-means method has been widely used in clustering due to its easy application, effectiveness and experimental success. There are two important parameters in the k-means method: (1) the number of clusters and (2) the center of clusters [11]. If these parameters are unique, the answer to the K-means method is also unique. However because these parameters are not unique, the answer of the K-means method is also different. As a result, different methods have been developed to determine the number of clusters. Therefore, clustering is one of the attractive methods in geosciences, especially geochemical data analysis. Therefore, many clustering methods have been used in the analysis of geochemical data, among which we can refer to the fcm method, hierarchical clustering, K-means, and self-organized map. For example, Sarparandeh and Hezarkhani [9, 10] used the SOM method for clustering rare earth elements in the Choghart iron mine. Ghezelbash et al. [12] applied the K-means method to enhance geochemical anomalies.

SOM is structurally similar to the K-means method, but the topology is much more complicated. This method is first trained using training data and then implemented on test data [13]. The SOM method is an unsupervised machine learning method that has a great ability to visualize and cluster geochemical data. The protective feature of the SOM method has made it a popular and widely used method. The most important advantage of SOM is that it has a good visualization ability for geochemical data and mapping it. In this research, the SOM method has been used to analyze drilling data in the Lar porphyry deposit area in Sistan and Baluchistan province. In addition, the K-means method has been used to compare the results. The purpose of this research is zoning and identifying high-grade zones in the Lar Porphyry copper deposit.

2. Geology and mineralization

Copper-Molybden Lar deposit is located in Sistan and Baluchistan province, 20 kilometers north of Zahedan and near the Iran-Pakistan border, and it has longitudes of 60° 51°, 55° 60° east, and latitudes of 40˚29 and 45˚29 north (Figure 1) [14].

Lar copper-molybdenum deposit is located in the Flyshi zone or Nahbandan zone – Khash zone in terms of geological location (Figure 2a). Its eastern limit is Pakistan and Afghanistan, its western limit is the Nahbandan fault, and its southern limit is the Makran zone of the Beshagard fault. The north-south extension of this zone is about 800 km and its width is about 200 km. On the north side, some of its branches extend to the northwest and west, i.e., Basiran and Birjand, and surround the Shah Kouh granite, and in the south of Birjand, it is separated from Lut by a fault. There are no sediments older than the Cretaceous in this zone. The Upper Cretaceous facies are a type of flysch sediments mixed with volcanic rocks, the thickness of which reaches up to 3000 meters. This area is covered by the Paleocene and Eocene flysch units and they are cut by the Lar igneous complex with a northwest-southeast trend (Figure 2b) [16]. Intermediate igneous units form the northeastern part and flysch deposits form the southeastern part of the field. The internal igneous rocks in the deposit are mostly syenite to monzonite, which have been crushed under the influence of tectonic factors and affected by hydrothermal and surface solutions, altered and weathered, and show copper-molybdenum mineralization. In addition, igneous rocks such as tuff and lamprophyre, as well as a smaller amount of metamorphic rocks such as hornfels, are among the other rocks in the deposit area. Lar region has at least two structural systems (Figure 2c). The first category is the northwest-southeast fault system, which is parallel to the Zahedan fault, which is the main fault system and extends with a steep slope to the southwest, in all the intrusive masses including the mineralized area [16]. The main focus of mineralization is related to this fault system. The other group is the northeast-southwest fault system, which is younger than the faults of the first group, which cut the sheared area of Mount Lar and the thick limestones of the northern layer of Mount Lar. In some cases, they have caused the displacement of mineralized zones [16].

Mineralization of copper and molybdenum has occurred in a scattered form and veinlets and veins, both siliceous and sulfide, which follow the fracture system of the host rocks and have a trend of north-northwest-south-southeast. The host rocks show variable degrees of endogenous and exogenous changes, which due to the similarity of the fluid composition with the host rocks, have limited expansion, lack zoning, and are generally siliceous and to a lesser extent potassic, phyllic, propylitic, and argillic. Mineralization in Lar can be divided into endogenous and exogenous types, which have sulfide and oxide subcategories. Endogenous mineralization can be recognized by the presence of chalcopyrite, pyrite, bornite, molybdenite, magnetite, hematite, and anargite, and exogenous mineralization is identified by cavelite, chalcocite, dignite, limonite, malachite, and azurite, which are most widespread in the surface areas. The exogenous enrichment zone in Lar is not visible as expected. The reason for this phenomenon is the small amount of pyrite and the lack of acid production necessary for washing copper, and probably the topography and special weather conditions of the region were also influential in the small expansion of this zone. The average grades of copper, molybdenum, and gold in the Lar deposit are 0.16%, 0.01%, and 0.66 ppm, respectively, which are spread in an area with a minimum area of 0.75 km2 [17]. It is worth mentioning that only a part of this mineralization is exposed on the surface and most of it is hidden.

The fluids involved in the quartzes of the Lar copper-molybdenum deposit are very fine and often two-phase, but despite this, due to the spatial and generative correlation with the intermediate porphyroid intrusive masses, the composition of the constituent minerals, the size, and style of mineralization, the uniform distribution of veinlets and veins. Siliceous and sulfide deposits, of low grade and large extent, are placed in the category of copper-molybdenum porphyry deposits in connection with Shoshone rocks. The said deposit is different from other porphyry deposits related to Shoshone rocks in terms of the composition of the alteration agent fluid, the physicochemical conditions of the formation such as depth and high pressure, and the amount of host rock subtraction that caused the abundance of molybdenum and the reduction of gold.

It is worth mentioning that supergene minerals can be seen up to a depth of 70 meters. In general, supergene processes visible at the surface are closely related to tectonic phenomena and are mostly developed along main faults and fractures, and joints. Their lateral expansion is also due to the influence and intensity of hydrothermal fluids that affect only large fractures and crushed areas. According to field investigations and microscopic studies, the alterations in Lar copper-molybdenum deposit are siliceous, potassic, phyllic, propylitic, and argillic, the most prominent of which is siliceous alteration. In general, the usual zoning of changes in porphyry deposits is not formed in this range. In the Lar copper-molybdenum deposit, there is a hypogene zone after a depth of 70 meters.

3. Data and methodology

In this study, 3234 samples taken from the drilling cores of the Lar copper mine were used. The location of x, y, and z were recorded for these samples. Then these samples were analyzed for Cu, Mo, Ag, Fe, S, Pb, and Zn with ICP-Ms method. The location of 17 boreholes and sample location are shown in Figure 3. The statistical parameters of the main elements of the Lar Porphyry deposit are presented in Table 1.

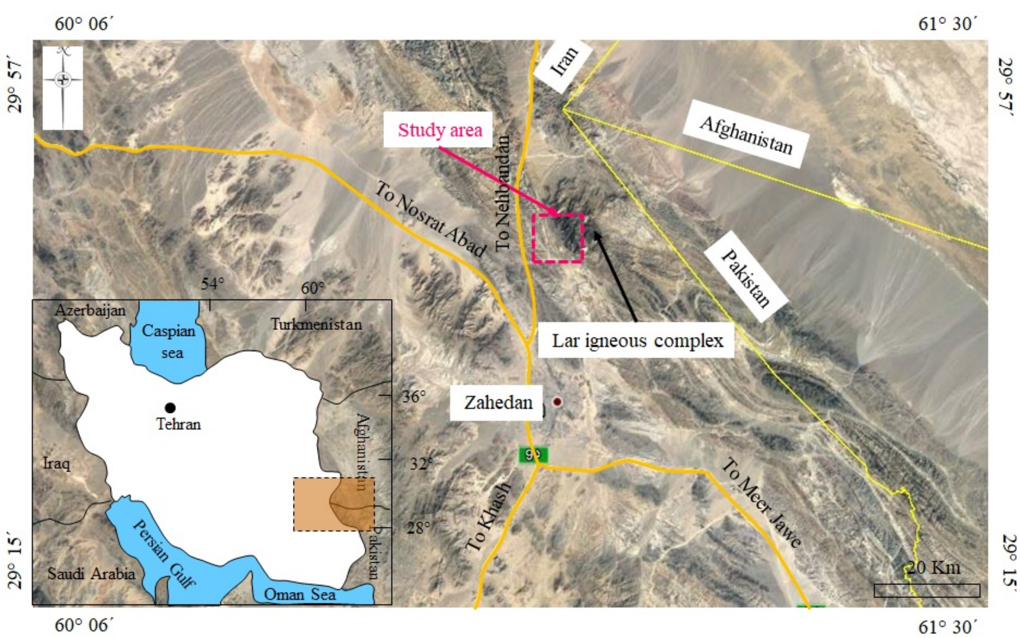

Figure 1. Geographical location and access ways to Lar copper-molybdenum deposit.

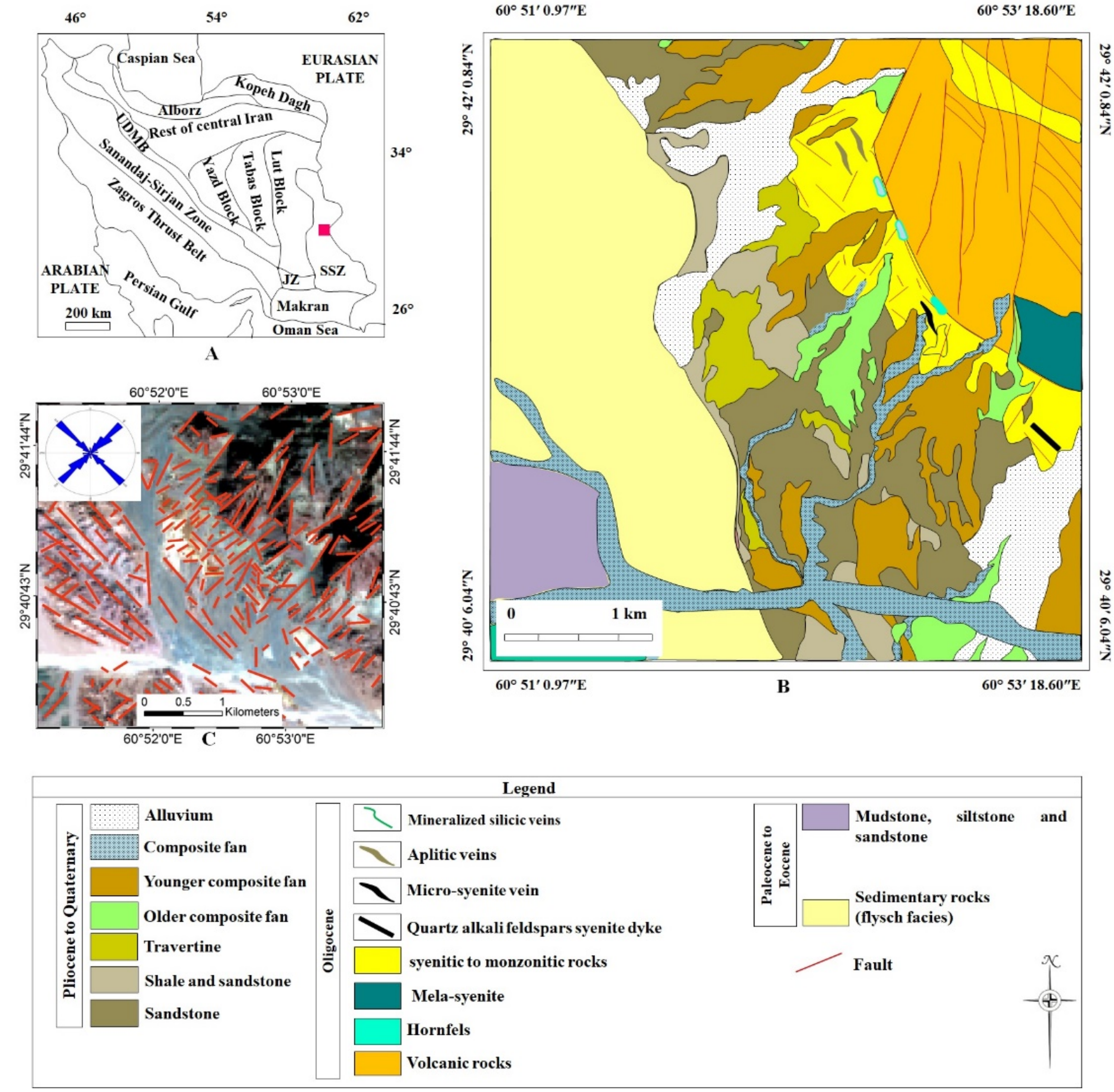

Figure 2. a) Map of sedimentary-structural zones of Iran b) Geological map of Lar copper-molybdenum deposit on a scale of 1:5000, c) Dominant structural trends in Lar copper-molybdenum deposit [13].

Table 1. Raw data statistical parameters.

Statistical Parameters	Cu	Mo	Ag	Fe	S	Pb	Zn
max	22787	3736.9	7.9	179582	24640	173	187
min	11	0.83	0.13	2454	51	3	4
std	1758.97	153.3062	0.43	7726.007	1818.4	7.80	16.23
median	630	28.1	0.25	26470	727	8	33
mean	2168.291	68.4487	0.35	45520.89	2254.8	10.33	37.09

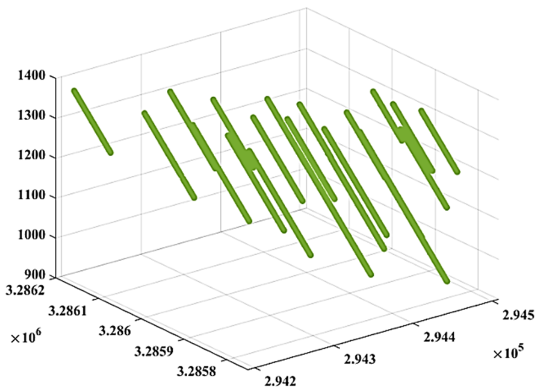

Figure 3. Location of boreholes and samples.

Data mining takes advantage of advances in artificial intelligence and statistics. Both of these fields work on pattern recognition and data classification problems. Therefore, it will be directly used in data mining. Data mining is not a substitute for previous statistical techniques, but rather their heir. It is the change and expansion of former techniques to adapt them to the volume of data and today's problems [18, 19].

Clustering is one of the branches of unsupervised learning. The application of clustering is to divide data into different groups. In this division, in-group data contain the most similarity and between-group data show the least similarity [20]. The K-means algorithm is one of the most important algorithms for clustering problems. In this algorithm, all data are divided into k clusters. The application of this algorithm is divided into parts. In the initial part, one point is considered as cluster center. It is determined which data has the shortest distance to the center of each cluster and that data is placed in that cluster. In general, in the K-means algorithm, the Euclidean distance method is used to calculate the distance from the center of the cluster. When all data have been placed in all groups, the first part of the K-means algorithm is finished. In the second part of the K-means application, the new cluster center is defined. Based on this new center, the clusters are updated and the distance from the cluster centers is calculated again. Clusters are updated again. This step continues until the clusters converge [21].

Network architecture and signal analysis used in modeling neural systems can be divided into three groups [22]. Feedforward networks convert a set of input signals into output signals. The desired input-output setting is often determined by supervised and external tuning of system parameters. In feedback networks, the input information defines the initial activity state of the system. In third-order networks, neighboring cells in a neural network compete in their activities using two-way lateral interactions [21], and according to specific identifiers, different patterns of a signal grow. This category is called competitive, unsupervised, or self-organized learning [23]. A self-organized map is a type of unsupervised artificial neural network that displays the samples under training in a separate space. This network uses an adjacency function to preserve the topological properties of the input space [22]. The Som network consists of neurons arranged in a low-dimensional, two- or three-dimensional network. The number of neurons may vary from a few tens to several tens. The algorithm is similar to the K-means algorithm, with the difference that in addition to the most suitable weight vector, its topological neighbor is also updated on the map. This means that the area of the mentioned vector in the network has a similar expansion. The final result is that the neurons in the network are regularized and the neighboring neurons get similar weight vectors [24- 26].

4. Results and discussion

The purpose of this research is to examine the data on the Lar Porphyry Copper Mine and its zoning. To achieve this goal, first, the geochemical data of this area were normalized, and then clustering methods were used for its zoning. The K-means and SOM methods were used for clustering geochemical data. The input data of these two algorithms includes 3500 samples for the grade of Cu, Mo, PB, Zn, and Ag elements. The optimum number of clusters was determined with the silhouette method. This method is based on calculating the distance between the members of each cluster and the distance between prototype clusters. The silhouette method shows the degree of similarity of each point with other points in its cluster compared to other points in other clusters. To determine the number of clusters, the number of clusters was considered from 2 to 10**.** The number of 4 clusters was considered as the optimal number of clusters. The results of K-means and SOM methods with silhouette graphical method for 4 clusters are shown in Figure 3. Positive values indicate that the clustering is done correctly, and the width of each sample shows its degree of certainty in clustering. However, in some samples, the amount of silhouette is negative. However, in some examples, the silhouette value is negative, which indicates that the clustering is not correct in these cases. However, according to the similarity of the silhouette diagram for the two K-means and SOM methods, the clustering results seem reasonable. The purpose of the SOM method is to place data with large dimensions in two-dimensional space. Therefore, in this research, the drilling data of Lar copper mine with 3500 samples and 10 characteristics including x, y, z, Cu, Mo, Fe, S, Pb, Zn, and Ag were surveyed in two-dimensional space. Based on the SOM method, the data from Lar Copper Mine were divided into four clusters. Based on this, the number of samples in each cluster for the defined zones is 799, 1066, 738, and 631 samples respectively (Figure 4). In the mentioned figure, each hexagon represents a neuron in the SOM neural network. In the present study, a two-by-two structure is used in the SOM neural network. The K-means method has also been used to cluster the data of the studied area. Four clusters were also considered for data clustering with the K-means method. The graphical method was used to evaluate the accuracy of four clusters. Based on the above two methods, four zones were defined for the Lar copper deposit. Based on the above two methods, four zones were defined for the Lar copper deposit. The average grade of each zone with SOM and K-means is presented in Tables 2 and 3. Also, the range of different elements is shown in Table 4 and Table 5. Zones 1 and 3 were defined as low-grade zones in the deposit under study. tThe average value of copper and silver in these two zones does not differ much, but there is a significant difference in the elements of lead and zinc as the elements of the deposit margin. Also, the concentration of iron and sulfur in Zone 3 is higher than in Zone 1. However, zones 2 and 4 are high-grade zones in the study area. In Zone 4, the average grade of copper, molybdenum, and lead is higher than in Zone 2. The concentration of iron, sulfur, and zinc in zone 2 is higher than in zone 4. U-matrix is a useful tool in the SOM algorithm. In this method, the line is used as a basis for comparing the distance of neurons. A lighter color reflects a closer distance between neurons and a darker color indicates a greater distance between neurons. Zones 2 and 3, which are respectively the highest and lowest grade mineralization zones, are the most distant. The shortest distance is between zones 1 and 3, which are the lowest-grade mineralization zones (Figure 5). Zone 4 in the upper part is the most valuable part of mineralization. Zone 4 in the upper part is the most precious part of mineralization, which is related to the stockwork and vein part. The location of the samples and zones in the drilling boreholes is shown in Figures 6 and 7. Figure 6 shows the location of samples and zones using the SOM method, and Figure 7 shows the location using the K-means method. Therefore, the difference between clusters based on geographic coordinates is shown in these figures.

Figure 9 shows the drilled boreholes in Lar copper mine. The numbers in this horizontal section indicate the grade of the samples. In this section, the four zones resulting from the clustering method are specified separately. As can be seen, the quality of the samples in the upper part is higher than in the deeper areas, and it confirms the results of the SOM clustering method. The grades in each borehole are randomly selected.

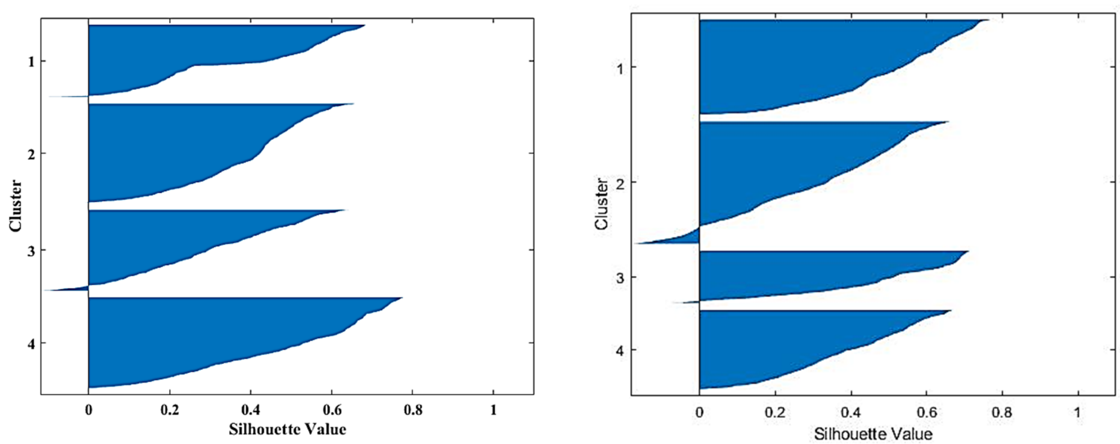

Figure 4. Silhouette plot a) number cluster with SOM b) number cluster with K-MEANS.

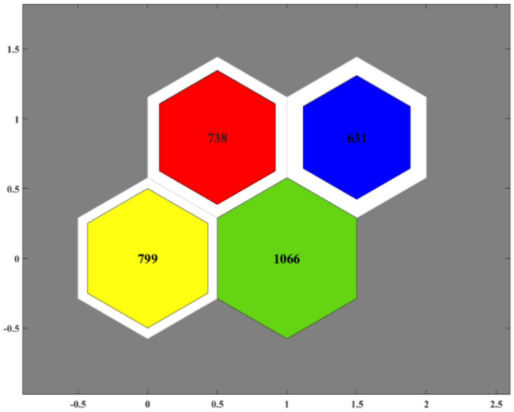

Figure 5. Determining the number of samples in each neuron by SOM method.

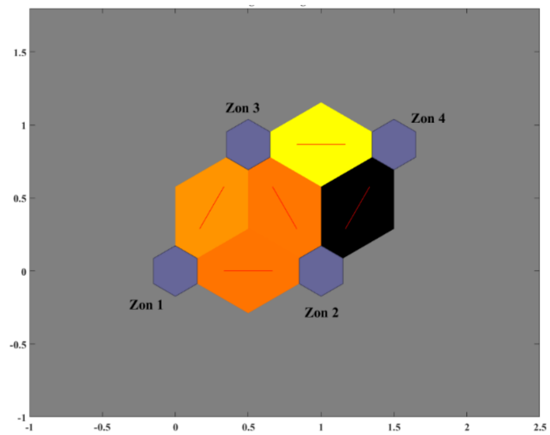

Figure 6. U-matrix for SOM method.

Table 2. Average of elements in each zone in the SOM clustering method.

	Cu	MO	Ag	Fe	S	Pb	Zn
Raw data	1204	59	0.35	27018	1224	10.27	36
Zone 1	689	41	0.26	26487	942	8.23	29
Zone 2	1822	71	0.45	27622	1330	9.74	43
Zone 3	558	50	0.24	28511	1409	10.2	36
Zone 4	1566	75	0.41	24924	1184	13.81	35

Table 3. Average of elements in each zone in the K-MEANS clustering method.

	Cu	MO	Ag	Fe	S	Pb	Zn
Raw data	1204	59	0.35	27018	1224	10.27	36
Zone 1	1171	65	0.3	26807.78	1080	12.16	39
Zone 2	524	41	0.24	26846	1085	8.43	29
Zone 3	1142	70	0.36	26908	1337	11	36
Zone 4	2067	63	0.50	27564	1455	8.74	44

Table 4. The range of different elements in different zones by the SOM method

	Cu	Mo	Ag	Fe	S	Pb	Zn
Zone 1	10759	1563.97	4.47	110857	13914	62	86
Zone 2	22769	3735.87	7.77	174936	17219	168	165
Zone 3	5585	1102.88	0.96	57257	22822	136	119
Zone 4	12423	1517.1	3.85	65842	24580	50	94

Table 5. The range of different elements in different zones by the K-MEANS method

	Cu	Mo	Ag	Fe	S	Pb	Zn
Zone 1	22746	3735.87	3.27	84290	12055	102	176
Zone 2	10761	1563.97	4.47	110857	13914	139	86
Zone 3	12425	1517.13	3.85	68896	24580	50	128
Zone 4	20394	3414.38	7.77	174936	17219	168	165

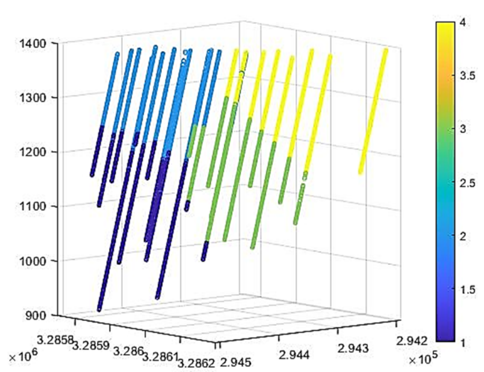

Figure 7. The Result of the SOM clustering method. The borehole data is divided into four zones.

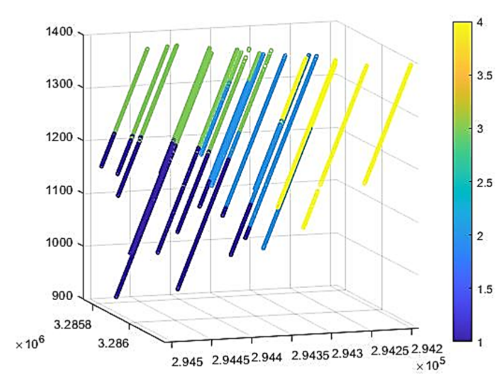

Figure 8. The result of the K-means clustering method. The borehole data is divided into four zones.

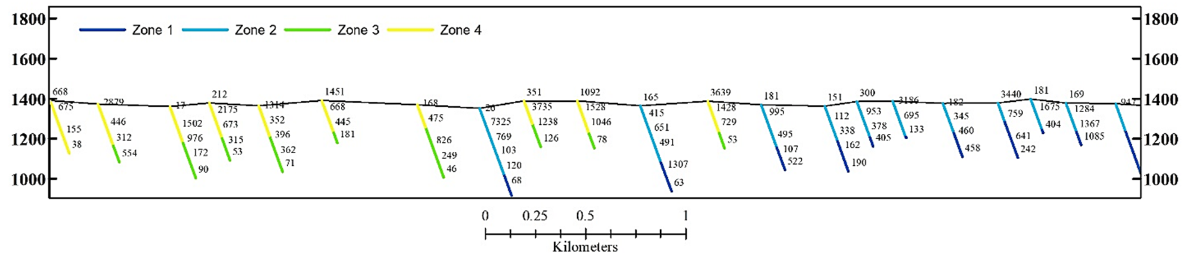

Figure 9. Cross section for validation of SOM clustering method.

5. Conclusions

Different methods have been used for clustering geochemical data, which were mentioned in the introduction. In this article, K-means and SOM methods are used for clustering drilling data in the Lar porphyry copper-molybdenum mine. SOM is an artificial neural network that transforms geochemical data from multidimensional space to 2D space. In this study, SOM clustering methods have been used for zoning the drilling data of the Lar copper mine and the results were compared with the K-means method. Using these two methods and relying on the silhouette method, the number of clusters was considered equal to 4. K-means and SOM methods show relative similarity in the number of clusters. However, the number of negative samples in clustering by the K-means method is more than the SOM method, which shows better results of SOM in clustering geochemical data in the studied area. Also, by comparing the average data related to each cluster in the two mentioned clustering methods, it is clear that the clustering operation and separation of sub-zones in the SOM method have been separated with higher accuracy. By using this method, it creates a perspective on the areas with higher grade distribution and also with lower grade distribution. This point of view can be used in additional discoveries and also in the extraction stage. According to the obtained view and considering that the high-grade zone (zone 4) of the wells has not been closed, it is possible to recommend the continuation of drilling operations in the east of this area.

References

[1]. Mostafaei K, Ramazi K (2018) 3D model construction of induced polarization and resistivity data with quantifying uncertainties using geostatistical methods and drilling (Case study: Madan Bozorg, Iran). *Journal of Mining & Environment* 4: 857-872.

[2]. Mostafaei K, Ramazi H (2019) Mineral resource estimation using a combination of drilling and IP-Rs data using statistical and cokriging methods. *Bulletin of the Mineral Research and Exploration* 160:177–195.

[3]. Sinclair, W (2007) Porphyry deposits. Mineral deposits of Canada: A synthesis of major deposit types, district metallogeny, the evolution of geological provinces, and exploration methods. *Geological Association of Canada, Mineral Deposits Division* 5: 223-243.

[4]. Takaew P, Cecilia Xia J, Doucet LS (2024) Machine learning and tectonic setting determination: Bridging the gap between Earth scientists and data scientists. *Geoscience Frontiers* 15:101726.

[5]. Farhadi, S Tatullo, MB Konari, P Afzal (2024) Evaluating StackingC and ensemble models for enhanced lithological classification in geological mapping. *Journal of Geochemical Exploration* 260, 107441

[6]. Farhadi, P Afzal, M Boveiri Konari, L Daneshvar Saein, B Sadeghi (2022) Combination of machine learning algorithms with a concentration-area fractal method for soil geochemical anomaly detection in sediment-hosted Irankuh Pb-Zn deposit, Central Iran. *Minerals* 12 (6), 689

[7]. Madhulatha TS (2012) AN OVERVIEW ON CLUSTERING METHODS. *IOSR Journal of Engineering* 02:719–725.

[8]. Templ M, Filzmoser P, Reimann C (2008) Cluster analysis applied to regional geochemical data: Problems and possibilities. *Applied Geochemistry* 23:2198–2213.

[9]. Zhang Z, Lin S, Ye Y,Xu z, Zhao H, Zhao Y and Sun J (2022) A Hybrid MCDM Model for Evaluating the Market-Oriented Business Regulatory Risk of Power Grid Enterprises Based on the Bayesian Best-*Worst Method and MARCOS Approach. Energies* (Basel) 15:2978.

[10]. Bigdeli A, Maghsoudi A, Ghezelbash R (2022) Application of self-organizing map (SOM) and K-means clustering algorithms for portraying geochemical anomaly patterns in Moalleman district, NE Iran. *Journal of Geochemical Exploration* 233:106923.

[11]. Žalik KR (2008) An efficient k′-means clustering algorithm. *Pattern Recognition Letters* | 29:1385–1391.

[12]. Sarparandeh M, Hezarkhani A (2017) Application of unsupervised pattern recognition approaches for

exploration of rare earth elements in Se-Chahun iron ore, central Iran. *Compu Computers & Geosciences ters & Geosciences* 6:537–546.

[13]. Sarparandeh M, Hezarkhani A (2016) Application of Self-Organizing Map for Exploration of REEs' Deposition. *Open Journal of Geology* 06:571–582.

[14]. Ghezelbash R, Maghsoudi A, Carranza EJM (2020) Optimization of geochemical anomaly detection using a novel genetic K-means clustering (GKMC) algorithm. *Computers & Geosciences* 134:104335.

[15]. Fu Q, Li Y, Albathan M (2023) A supervised method to enhance distance-based neural network clustering performance by discovering perfect representative neurons. *Granular Computing* 8:1051–1065.

[16]. Boomeri M, Moradi R, Stein H, Bagheri S (2019) Geology, Re-Os age, S and O isotopic composition of the Lar porphyry Cu-Mo deposit, southeast Iran. *Ore Geology Reviews* 104:477–494.

[17]. Nakisa M (2002) Results of Exploration Studies and Reserves Estimation of the Lar Cu Ore Deposit-Zahedan. Tehran, Iran: Ministry of Industries and Mines, *National Iranian Copper Industries Co*.

[18]. MORADI R, BOOMERI M, BAGHERI S, NAKASHIMA K (2016) Mineral chemistry of igneous rocks in the Lar Cu-Mo prospect, southeastern part of Iran: implications for P, T, and fO2. *turkish journal of earth sciences* 25:418–433.

[19]. Chen K, Abtahi F, Carrero J-J, Fernandez- Llatas C and Seoane F (2023) Process mining and data mining applications in the domain of chronic diseases: A systematic review. *Artificial Intelligence in Medicine* 144:102645.

[20]. Shu X, Ye Y (2023) Knowledge Discovery: Methods from data mining and machine learning. Soc Sci Res 110:102817.

[21]. Chaudhry M, Shafi I, Mahnoor M, Ramirez D. Bautista E and Ashraf I (2023) A Systematic Literature Review on Identifying Patterns Using Unsupervised Clustering Algorithms: A Data Mining Perspective. *Symmetry (Basel)* 15:1679.

[22]. Abdulnassar AA, Nair LR (2023) Performance analysis of Kmeans with modified initial centroid selection algorithms and developed Kmeans 9+ model. Measurement: *Sensors* 25:100666.

[23]. Kohonen T (1990) The self-organizing map. *Proceedings of the IEEE* 78:1464–1480.

[24]. Jensen CA, Reed RD, Marks RJ, El-Sharkawi MA, Jung JB, Miyamoto RT, Anderson G and Eggen C, (1999) Inversion of feedforward neural networks: algorithms and applications. *Proceedings of the IEEE* 87:1536–1549

[25]. Li L, Wang Y (2014) What drives the aerosol distribution in Guangdong - the most developed province in Southern China? *Scientific Reports* 4:5972.

[26]. Vesanto J (1999) SOM-based data visualization methods. *Intelligent Data Analysis* 3:111–126.

کاربرد روش های خوشه بندی برای شناسایی مناطق با عیار بالا، مطالعه موردی: کانسار پورفیری لار، استان سیستان و بلوچستان، جنوب شرق ایران

مسلم جهانتیغ و حمیدرضا رمضی*

بخش مهندسی معدن، دانشگاه امیر کبیر، تهران، ایران

ارسال ۲۰۲۴/۰۵/۲۴، پذیرش ۲۰۲۴/۱۰/۱۳

* نویسنده مسئول مکاتبات: Ramazi@aut.ac.ir

چکیده:

روش‌های مختلفی برای خوشه‌بندی کلان داده‌ها استفاده شده است. روش‌های تشخیص الگو روش‌های مناسبی برای خوشه‌بندی این داده‌ها هستند. با توجه به حجم بالای نمونه‌های برداشت شده در حفاری معادن و تجزیه و تحلیل آنها برای عناصر مختلف، این دسته از داده‌های ژئوشیمیایی را می‌توان کلان داده در نظر گرفت. بررسی و ارزیابی داده‌های حفاری در معدن مس لار در استان سیستان و بلوچستان واقع در جنوب شرق ایران مستلزم استفاده از این روش‌ها است. بنابراین، هدف اصلی مقاله، خوشه‌بندی داده‌های حفاری در معدن مذکور و پهنه‌بندی آن از داده‌های ژئوشیمیایی است. برای دستیابی به این هدف از ۳۵۰۰ نمونه برگرفته از هسته‌های حفاری استفاده شده است. تجزیه و تحلیل عنصری برای شش عنصر با استفاده از روش ICP-Ms انجام شده است. روش‌های تشخیص الگو شامل SOM و K-MEANS برای ارزیابی رابطه بین این عناصر استفاده شده است. برای تعیین و ارزیابی تعداد خوشه‌ها از روش سیلوئت استفاده شده است. با استفاده از این روش ۴ خوشه برای داده‌های مذکور در نظر گرفته شده است. با توجه به این روش مشخص شد که دقت خوشه‌بندی در روش SOM بیشتر است. با در نظر گرفتن ۴ خوشه، ۴ پهنه با استفاده از روش‌های خوشه‌بندی شناسایی شد. با مقایسه نتایج دو روش و با استفاده از روش گرافیکی مشخص شد که روش SOM عملکرد بهتری برای خوشه‌بندی داده‌های ژئوشیمیایی در منطقه مورد مطالعه دارد. بر این اساس، مناطق ۲ و ۴ به عنوان زون‌های درجه بالا در این منطقه شناخته شدند.

کلمات کلیدی: خوشه بندی، SOM، K-MEANS، حفاری، سیلوئت.