

# **Integration of Airborne Geophysics Data with Fuzzy c-means Unsupervised Machine Learning Method to Predict Geological Map, Shahr-e-Babak Study Area, Southern Iran**

Moslem Jahantigh, and Hamidreza Ramazi\*

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



## **1. Introduction**

Geophysical surveys include magnetic, radiometric, gravimetric, etc. Surveying several airborne geophysics methods at the same time supplies big data and precision information from the study area. Today geoscientists try to extract appropriate information from complex geological information. To achieve this goal, applying machine learning methods are useful tools. Geophysics data is usually interpreted separately. This paper tries to integrate geophysical data with cluster analysis and with machine learning methods. Before machine learning methods, knowledge-driven methods like WOFE, AHP, VIKOR, etc. were widely used to integrate data, perform geological mapping and identify high potential areas for more exploration [1,2]. These methods are mainly known as knowledge-based methods. Knowledge-based approaches are employed in the initial phases of exploring

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

minerals in regions with appropriate geological characteristics but limited past exploration history [3]. It is typical to combine the findings from various techniques within a linked structure to identify the most promising targets for future research initiatives [4]. These methods need to prepare a thematic map and the integrated layer should contain known deposit points. These knowledge-driven methods especially the WOFE method are not appropriate for small areas [1]. Also, these methods contain high uncertainty about 3D geological modeling [5]. ML methods, also known as machine learning approaches, have been developed and enhanced since the 1980s [6]. In recent years unsupervised machine learning methods like fuzzy means, K-means, DBSCAN, etc. have been widely used to integrate geoscience data and clarify nonlinear relation between geological structures and geophysics data [7,8].

Supervised machine learning methods require known indices to train models especially when there is big data from the study area. However, in unsupervised machine learning methods like the FCM method, there is no need for known indices to train the model. These methods extract clusters based on similarities [9]. In crisp clustering unsupervised machine learning methods, such as the K-means method, each data point is related to the closest cluster center without any hesitation. However, in fuzzy c-means method, each data was allowed to have a degree of fuzzy membership with respect to other clusters. The fuzzy membership score reflects the degree of membership of each data point belonging to a certain cluster. Different studies were done with crisp machine learning methods to integrate geophysical data to extract high favorability areas and geological units. In recent decades, the FCM method has been transformed into a popular method for quantitative evaluation of identified clusters [10,11]. Similar to crisp methods, the FCM method is widely used to integrate geophysical data for the automatic identification of geological units and mineral exploration.

There are four primary categories of ML algorithms: supervised, unsupervised, semisupervised, and reinforcement learning [12]. In this paper, the unsupervide machine learning method (i.e. FCM) has been applied to integrated airborne magnetic and airborne radiometric data to predict geological mapping in southeastern Iran, Shahr-e-Babak study area. In the first step, feature engineering has been performed on airborne geophysics data to provide input for unsupervised machine learning method. The number of cluster has been determined with Calinski- Harbaz and Davis-Bouldin methods. The geological cluster map (pseudo geological map) was created with the FCM method. In the final step, the fuzzy membership for each input layer was calculated to evaluate the relation between each input data and the pseudo geological map.

## **2. Geology of Study area**

Shahr-e-Babak is located in southern Iran. This area is part of the Urumia-Dokhtar magmatic Arc (UDMA). This area is known as the Central Iranian Volcanic Belt [13]. The studied area is located in the Alpine-Himalayan mountain range, a region that is formed due to the Neotethyan Ocean closing between Arabia and Eurasia. The UDMA is known as one of the most significant regions in the world for copper deposits. It holds three large and ten medium-sized porphyry copper, molybdenum, and gold deposits with a combined copper reserve of over 40 million tons. The major geological units in Tertiary time in the Shahr-e-Babak study area include granodiorite, diorite, tonalite, and monzonite belonging to Miocene-Oligocene time and in the shape of dyke and stocks. These igneous rocks intruded into the Eocene Razak, Oligocene Hezar, and Bahar Aseman volcano- sedimentary volcanic complex. The geological units mentioned above were covered with Mosahem stratovolcano volcanic and sub-volcanic rocks [14]. The oldest geological unit in the study area is flysch sedimentary rocks in the eastern part of the study area. The youngest geological units are quaternary units consisting of quaternary alluvial and gravel fans surrounding the volcanic belt. Figure 1 shows the location of the Shahr-e-Babak study area and Figure 2 shows the geological map of the study area.

## **3. Data and methodology**

To predict the geology map of the study area, we used airborne magnetic and radiometric data. The flow chart of the FCM method to create a geology map is shown in Figure 3. The Atomic Energy Organization of Iran (AEOI) obtained this dataset in 1977 and 1978. Magnetic data were collected from flight lines spaced 500 m apart and at an altitude of 120 m. These airborne geophysics data were surveyed north-south oriented lines spaced 1 km apart. The airborne magnetic data were corrected with an International Geomagnetic Reference Field (IGRF) and earth magnetic field was removed and magnetic anomalies were obtained. Also, to separate regional and residual anomalies, an upward continuation filter was done to obtain long-wavelength magnetic anomalies. The regional and residual components contain different anomalies originating from different depths and different magnetization. The long wavelength anomalies were subtracted from magnetic anomalies to calculate residual magnetic anomalies. Figure 4 shows the grid of airborne geophysics data consisting of (a) airborne magnetic data (b) potassium concentration and (c) thorium concentration.



**Figure 1. a) Geographical location of the Shahr-e-Babak area. b) The location of the studied area in the structural zones of Iran.**



**Figure 2. Geological map of Shahr-e-Babak area taken from Shahr-e-Babak 1:100000 geological map.**



**Figure 3. Flowchart of multigeophysical data integration.** 



**Figure 4. Airborne geophysics data of Shahr-e-Babak study area (a) magnetic data (b) potassium (c) thorium.** 

## **4. Fuzzy C-means (FCM) unsupervised clustering algorithms**

FCM is an unsupervised machine learning method which extracts clusters and structures in groups and between groups, and divides data into two or more clusters. This method is widely applied in pattern recognition. In the FCM, method the following function should be minimized [15].

$$
J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^{m} ||x_i - x_j||^2
$$

m is a real number and it should be greater than 2.  $u_{ij}$  is the degree of membership of  $x_i$  in cluster j;  $x_i$  is the d-dimensional measured data;  $c_j$  is the d-dimensional center of the cluster and  $||*||$  is any norm stating the similarity between measured data and center of the groups. Fuzzy c-means will calculate through iteration optimization of the above function. During this optimization.  $u_{i,i}$  and  $c_i$ parameters will be updated with the following equations:

$$
u_{ij} = \frac{1}{\sum_{k=1}^{1} \left( \frac{||x_i - c_j||}{||x_i - c_k||} \right)^{\frac{2}{m-1}}}
$$

$$
c = \frac{\sum_{i=1}^{N} u_{ij}^m \cdot x_i}{\sum_{i=1}^{N} u_{ij}^m}
$$

This iteration will stop when

$$
max_{ij} = \left\{ \left| u_{ij}^{(k+1)} - u_{ij}^{(k)} \right| \right\}
$$

Where  $\varepsilon$  is a termination criterion between  $0$ and 1 and k are the iteration steps.

#### **5. Feature Engineering**

To achieve the best performance of machine learning performance, the input data should be prepared appropriately. This preparing of input data is feature engineering [16]. Feature engineering consists of transforming data to a suitable form to prepare input for machine learning algorithms. It is

necessary to pre-process raw data before pttting input in machine learning algorithms. This preprocessed operation will be done with imputation, binning, outlier handling, filtering, log transformation, scaling, etc [7]. This operation is caused by better performance for machine learning algorithms. In recent research, the FCM unsupervised machine learning method was used to integrate airborne geophysics data consisting of aeromagnetic data and radiometric data (Uranium, Thorium, and Potassium layers) to predict the geological map of Shahr-e-Babak study area. The FCM method operated based on Euclidean distance and clusters tend to show spherical shape. The abnormal data with skewness and high kurtosis caused inappropriate results. As a result of the above description, the input data should be transformed to normal data without any longtail and skewness. Airborne geophysics datasets are shown in Figure 4. Different feature engineer methods were used on these datasets. Feature engineering of airborne magnetic data is complicated because airborne magnetic data contain a dipolar nature and this dipolar anomaly belongs to the same source. The step-by-step feature engineering of airborne magnetic data is shown in Figure 5. In the first step,vertical derivative filter was applied to airborne magnetic data (Figure 5a). This filter removes long wavelength anomalies, enhances the edge anomalies, and sharpens shallow source anomalies. The distribution of the first vertical derivative of magnetic intensity shows a more dense distribution. The absolute value of the first vertical

derivative data decreases the dipolar nature of magnetic data (Figure 5b). The visual surveying of the resulting map confirms all magnetic anomalies are still located in the correct position. The distribution of result data is abnormal too. The logarithmic transformation was applied to the absolute value of the first vertical derivative data. In the final step lowpass filter was applied to decrease noise (Figure 5c). The distribution of lowpass filter data is normal. For airborne radiometric data, the moving box type convolution filter was applied to remove stochastic noises. The feature engineered of airborne geophysics data is presented in Figure 6. For further evaluating the normal distribution of feature engineering of airborne geophysics data, the cross-plot of raw data shows outlier data and long-shape form but the cross-plot of feature engineer data shows round scatter pattern (Figure 7). Some outlier data in Figure 7a and in Figure 7b can be seen as normalized and well-rounded data. This wellrounded scatter pattern is an appropriate shape for input in the FCM machine learning method. Then the feature engineered data were integrated with the FCM method. The feature engineer of airborne magnetic data, thorium, and potassium layers was selected for the input in the machine learning method. Because uranium is a mobile radioelement, it was removed from this data integration. Also, the histogram of feature engineered of airborne geophysics data is presented in Figure 8. These histograms confirm the normal distribution of feature engineered data.



**Figure 5. Feature engineered of airborne magnetic data (a) first vertical derivative of airborne magnetic data (b) the absolute value of the first vertical derivative of magnetic data (c) logarithmic transformation of airborne magnetic data.** 



**Figure 6. Feature engineered map of airborne geophysics data (a) magnetic data (b) potassium (c) thorium.**



**Figure 7. Cross-plot of airborne magnetic data versus potassium (a) before feature engineering (b) after feature engineering.** 



**Figure 8. Histogram of Feature engineered of airborne geophysics data (a) magnetic data (b) potassium (c) thorium.**

#### **6. The optimal number of clusters**

To operate FCM unsupervised machine learning method on airborne geophysics data, the number of clusters should be determined. There is no fixed method to determine the number of clusters. The best method is running the FCM algorithm with different numbers of clusters and identifying the great number of clusters with a given criterion. Different methods were proposed to select an optimal number of clusters such as the Silhouette score [9], the Calinski-Harabasz [17] score, the Elbow method [18] and the DBI method [19].

In this study, two methods were applied to determine an optimal number of clusters. Figure 9 shows WCSS values obtained with Calinski-Harabasz (Figure 9a) and Davis and Bouldin method (Figure 9b). The curves of these two methods confirm the optimal cluster for the FCM machine learning method is six. It can be mentioned that the determination of optimal clusters depends on the method used for clustering and the expert's experience (Duan et al., 2022, 2023). The 2D cross-plot method was used to show number clusters with Calinski-Harabasz and Davis and Bouldin methods (Figure 10).



**Figure 9. Assisting in cluster performance using (a) the Calinski-Harabasz method and (b) Davis and Bouldin method.**



**Figure 10. 2D cross-plot of several clusters with (a) the Calinski-Harabasz method and (b) the Davis and Bouldin method.** 

#### **7. Results and Discussion**

The histogram distributions of feature engineer of aeromagnetic data and aero radiometric data are normal distribution and the range of feature engineer of magnetic data is -2.5 to 0.1. The range of feature engineered of potassium is between 4.2 to 15.5 and the range of thorium is between 1.9 to 12.3 (Figure 8). The cross-plot of feature engineered magnetic data and potassium is shown in Figure 10.

The relative feature engineered values are more important to discrete geological units than absolute values. So, the higher feature engineered values belong to the higher magnetic, potassium, and thorium values. And lower feature engineered values belong to the lower magnetic, potassium, and thorium values.

Figure 11 shows the geological map resulting from the FCM machine learning method. Table 1 shows the correlation between geological units with each of the resulting clusters by using a

predicted geology map with FCM. Cluster 1 shows intermediate, higher, and higher magnetic, thorium, and potassium anomalies respectively. Cluster 2 represents the highest magnetic anomaly and higher thorium and potassium anomalies. Cluster 3 can be attributed to the highest magnetic anomaly and intermediate thorium and potassium anomalies. Cluster 4 demonstrates intermediate magnetic anomalies and the highest thorium and potassium anomalies. Also, cluster 5 represents the lowest magnetic anomalies and intermediate thorium and potassium anomalies, and Cluster 6 represents the intermediate, lowest, and lowest magnetic, thorium, and potassium anomalies respectively. Based on this integration, some surface and near-surface information can be gathered. Radioelements data help to gather surface information and airborne magnetic anomalies help to earn near-surface structures thatt are covered with sedimentary rocks.







**Figure 11. Predicted geology map with unsupervised machine learning method.** 

To clarify the correlation of each geology unit with the related cluster, the geology map was overlayed with the cluster map resulting from the FCM method (Figure 12 and Table 1). Cluster 1 shows a correlation with Trachy andesite, Trachy Basalt, Granodiorite, and Phenoandesite units. Cluster 2 is attributed to the Flysh, conglomerate, and Pyroclastic geology units. Cluster 3 is associated with Red tuff and tuffaceous sediments, and Trachy andesite. Cluster 4 is attributed to the Trachy andesite, Dacite, and Granodiorite. Cluster 5 shows Pyroclastic, Quaternary units and Cluster 6 represents Conglomerate, Alluvium, and Flysch units.



Figure 12. Association between geological units and FCM clusters (Validation Map).

Six clusters resulting from the FCM method show an appropriate correlation with the geological map. So, multi-geophysical integration with the FCM method can be a useful tool for predicting geology maps. To quantify each cluster and determine the relation between geological units and resulting clusters, the fuzzy score method was

used. A fuzzy score of each cluster is presented in Figure 12. All fuzzy memberships are higher than 0.72. This parameter shows an appropriate correlation between geological units and the resulting clusters with the FCM machine learning method.



**Figure 13. Fuzzy score of each cluster ( Cluster 1-6).** 

#### **8. Conclusions**

In the recent research, the unsupervised machine learning method (FCM) has been used to predict geological units in the Shahr-e-Babak area in southeastern Iran. To reach this aim, airborne multigeophysical data consisting of aeromagnetic, potassium, and thorium concentration have been used. The following results have been obtained.

- $\triangleright$  Six clusters have resulted with fuzzy c-means machine learning method.
- $\triangleright$  The fuzzy scores of all clusters are higher than 0.72. This event shows an appropriate

correlation between geological units and airborne multi geophysical data.

- $\triangleright$  With attention to the predicted geology map, there are good correlation between the geology map and resulted clusters. However, there are inconsistencies in some areas .
- $\triangleright$  In multi-geophysical data integration with machine learning methods, the role of preprocessing data is very important.
- $\triangleright$  The results of the FCM method provide a suitable perspective for preparing a geological map.
- $\triangleright$  Based on FCM and airborne geophysics data surface and near-surface structures can be enhanced.

#### **References**

[1]. Abedi, M., Mohammadi, R., Norouzi, G.-H., & Mohammadi, M. S. M. (2016). A comprehensive VIKOR method for integration of various exploratory data in mineral potential mapping. *Arabian Journal of Geosciences*, *9*(6): 482.

[2]. Montsion, R. M., Saumur, B. M., Acosta-Gongora, P., Gadd, M. G., Tschirhart, P., & Tschirhart, V. (2019). Knowledge-driven mineral prospectivity modelling in areas with glacial overburden: porphyry Cu exploration in Quesnellia, British Columbia, Canada. *Applied Earth Science*, *128*(4): 181–196.

[3]. Bahri, E., Alimoradi, A., Yousefi, M. (2021). Mineral Potential Modeling of Porphyry Copper Deposits using Continuously-Weighted Spatial Evidence Layers and union Score Integration Method, Journal of Mining and Environment, 12 (3), 743- 751.Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-1*(2): 224– 227.

[4]. Bahri, E., Alimoradi, A., Yousefi, M. (2023). Investigating the performance of continuous weighting functions in the integration of exploration data for mineral potential modeling using artificial neural networks, geometric average and fuzzy gamma operators, *International Journal of Mining and Geo-Engineering,* doi:

10.22059/IJMGE.2023.361593.595080.

[5]. Yuan, F., Li, X., Zhang, M., Jowitt, S. M., Jia, C., Zheng, T., & Zhou, T. (2014). Three-dimensional weights of evidence-based prospectivity modeling: A case study of the Baixiangshan mining area, Ningwu Basin, Middle and Lower Yangtze Metallogenic Belt, China. *Journal of Geochemical Exploration*.

[6]. 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*, 689.

[7]. Kumar, S., Arasada, R. C., & Rao, G. S. (2023a). Multi-Scale Potential Field Data Integration Using Fuzzy c-Means Clustering for Automated Geological Mapping of North Singhbhum Mobile Belt, Eastern Indian Craton. *Minerals*, *13*(8): 1014.

[8]. Wang, Y., Ksienzyk, A. K., Liu, M., & Brönner, M. (2021). Multigeophysical data integration using cluster analysis: Assisting geological mapping in Trøndelag, Mid-Norway. *Geophysical Journal International*.

[9]. Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*: 53–65.

[10]. Paasche, H., & Eberle, D. G. (2009). Rapid integration of large airborne geophysical data suites using a fuzzy partitioning cluster algorithm: a tool for geological mapping and mineral exploration targeting. *Exploration Geophysics*, *40*(3): 277–287.

[11]. Sun, J., & Li, Y. (2011). Geophysical inversion using petrophysical constraints with application to lithology differentiation. *SEG Technical Program Expanded Abstracts*.

[12]. Afzal, P., Farhadi, S., Shamseddin Meigooni, M., Boveiri Konari,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 doi: 10.22059/GEOPE.2022.336072.648644.

[13]. Safari, M., Maghsoudi, A., & Pour, A. B. (2018). Application of Landsat-8 and ASTER satellite remote sensing data for porphyry copper exploration: a case study from Shahr-e-Babak, Kerman, south of Iran. *Geocarto International*, *33*(11): 1186–1201.

[14]. Tangestani, M. H., & Moore, F. (2001). Porphyry copper potential mapping using the weights-ofevidence model in a GIS, northern Shahr-e-Babak, Iran. *Australian Journal of Earth Sciences*, *48*(5): 695–701.

[15]. Hashemi, S. E., Gholian-Jouybari, F., & Hajiaghaei-Keshteli, M. (2023). A fuzzy c-means algorithm for optimizing data clustering. *Expert Systems with Applications*.

[16]. Kumar, V., Kedam, N., Sharma, K. V., Mehta, D. J., & Caloiero, T. (2023b). Advanced Machine Learning Techniques to Improve Hydrological Prediction: A Comparative Analysis of Streamflow Prediction Models. *Water (Switzerland)*.

[17]. Calinski, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics - Theory and Methods*, *3*(1): 1–27.

[18]. Thorndike, R. L. (1953). Who belongs in the family? *Psychometrika*, *18*(4): 267–276.

[19]. Davies, L., Bouldin, W. (1979). A Cluster Separation Measure". *IEEE Transactions on Pattern Analysis and Machine Intelligence. PAMI-1 (2)*: 224– 227.

[20]. Duan, X., Ma, Y., Huang, H., & Wang, B. (2022). A Novel Cluster Validity Index Based on Augmented Non-Shared Nearest Neighbors. *SSRN Electronic Journal*.

[21]. Duan, X., Ma, Y., Zhou, Y., Huang, H., & Wang, B. (2023). A novel cluster validity index based on augmented non-shared nearest neighbors. *Expert Systems with Applications*.

## **ادغام داده هاي ژئوفیز یک هوابرد با روش یادگ ی ر ي ماشین ی بدون نظارت means-c فازي برا ي پیش بینی نقشه زمین شناسی، منطقه مورد مطالعه شهربابک، جنوب ا یران**

**مسلم جهانتیغ و حمیدرضا رمضی\*** 

**بخش مهندسی معدن، دانشگاه صنعتی امیرکبیر، تهران، ایران** 

ارسال /08 ،2024/02 پذیرش /21 2024/04

 $\mathrm{Ramazi}\allowbreak(\overline{\mathscr{Q}}\allowbreak\mathrm{aut.ac}.\text{ir}$  :ما مسئول مکاتبات  $*$ 

#### **چکیده:**

فازي (FCM (means-c یک الگوریتم یادگیري ماشـ ینی بدون نظارت اسـت. این روش به یکپارچه سـازي داده هاي ژئوفیزیک هوابرد و اسـتخراج نقشـه زمین شـناسـی خودکار کمک می کند. این مقاله سـعی دارد با ترکیب داده های ژئوفیزیک هوابرد متشـکل از لایه های مغناطیس هوایی، پتاسـیم و توریم، نقشـه سـنگ شـناسـي منطقه شـهربابک را به عنوان منطقه پورفيري درجه یک در جنوب ایران طبقه بندي کند. خوشـه هاي حاصـل با FCM همخواني مناسبی را با نقشـه زمین شـناسـی منطقه مورد مطالعه نشـان می دهند. خوشـهها با ناهنجاریهای مغناطیسـی بالا مربوط به سنگـهای آتشفشانی مافیک و خوشـههایی با اثر پرتوسـنجی بالا مرتبط با سنگـهای آذرین هستند. خوشه با ناهنجاری مغناطیسی کم و غلظت کم عناصر رادیواکتیو نشان دهنده سنگ های رسوبی است. برخی از خوشه ها با دو یا چند سازنده سنگ شناسی همراه هستند که به دلیل خواص ژئوفیزیکی مشابه می باشد. امتیاز فازی در همه خوشـهها بالای ۰.۷۱ است که نشـاندهنده همبسـتگی بالایی بین واحدهای زمینشـناسی و دادههای ژئوفیزیکی است. این مطالعه نشـان می دهد که داده هاي ژئوفیزیکی تحلیل شده با روش یادگیري ماشینی می توانند واحدهاي زمین شناسی را آشکار کنند.

**کلمات کلیدي:** یادگیري ماشین، FCM، شهربابک، ژئوفیزیک هوابرد، نقشه زمین شناسی.