International Journal of Mining and Geo-Engineering

IJMGE

IJMGE 54-1 (2020) 33-38

DOI: 10.22059/ijmge.2019.273558.594778

CRFA-CRBM: a hybrid technique for anomaly recognition in regional geochemical exploration; case study: Dehsalm area, east of Iran

Ahmad Aryafar [a, *], Hamid Moeini [b], Vahid Khosravi [a]

[a] *Department of Mining, Faculty of Engineering, University of Birjand, Birjand, Iran*

[b] *Department of Mining and Metallurgy, Faculty of Engineering, University of Yazd, Yazd, Iran*

Article History:

Received: 11 January 2019,
Revised: 29 May 2019
Accepted: 30 May 2019.

ABSTRACT

Identification of geochemical anomalies is a crucial step in regional geochemical explorations. In this regard, new techniques have been developed based on deep learning networks. These simple-structure-networks act as human brains in processing the data by simulating deep layers of thinking. In this paper, a hybrid compositional-deep learning technique was applied to identify anomalous zones in the Dehsalm area located in 90 km of SW-Nehbandan, a town in South Khorasan province, Iran. The compositional robust factor analysis (CRFA) was applied as a tool to select a meaningful subset as an input to Continuous Restricted Boltzmann Machine (CRBM). The dataset consists of 635 stream sediment geochemical samples analyzed for 21 elements. Using CRFA, the 3rd factor (i.e. Pb, Zn, Cu, Ag, Sb, Sr, Ba, Hg, and W), which indicates the occurrence of epithermal mineralization in the area, was considered as an input set to CRBM. The best-performed CRBM with 80 hidden units and stabilized parameters at 150 iterations was finalized and trained on all the geochemical samples of the study area. The average square contribution (ASC) and average square error (ASE) values were determined as anomaly identifiers on the reconstructed error of the trained CRBM. A statistical threshold was applied to the values of the criteria (ASC & ASE), and the resulting outputs were mapped to delineate the anomalous samples. The maps indicated that ASC and ASE had the same performance in multivariate geochemical anomaly recognition. The anomalies were confirmed spatially using mineral prospects of Pb, Zn, Cu, and Sb, as well as several active lead and copper mines in the study area.

Keywords : *Geochemical exploration, compositional data, Robust factor analysis, Deep learning, CRBM, Dehsalm*

1. Introduction

One of the essential steps in mineral exploration is distinguishing prospective areas within the region of interest [1]. The main goal of geochemical explorations is to present a program to distinguish as much hidden mineralization as possible while taking economic and physical restrictions into consideration. In the past decades, numerous methods have been used to identify promising areas ranging from traditional parametric statistical methods to non-parametric techniques [2, 3, 4, 5, 6]. Many researchers have applied univariate statistical and graphical methods (such as probability graphs, fractal geometry, spatial U-statistic, and box plot) and multivariate techniques, particularly factor analysis, in geochemical anomaly recognition [7, 8, 9, 10, 11, 12]. Multivariate methods commonly involve the multivariate geochemical background of a sample population to gratify an acknowledged statistical distribution, such as multivariate normal distribution, so that the general properties of the multivariate geochemical background can be easily designated by a predefined simple function. However, complex geological settings often result in mysterious complex multivariate probabilities [13]. Mineral occurrence data are used chiefly to understand and validate mapped geochemical anomalies. [6] Believe that each method has advantages and disadvantages, and each method may be suitable for a specific geological environment or a regional exploration scenario. Previous investigations [3, 4, 5, 6] have revealed that due to geochemically complex nature of arid areas, conventional statistical approaches are not appropriate for the detection of geochemical anomalies. Additionally, the compositional nature of stream sediment geochemical data should be considered for geochemical anomaly discrimination. After Aitchison presented the idea of compositional data analysis (CoDa) to discover the relationships among closed system datasets [14], the idea was used in some researches to identify multivariate geochemical anomalies [2, 15, 16, 17, 18]. Thereafter, it has become an interesting topic in the field of multivariate statistics [19, 20]. In past years, the use of data-mining techniques and metaheuristic algorithms has become common. Several tasks, formerly done at the expense of significant amounts of time and money, can be performed more efficiently utilizing these techniques and algorithms [21, 22].

Machine learning algorithms improve approaches that can automatically recognize patterns in data and then use the revealed patterns to predict future data or other outcomes of interest [23]. Pattern recognition is troubled with automatic discovery of regularities in data through the use of computer algorithms and these regularities to take actions [24]. Many typical machine learning techniques abuse narrow architectures. The most mention of these methods includes kernel regression, random forest (RF), conventional hidden Markov models, maximum entropy models, support vector machines, and multilayer perceptron with a single invisible layer. The shared property in these light learning models is the simple architecture that comprises only one layer responsible for transforming the raw input signals or structures into a problem-specific structure [11, 25]. Artificial neural networks can be used as advantageous methods to investigate mineral resources [26]. Neural networks are highly complex, nonlinear systems with high degrees of freedom that employ various principles of

* Corresponding author. Tel:+98-9155616379, *E-mail address*: aaryafar@birjand.ac.ir (A. Aryafar).

information processing from nonlinear computing systems [6]. In the past decades, deep learning models have been used broadly in computer vision and image processing. It has been proven that deep learning neural networks outperform traditional classification schemes in many usage cases [27]. The main characteristics of deep learning models are (1) the generative nature of the model, which generally requires an additional top layer to perform the discriminative task; (2) An unsupervised pre-training stage that creates an effective use of large amounts of unlabeled training data for extracting structures and regularities in the input signals [11]. In other words, deep learning is a multi-algorithm technique on learning multiple levels of presentation of the data. Deep belief models (DBM) are known as the most implemented deep learning networks. A restricted Boltzmann machine (RBMs) as a greedy algorithm is known as the core of deep belief networks. The structure of RBM network involves two layers: the visible and hidden layers. It is necessary to modify the model's variables so that the generated vector is as close to the input vector as possible. The generated vector is the vector obtained from probabilistic inference from the hidden layer, the values of which are in turn gained by the probabilistic inference from the visible layer, i.e. from the original vector [2, 11, 28, 29]. Therefore, selecting a suitable method or algorithm is essential in producing an accurate mineral potential map. It depends, mainly, on the capacity of the algorithm to learn complex relationships between the input evidential features and the occurrence of mineral deposits. However, interpretability and transparency must also be considered [25]. In this paper developing a hybrid model (compositional robust factor analysis before applying a deep learning network) was studied to recognize the anomalous areas from stream sediment geochemistry data in Dehsalm 1:100,000 geological map located near the Nehbandan city in the East of Iran. The resulting maps were confirmed with several mineral prospects of the area.

2. Geological Setting of the Study Area

The Dehsalm area (31°-31.5° N, 59°-59.5° E) is located in 90 km of SW-Nehbandan City, South Khorasan Province, Iran. It is a part of the Lut block and is divided into two parts: one part contains tertiary volcanic rocks, and the other part is composed of metamorphic rocks. The Cretaceous units are dominant in the Northeastern and Eastern parts of the study area, which form the highest elevation point in the region. The lowest point (740m high) is located in the basin (playa). The Dehsalm area is an arid region with highly variable temperature, with a maximum temperature of 50^∘C in summers, and average annual precipitation of about 80mm. Fig. 1 shows the modified 1:100,000 geological map of the study area.

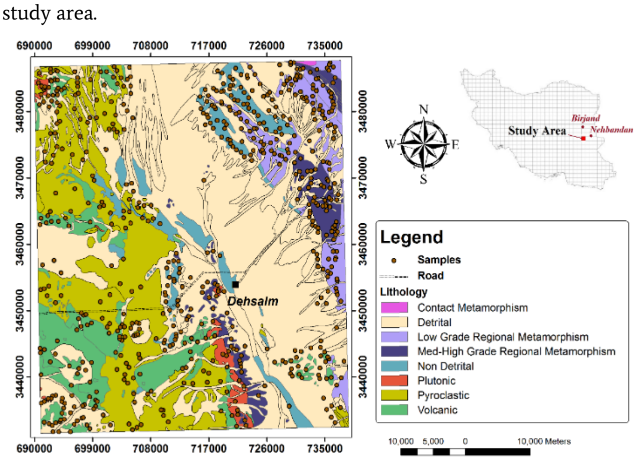

Fig. 1. The modified 1:100,000 geological map of the study area with the locations of taken stream sediment samples.

This area has a high mineral potential. Several porphyroblast minerals, including garnet and andalusite, have occurred in metamorphic rocks. Due to their specific features such as hardness and high melting point, each porphyroblast is minerally remarkable. Of these potentials, it can be referred to the Northeastern Andalusite, Ltd. Also, a considerable volume of Garnet crystals has occurred in the east and south of the area. Another mineral prospect has occurred within the skarns of NW-Dehsalm. The intrusion of the granitoid suites as well as tectonic activities have caused copper mineralization in the boundary of the faults. Some Ag-bearing Galena prospects in the north and east of Dehsalm are currently being extracted by local miners.

3. Approaches and Materials

3.1. Data Collection and Preparation

A regional geochemical survey was carried out by the Industry, Mine and Trade Organization (IMTO) of South Khorasan province over the 1:100,000 geological map of Dehsalm. Totally, 635 stream sediment samples (Fig. 1) were analyzed for 21 elements (Au in ppb and other elements in ppm).

The overall study procedure is divided into three parts: 1) compositional robust principal factor analysis to select a proper input set; 2) applying continuous restricted Boltzmann machine on raw (untransformed) data of the input set; 3) checking the resulting maps with field evidence.

3.2. The Hybrid Method

3.2.1. Compositional Robust Factor Analysis

Compositional data (CoDa) such as stream sediment data are the observations present parts of a whole. The true information of compositional data is included in the ratios between the components of each observation whether the whole composition is analyzed or not. In geochemistry, all analytical results are thus intrinsically compositional [18, 30]. The first step was exploratory data analysis. The best visual tool for this task is CoDa-biplot. It is used to investigate the codependence between parts of the observations and principal components [31].

After imputation of the missing data using the ilr-EM method [32], the robust factor analysis (RFA) is applied to the compositional data. The centered log-ratio (clr) transformation; y=clr(x) is used to open the matrix x of compositional observations (with D dimensions or variables) [20]. Then, the factor analysis model is applied to y as below:

$$y=\Lambda f+e \tag{1}$$

In Eq. 1; f, e, and Λ are factors (less than D dimensions), error and loadings matrices, respectively. After rewriting Eq. 1 based on covariance, Eq. 2 is resulted [20]:

$$Cov(y)=\Lambda\Lambda^{T}+Cov(e) \tag{2}$$

In Eq. 2, Cov(e) is a diagonal matrix with elements called unique variances that include the variance of the model's components that is not explained by the factors [20]. In order to solve the singularity problem of Cov(y) due to clr properties, [20] suggested that the diagonal matrix of Cov(e) should be replaced by Ψ^{*}=HCov(e)H in Eq. 2 [20].

$$H=(I_D-\frac{1}{D}J_D) \tag{3}$$

Where, I and J are known as the identity and one matrices, respectively. The problem then turns into estimating the parameters Λ and Ψ^{*} in a multi-step algorithm proposed in [20]. It continues until the elements in Ψ^{*} stabilize. In order to make a better interpretation, the loadings matrix Λ is then rotated orthogonally.

Both loadings values and the classical factor analysis have the same interpretation; the high amounts of loadings will address the high influence of corresponding clr variables on the factor. However, if a factor has similar loadings, it will not change the ratio x_i/x_j, i.e. this factor is not related to any relative enrichment or depletion of x_i with respect to x_j [14]. Finally, the regression techniques are applied for factor scores estimating.

In order to minimize the role of outliers on the estimation of the parameters, it is suggested that a robust covariance matrix should be used instead of the sample covariance matrix (Cov(y) in Eq. 2). A popular choice is the minimum covariance determinant (MCD) regressor, for which a rapid algorithm is stated by [33].

The MCD regressor searches a subset h out of n observations with the minimum determinant of their sample covariance matrix. The values are univariately ilr-transformed and then used in the MCD estimator (i.e. z=ilr(x)). In the following, the covariance matrix of the new variable (z) is back-transformed into the clr-domain. This robust version of Cov(y) can be used for the mentioned algorithm of the parameter estimation in the factor analysis [20]. All mentioned CRFA operations can be directly carried out using the 'robCompositions' package in R software [34].

3.2.2. Continuous Restricted Boltzmann Machine

Nowadays, the theory and practice of artificial neural networks have been developed through successful deep learning methods. The term refers to multi-layer neural networks similar to those used in the past except that now it is possible to use more hidden layers than before. Deep learning is a multi-algorithm based on learning multiple steps of data production. Before training the net, the data must be normalized to fit in the range (-1, +1) or (0,1). Mostly, Eq. 4 is used that in this study was applied too as below:

$$\tilde{x}_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \bullet i \in [\text{observations}] \tag{4}$$

Restricted Boltzmann Machines (RBMs) have been widely used for multivariate distribution modeling, high-dimensional temporal sequence modeling, and geochemical anomaly recognition [13, 28, 35, 36, 37]. Typically, RBMs work properly based on binary or Gaussian variables, but their applications for other kinds of variables such as continuous-value non-Gaussian inputs have been limited [35, 36, 37]. Continuous restricted Boltzmann machine (CRBM) is a well-known deep learning net with a stochastic network of units where each unit has some random behaviors when activated. It comprises one visible layer and one hidden layer with only interlayer connections. The probability of creating a visible vector is in agreement with the product of the probabilities that the visible vector would be generated by each of the hidden units acting alone. Therefore, the model is an output of experts with one expert per hidden unit [13]. Fig. 2 illustrates the structure of CRBM with six visible units, three hidden units, and two permanently-on bias units.

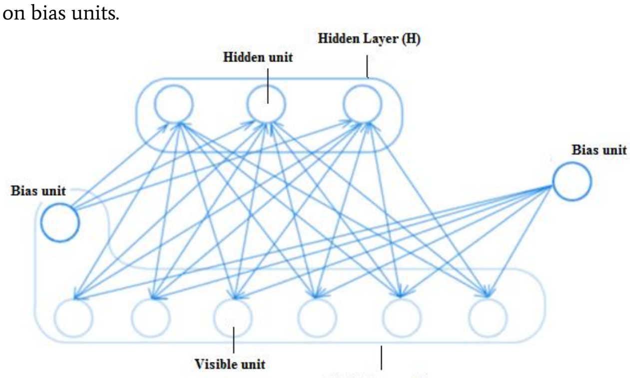

Fig. 2. Restricted Boltzmann machine structure with six visible units and three hidden units (modified after [29]).

The visible and hidden units have continuous states generated by adding a zero-mean Gaussian noise to the input of a sampled sigmoid-activated unit. They are connected by a weight matrix. In each training repeat, a contrastive divergence is updated between the situation of the visible and hidden units. In practice, the model identifies and reconstructs the training samples population that is sequentially presented to it while the interlayer connection weights are modified. CRBM can be taken as an associative memory that is powerful to encode a complex non-linear non-Gaussian training data distribution [13]. Therefore, CRBM is appropriate for modeling the complex multivariate probability distribution of a geochemical sample population taken from a complex geological setting. Careful training of CRBMs is essential in obtaining successful practical results [11]. The resulting difference between the final model after training and the input model is called "reconstructed error".

Based on the reconstructed error, two essential indicators are defined to detect multivariate geochemical anomalies from the background; average square contribution (ASC) and average square error (ASE) [13]. In Eq. 5 & 6, p and q are the number of visible (except bias) and hidden (except bias) units respectively [13].

$$ASC = \frac{1}{pq}\sum_{ij} (v_i h_j - \hat{v}_i \hat{h}_j)^2 \tag{5}$$

$$ASE = \frac{1}{p}\sum_{i=1}^{p} (v_i - \hat{v}_j)^2 \tag{6}$$

In a stabilized model, they can be calculated for every observation. The threshold values of the indicators (ASC and ASE) are the 90th percentile. Some parameters can be set to manage the efficiency of CRBM and its stability. They include the number of hidden units and iterations, noise control coefficients for visible and hidden units, learning momentum, learning cost, and learning rates. A practical way is to train the network on the whole data with high iteration and then to decide about the change of parameters to reach stability [13].

4. Results and Discussion

The stream sediment samples were analyzed primarily for compositional statistics of the simplex (Table 1). The total variance is 5.2834. The statistics show a relatively high clr-variance of W (4.339), which might be due to hydrothermal genesis of the dominant mineralization in the region of the study.

Then, the robust factor analysis (CRFA) was applied to the compositionally scaled data to reduce the dimension and choose an input set to the CRBM network. Six factors were extracted as the most meaningful ones according to the geology of the region and the lowest uniqueness values. They explained about 92% of the variation.

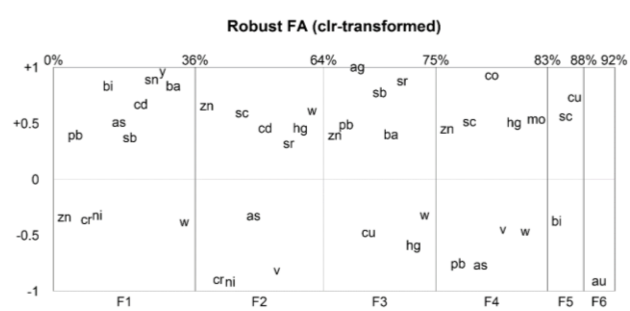

Fig. 3. Loadings plot of the robust factor analysis.

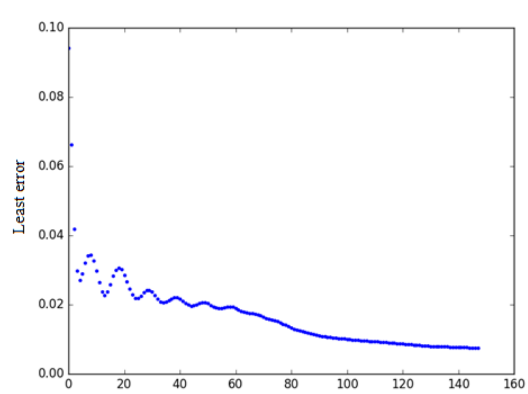

Fig. 4. MSE plot of CRBM after 150 epochs.

Table 1. Compositional statistics summary

	Zn	Pb	Ag	Cr	Ni	Bi	Sc	Cu	As	Sb	Cd	Co	Sn	Y	Ba	V	Sr	Hg	W	Mo	Au
Center	0.068	0.024	0.001	0.147	0.068	0.000	0.006	0.043	0.010	0.001	0.000	0.020	0.005	0.023	0.341	0.073	0.158	0.000	0.008	0.001	0.001
Median	0.066	0.023	0.001	0.144	0.066	0.000	0.006	0.042	0.010	0.001	0.000	0.019	0.005	0.022	0.328	0.073	0.151	0.000	0.019	0.001	0.001
clr-Var	0.020	0.022	0.040	0.084	0.117	0.034	0.044	0.021	0.060	0.047	0.026	0.046	0.018	0.021	0.015	0.096	0.015	0.070	4.339	0.072	0.077

As it can be seen from the loadings plot in Fig. 3, the 3rd factor with Pb, Zn, Cu, Ag, Sb, Sr, Ba, Hg, and W, represents the low-sulfide epithermal mineralization in the area which is a part of the polymetallic belt.

A new dataset was formed from the raw data by removing all elements except for factor 3. It was then normalized as the input of the CRBM network. The net parameters were set experimentally with trial and error in different iterations until the best performance was obtained and, at the least error level of about 0.01 was stable. The performance shows that the net has been stable with 80 hidden units, the standard deviation of 0.2 for normal noise, coefficient of 0.5 for noise controls of both hidden and visible units, learning momentum of 0.9, learning cost of 0.00001 for 150 iterations (Fig. 4).

After applying ASE and ASC criteria, anomalous samples were determined by putting thresholds on the values (Fig. 5). The horizontal axis represents the samples indexes, and the red lines are the 90th percentiles (0.0127 for ASC and 0.0549 for ASE). The values above the lines are related to anomalous samples. The results were mapped using binary classification based on the thresholds and also stretched along the color map to give a better view of anomalous halos (Fig. 6). The low and high values are colored blue and red, respectively.

By comparing the anomalies and field mineral prospects in Fig. 6, it is confirmed that CRBM has favorably detected the anomalous samples associated with hydrothermal mineralization in skarns. This is a remarkable advantage of CRBM that can reconstruct any statistical distribution of input samples and then detect its anomalies.

However, CRBM has drawbacks; for instance, various parameters have to be set and CPU dependency in case of big data. Some parameters are more important than the others; increasing the iteration can give a more stable model as well as the sufficient number of hidden units that can prevent overgeneralization and spurious results.

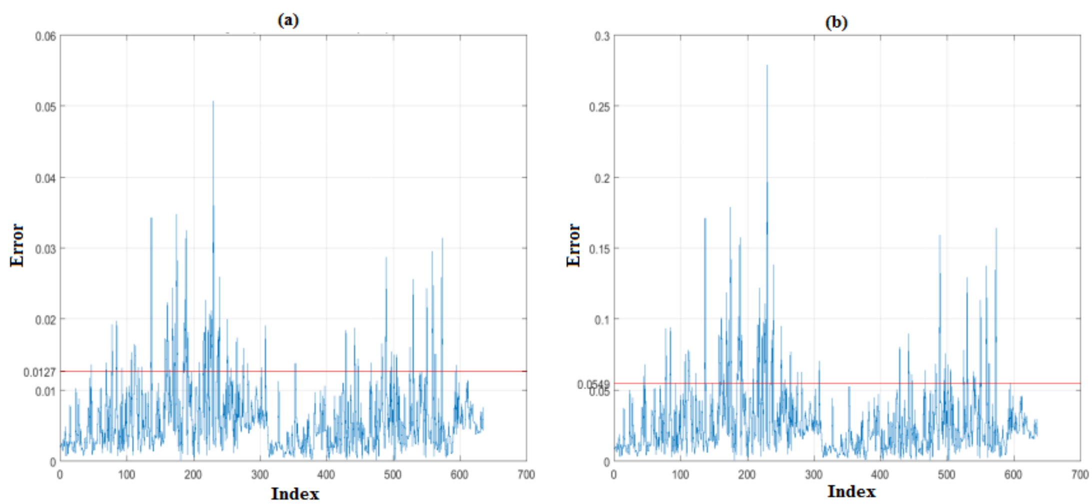

Fig. 5. Anomaly indicators of the stable network (ASC (a) and ASE (b))

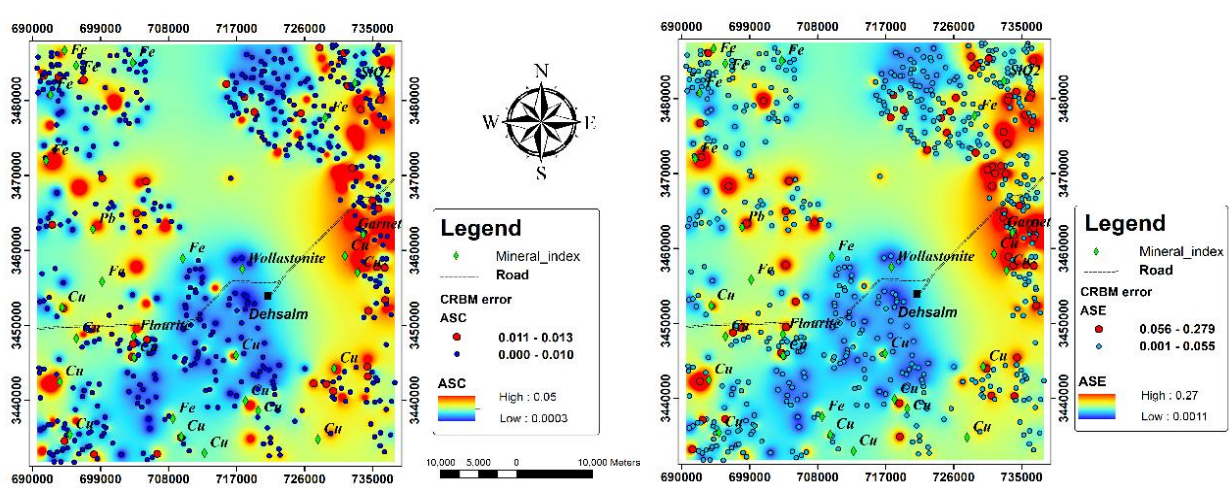

Fig. 6. Anomaly maps of ASC and ASE indicators in the study area. The mineral prospects and occurrences are marked on the maps.

5. Conclusion

The critical challenge during regional geochemical exploration is to detect geochemical anomalies from the background. Many algorithms have been presented, from statistical thresholds to advanced data mining methods, to deal with this challenge. One of the most effective and recent techniques is machine learning through deep belief networks. It has been widely used in many scientific fields and has provided successful results in anomaly recognition. Concerning the compositional nature of geochemical data, a hybrid compositional-deep learning approach was constructed and applied on 635 stream sediment samples taken over the Dehsalm 1:100,000 geological map in South Khorasan province of Iran for the identification of geochemical anomalies.

After implementing the compositional robust factor analysis (CRFA) on the study area, a subset of raw data with the elements detected in the 3rd factor (Pb, Zn, Cu, Ag, Sb, Sr, Ba, Hg, and W) was specified as input variables to the CRBM. Consequently, after normalizing the data, the CRBM with the best performance (least error level of about 0.01) was confirmed. Furthermore, the number of visible and hidden units, noise-control coefficients for both hidden and visible units, the standard deviation of normal noise, learning rates, learning cost, and learning momentum were predetermined. In geochemical prospecting, generally, anomalous samples happen much less than the background samples. Same as the small probability samples, the geochemical anomaly samples can be recognized using the trained CRBM from the training geochemical sample population. The ASC and ASE techniques were applied to detect multivariate anomalous samples from the training geochemical sample population. The threshold values of ASC and ASE were estimated using the 90th percentile equal to 0.0127 and 0.0549, respectively. The higher values were considered as anomalous samples. The obtained anomaly halos through this method were checked with mineral prospects on the field. The indications of hydrothermal minerals such as Flourite (as an active mine in the area), Galena (various occurrences), Calcite, Malachite and Sphalerite (as numerous copper prospects in the central and southern parts of the area) are evident in most of the high-valued halos marked in the ASE and ASC maps. This indicates that CRBM has performed favorably in detecting the anomalous samples which is considered as a notable advantage of CRBM. It can rebuild any statistical distribution of the input samples and then identify its anomalies. The identified geochemical anomaly samples represented the low-sulfide epithermal mineralization in the area which is a part of the polymetallic belt. The presence of active lead and copper mines confirms the mineralization in the study area.

Acknowledgment

The authors express their special thanks to Industry, Mine & Trade Organization of South Khorasan province for providing the data. Also, the financial support of University of Birjand is grateful.

REFERENCES

[1] Abedi, M., & Norouzi, G. H. (2012). Integration of various geophysical data with geological and geochemical data to determine additional drilling for copper exploration. Journal of Applied Geophysics, 83, 35–45.

[2] Aryafar, A., & Moeini, H. (2017). Application of continuous restricted Boltzmann machine to detect multivariate anomalies from stream sediment geochemical data, Korit, East of Iran. Journal of Mining and Environment, 8(4), 673–682.

[3] Ziaii, M., Abedi, A., & Ziaii, M. (2007). Prediction of hidden ore bodies by new integrated computational model in marginal Lut region in east of Iran. In Proc. Exploration. 7: 957–961.

[4] Ziaii, M., Pouyan, A. A., & Ziaei, M. (2009a). Neuro-fuzzy modelling in mining geochemistry: identification of geochemical anomalies. Journal of Geochemical Exploration, 100(1), 25–36.

[5] Ziaii, M., Soleymani, A. A., Kamkar-Rouhani, A., & Modarres, H. R. (2009b). Geochemical data analysis of Deh-Salm area using neuro-fuzzy networks. Geochimica et Cosmochimica Acta Supplement, 73, A1533.

[6] Ziaii, M., Doulati-Ardejani, F., Ziaei, M., & Soleymani, A. A. (2012). Neuro-fuzzy modeling based genetic algorithms for identification of geochemical anomalies in mining geochemistry. Applied geochemistry, 27(3), 663–676.

[7] Cheng, Q. (2007). Mapping singularities with stream sediment geochemical data for prediction of undiscovered mineral deposits in Gejiu, Yunnan Province, China. Ore Geology Reviews, 32(1–2), 314–324.

[8] Deng, J., Wang, Q., Yang, L., Wang, Y., Gong, Q., & Liu, H. (2010). Delineation and explanation of geochemical anomalies using fractal models in the Heqing area, Yunnan Province, China. Journal of Geochemical Exploration, 105(3), 95–105.

[9] Wackernagel, H. (2013). Multivariate geostatistics: an introduction with applications. Springer Science & Business Media.

[10] Wang, J., Zuo, R., & Caers, J. (2017). Discovering geochemical patterns by factor-based cluster analysis. Journal of Geochemical Exploration, 181, 106–115.

[11] Yu, D., & Deng, L. (2011). Deep learning and its applications to signal and information processing. Signal Processing Magazine, 28(1), 145–154.

[12] Zuo, R., Cheng, Q., Agterberg, F. P., & Xia, Q. (2009). Application of singularity mapping technique to identify local anomalies using stream sediment geochemical data, a case study from Gangdese, Tibet, western China. Journal of Geochemical Exploration, 101(3), 225–235.

[13] Chen, Y., Lu, L., & Li, X. (2014). Application of continuous restricted Boltzmann machine to identify multivariate geochemical anomaly. Journal of Geochemical Exploration, 140, 56–63.

[14] Aitchison, J. (1986). The statistical analysis of compositional data. Chapman and Hall London.

[15] Filzmoser, P., & Gschwandtner, M. (2012). mvoutlier: Multivariate outlier detection based on robust methods. R package version, 1(7).

[16] Gallo, M., & Buccianti, A. (2013). Weighted principal component analysis for compositional data: application example for the water chemistry of the Arno river (Tuscany, central Italy). Environmetrics, 24(4), 269–277.

[17] Hosseinpoor, M. K., & Aryafar, A. (2016). Using robust staged R-mode factor analysis and logistic function to identify probable Cu-mineralization zones in Khusf 1: 100,000 sheets, east of Iran. Arabian Journal of Geosciences, 9(2), 157.

[18] Moeini, H., Torab, F. M. (2017). Comparing compositional multivariate outliers with autoencoder networks in anomaly detection at Hamich exploration area, east of Iran". Journal of Geochemical Exploration, 180, 15–23.

[19] Filzmoser, P., Garrett, R. G., & Reimann, C. (2005). Multivariate outlier detection in exploration geochemistry". Computers & Geosciences, 31, 579–587.

[20] Filzmoser, P., Hron, K., Reimann, C., & Garrett, R. (2009). Robust factor analysis for compositional data. Computers and Geosciences, 35(9), 1854–1861.

[21] Hajighorbani, M., Hashemi, S. M. R., Minaei-Bidgoli, B., & Safari, S. (2016). A review of some semi-supervised learning methods. In First International Conference on New Research Achievements in Electrical and Computer Engineering.

[22] Hashemi, S. M. R., & Broumandnia, A. (2015). A Review of Attention Models in Image Protrusion and Object Detection. The Journal of Mathematics and Computer Science, 15(4), 273–283.

[23] Murphy, K. P. (2012). Machine Learning: A Probabilistic Perspective. Adaptive Computation and Machine Learning. MIT press.

[24] Bishop, C. (2006). Pattern Recognition and Machine Learning. Springer-Verlag New York.

[25] Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. (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.

[26] Bárdossy, G., & Fodor, J. (2004). Evaluation of uncertainties and risks in geology: new mathematical approaches for their handling. Springer Science & Business Media.

[27] Chamatidis, I., Katsika, A., & Spathoulas, G. (2017). Using deep learning neural networks for ECG based authentication". In Security Technology (ICCST), International Carnahan Conference on (pp. 1–6).

[28] Hinton, G. E. (2012). A practical guide to training restricted Boltzmann machines. In Neural networks: Tricks of the trade (pp. 599–619). Springer.

[29] Kislov, K. V, & Gravirov, V. V. (2018). Deep Artificial Neural Networks as a tool for the Analysis of Seismic Data. Seismic Instruments, 54(1), 8–16.

[30] Reimann, C., Filzmoser, P., Hron, K., Kynčlová, P., & Garrett, R. G. (2017). A new method for correlation analysis of compositional (environmental) data--a worked example. Science of The Total Environment, 607, 965–971.

[31] Pawlowsky-Glahn, V., & Buccianti, A. (2011). Compositional Data Analysis; Theory and Applications. Wiley & Sons, Ltd.

[32] Palarea-Albaladejo, J., & Martín-Fernández, J. A. (2015). zCompositions—R package for multivariate imputation of left-censored data under a compositional approach. Chemometrics and Intelligent Laboratory Systems, 143, 85–96.

[33] Rousseeuw, P. J., & Driessen, K. V. (1999). A fast algorithm for the minimum covariance determinant estimator. Technometrics, 41(3), 212–223.

[34] Templ, M., Hron, K., & Filzmoser, P. (2011). robCompositions: an R-package for robust statistical analysis of compositional data. Compositional Data Analysis: Theory and Applications. John Wiley and Sons.

[35] Chen, H., & Murray, A. F. (2003). Continuous restricted Boltzmann machine with an implementable training algorithm. IEE Proceedings-Vision, Image and Signal Processing, 150(3), 153–158.

[36] Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. science, 313(5786), 504–507.

[37] Xiong, Y., & Zuo, R. (2016). Recognition of geochemical anomalies using a deep autoencoder network. Computers & Geosciences, 86, 75–82.