Geochemical Anomaly Detection in the Irankuh District Using Hybrid Machine Learning Technique and Fractal Modeling

Original
Geochemical Anomaly Detection in the Irankuh District Using Hybrid Machine Learning Technique and Fractal Modeling / Afzal, Peyman; Farhadi, Sasan; Boveiri Konari, Mina; Shamseddin Meigooni, Mojtaba; Daneshvar Saein, Lili. - In: GEOPERSIA. - ISSN 2228-7817. - 12:1(2022), pp. 191-199. [10.22059/GEOPE.2022.336072.648644]

Availability:
This version is available at: 11583/2974144 since: 2022-12-23T22:55:23Z

Publisher:
University of Tehran

Published
DOI:10.22059/GEOPE.2022.336072.648644

Geopersia 2022, 12(1): 191-199
DOI: 10.22059/GEOPE.2022.336072.648644

RESEARCH PAPER

Geochemical anomaly detection in the Irankuh District using Hybrid Machine learning technique and fractal modeling

Peyman Afzal [1], Sasan Farhadi [2,*], Mina Boveiri Konari [3], Mojtaba Shamseddin Meigoony [4], Lili Daneshvar Saein [5]

[1] Department of Petroleum and Mining Engineering, South Tehran Branch, Azad University, Tehran, Iran
[2] Department of Structural, Geotechnical and Building Engineering, Polytechnic University of Turin, Italy
[3] Department of Economic Geology, Tarbiat Modares University, Tehran, Iran
[4] Department of Earth Sciences, Science and Research Branch, Azad University, Tehran, Iran
[5] Department of Geology, Payame Noor University, Tehran, Iran

Received: 23 December 2021, Revised: 07 March 2022, Accepted: 14 March 2022

Abstract
Prediction of elemental concentrations is essential in mineral exploration as it plays a vital role in detailed exploration. New machine learning (ML) methods such as hybrid models are robust approaches infrequently used concerning other methods in this field; therefore, they have not been examined properly. In this study, a hybrid machine learning (HML) method was proposed based on combining K-Nearest Neighbor Regression (KNNR) and Random Forest Regression (RFR) to predict Pb and Zn grades in the Irankuh mining district, Sanandaj-Sirjan Zone. The aim of the proposed study is to employ the hybrid model as a new method for grade distribution. The KNNR-RFR hybrid model results have been applied for the Pb and Zn anomalies classification. The hybrid (KNNR-RFR) method has shown more accurate prediction outputs based on the correlation coefficients than the single regression models with 0.66 and 0.54 correlation coefficients for Pb and Zn, respectively. The KNN-RF results were used for the classification of Pb and Zn anomalies in the study area. The concentration-area fractal model separated the main anomalous areas for these elements. The Pb and Zn main anomalies were correlated with mining activities and core drilling data. The current study demonstrates that the hybrid model has a substantial potential for the ore elemental distribution prediction. The presented model expresses a promising result and can predict ore grade in similar investigations.

Keywords: Hybrid Machine Learning, Geochemical Anomaly Detection, K-Nearest Neighbor Regression (KNNR), Random Forest Regression (RFR), Fractal Modeling.

Introduction

One of the essential tasks in mineral exploration is identifying the geochemical anomalies. Prediction of elemental concentration has a crucial role in mineral exploration development. For this purpose, there are several machine learning techniques approaches have been developed in the last decades, such as artificial neural networks (ANN; (Zaremotlagh & Hezarkhani, 2017; Zhang et al., 2021)), support vector machine (SVM; (Zuo & Carranza, 2011; Lin et al., 2021)), and Random Forest (RF; (Carranza & Laborte, 2016; Wang et al., 2020)).

In recent years, ML techniques have developed as a vital tool in various categories of science geoscience research (Dramsch, 2020; Cunxiao et al., 2021; Lin et al., 2021; Richter-Laskowska

* Corresponding author e-mail: sasan.farhadi@polito.it

et al., 2021). The ML algorithms can be divided into four main groups: supervised, unsupervised, semi-supervised, and reinforcement learning (Murphy, 2012; Burkov, 2019; Zuo et al., 2021). One of the most successful algorithms of ML is supervised learning which has two main subdivisions of regression and classification (Muller & Guido, 2016). One of the most successful algorithms of ML is supervised learning which has two main subdivisions of regression and classification (Muller & Guido, 2016).

The fractal methodology has been used to interpret the geochemical data; particularly, for geochemical anomalies classification. This approach is utilized simultaneously to separate anomalous zones determined by their concentrations and occupied spaces (Afzal et al., 2017; Aliyari et al., 2020; Shahsavar et al., 2020; Farshid Koohzadi et al., 2021; Heidari et al., 2021; Pourgholam et al., 2021; Shamseddin Meigooni et al., 2021). As a well-known fractal approach, the concentration-area (C-A) fractal model, was proposed by Cheng et al. (1994) for anomalous area detection. Many researchers have been used and developed this method, e.g., (Meigoony et al., 2014; Saein & Afzal, 2017).

In this article, a hybrid ML approach has been utilized to estimate ore grades (Pb and Zn) in the Irankuh area, Central Iran. Initially, two regression models have been developed, including K-nearest neighbor and random forest, which have been rarely used to predict the ore grades. The obtained results were classified with the C-A fractal model, and at the end, the major Pb and Zn anomalies were selected.

Geological investigation

The Irankuh Mining District formed within the Late Jurassic-Lower Cretaceous back-arc extensional setting, and it is one of Iran's most important sediment-hosted Zn-Pb (Ag-Ba) deposits (Fig. 1). Mineralization occurred as stratiform and stratabound bodies of sulfide and non-sulfide ores within Lower Cretaceous dolostone, siltstone, and crystal lithic tuff. This district includes Tappehsorkh, Rowmarmar, Gushfil and Kolahdarvazeh deposits, Baghabrisham, Gowdezendan, Khanehgorgi and Tofangchiha occurrences. Mining activities occurred as both open pit and underground excavations.

Sulfide mineralization formed as massive and semi-massive replacement, brecciated, laminated, vein-veinlet, disseminated, colloform and framboidal. Mineralization is accompanied by hydrothermal alteration such as dolomitization, silicification and sericitization. Other style of mineralization in the district is known as non-sulfide mineralization where supergene and sometimes later sulfide mineralization were formed. The later style of mineralization is well observed along and around the Gushfil fault where deformation textures resulting from reactivation of fault during orogeny caused increasing the grain size of sulfide minerals, forming strain fringes texture as well as foliation-like textures of sphalerite and galena all of them confirmed later deformation of early sulfide mineralization. Some non-sulfide minerals including smithsonite, hydrozincite, hemimorphite and cerussite also formed along some later domino-normal type faults. Thus, (Boveiri Konari et al., 2017, 2020) and (Boveiri Konari & Rastad, 2018), suggested that the IMD formed as sub-sea floor replacement Sedex-type deposit at Late Jurassic-Lower Cretaceous time and then influenced by orogenic event (Upper Cretaceous) which caused later sulfide and non-sulfide mineralization.

Material and methods

Dataset

In this study, 804 in-situ soil samples have been selected for geochemical exploration. These were analyzed by the ICP-MS method for 35 elements according to Pb-Zn and associated paragenesis elements. These samples were sent by the Bama Company to ALS Chemex (Canada), following a microwave digestion with a 1:1 nitric-hydrochloric acid mixture. Location of these samples is revealed in Fig. 1.

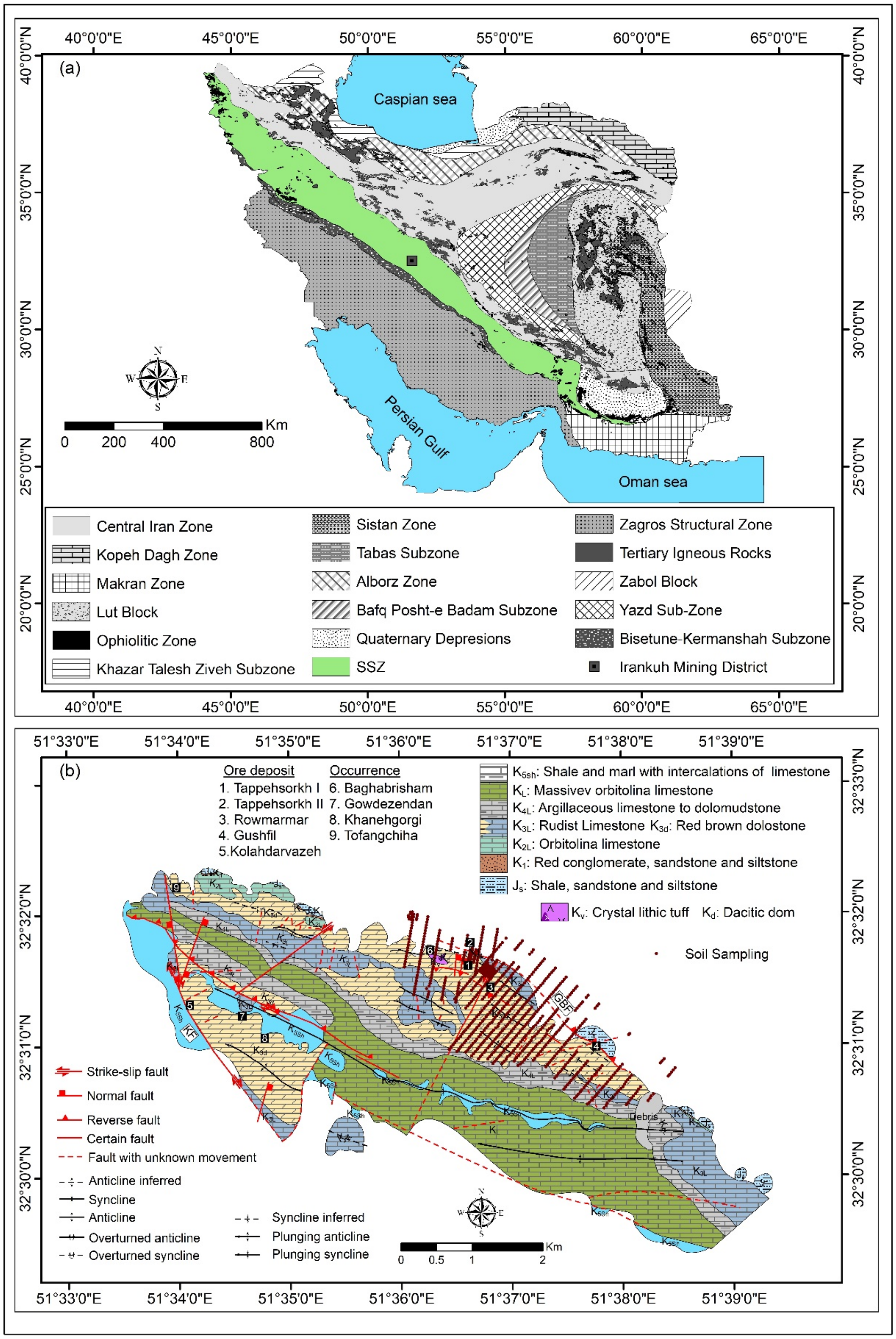

Figure 1. a. Geological-structural map of Iran and location of Irankuh Mining District (IMD) (base map from (Aghanabati, 2004; Alavi, 1994)). b. Geological map of the IMD and location of soil samples (GBF: Gushfil-Baghabrisham fault, KF: Kolahdarvazeh fault)

K_Nearest Neighbor

The KNNR method is recognized as one of the most effective ML algorithms. This method has appropriate performance without many adjustments, and this advantageous makes it a right baseline before employing any advanced technique (Biau et al., 2019; Song et al., 2017). In general, the algorithm finds the nearest point in the training dataset and makes the required prediction for these values.

In general, the algorithm finds the nearest point in the training given dataset (Y) and makes the required prediction for these values. The predicted outcomes are considered f(X), which is computed as the mean response among the K nearest neighbor (J. Chen & Lau, 2016; R. Chen & Paschalidis, 2019).

$$f(X) = \frac{1}{k} \sum_{i=1}^{k} Y_i(X) \quad (1)$$

Random Forest

The initial algorithm for random forests was postulated by (Tin Kam Ho, 1995) implementing the random subspace method, and an extension of this algorithm was developed by (Breiman, 2001) around twenty years ago. This technique is among the most successful and effective existing supervised methods (Wang et al., 2018). In principle, this method gathers many decisions tree analyses (Kombo et al., 2020). Besides, it doesn't require any specific assumptions for the statistical distribution of the given data. The aim of this method is to collect the computed outputs from multiple trees, which weaken the overfitting problem in the ultimate model. The prediction function f(X) in RFR can be expressed as below (Hastie et al., 2009):

$$f(x)=\frac{1}{J}\sum_{j=1}^{J} H_j(X) \quad (2)$$

Where H(X) is the base learners, and j^{th} is a tree.

Hybrid KNNR-RFR

Although KNNR and RFR are considered effective and popular methods for prediction, they do not perform efficiently when there is a tiny amount of data. Hence, a hybrid model which a combination of both methods can be built to overcome this issue. The hybrid model equation can be expressed as below (Kombo et al., 2020):

$$\mu(X)=\frac{1}{N} \sum_{n=1}^{N} \omega_n P_n(X) \quad (3)$$

Where μ(X) is the weighted average result of the model, ω_n is the weight assigned to the n^{th} regressor, P_n is the prediction related to the nth model, and X is the sample data.

C-A fractal model

Cheng et al. (Cheng et al., 1994) proposed a C-A fractal model to delineate geochemical anomalies from the background. This method describes the relationships between the elemental concentrations and the geological data, based on the amount of area that each specific concentration occupies in the study area; keeping in mind that by an increase in element concentration, the occupied area decreases (Cheng et al., 1994; Afzal et al., 2010, 2012; Saein & Afzal, 2017; Ahmadfaraj et al., 2019). This model has the general form of relation one as shown below (Cheng et al., 1994):

$$A(\rho \leq \vartheta) \infty \rho^{-a_1} ; A(\rho \geq \vartheta) \infty \rho^{-a_2} \quad (4)$$

where $A(\rho\leq\vartheta)$ and $A(\rho\geq\vartheta)$ denote the area with concentration values ρ that are respectively, more minor, and more significant than contour value ρ defining areas υ, which represents threshold value, and a_1 and a_2 are characteristic exponents for both criteria. The area A (ρ) for a given ρ is equal to the size of cells with grade levels higher than ρ multiplied in the number of cells. The average concentration value is used for cells with more than one sample. Fractures between straight-line segments on the concentration-area log-log plot and the corresponding values of ρ are used as thresholds for separating geochemical values among various components.

Results and discussion

All the models have been built using Python programming language version 3.9.0. In this study. Cell dimensions determination has an essential role in mineral exploration; therefore, cell sizes have been considered 40 m×40 m respectively for X and Y (Rezaie & Afzal, 2016). In this study, the KNN-RF model was employed, and the results were compared with two other methods. The developed models have been utilized for Pb and Zn ore grade estimation in the IMD, Central Iran.

In these models, 80% and 20% of the data have been considered for the model training and testing, respectively. Optimal hyperparameters are required to make a robust model, greatly depending on the specific dataset (Schratz et al., 2019). The utilized hyperparameters are listed in Table 1. Once the hyperparameters were selected, the required models were made, and the validation dataset can evaluate the models' performance.

The obtained results based on the correlation coefficients of each model have been presented in Table 2. The hybrid model has been improved the correlation for both Pb and Zn grade estimation. Accordingly, the results obtained through the KNNR-RFR model have been selected for the C-A fractal modeling. Based on the C-A fractal modeling, there are seven populations for Pb and Zn. Major elemental anomalies commence from 0.2% and 0.24% for Pb and Zn, respectively.

Table 1. The Hyperparameters for the applied ML models to estimate of Pb and Zn.

Model	Hyperparameters	Search Range	Model Parameters
KNNR	N_neighbor	1-20	5
	Leaf_size	1-30	10
	Metric	'euclidean' 'manhattan' 'chebyshev' 'minkowski'	'euclidean'
RFR	N_estimator	100-1000	400
	Max_depth	0-20	10
	Max_features	'auto' 'sqrt' 'log2'	log2

Table 2. Comparison of the between estimated and raw data based on the ML models by the correlation coefficient

Correlation Coefficients	KNNR	RFR	Hybrid KNNR-RFR
Pb	+0.59	+0.64	+0.66
Zn	+0.48	+0.51	+0.54

Validation

The elemental distribution maps were created for Pb, and Zn based on the IDW method, as depicted in Figs. 2 and 3. Major Zn and Pb anomalies commence from 0.31% and 0.30%, respectively (Figs. 2 and 3). The main Pb anomalies exist in this area's central-northwestern and southern parts, as shown in Fig. 2. In addition, high extensive anomalous parts of Zn occurred in the NW, central and eastern in the IMD (Fig. 3).

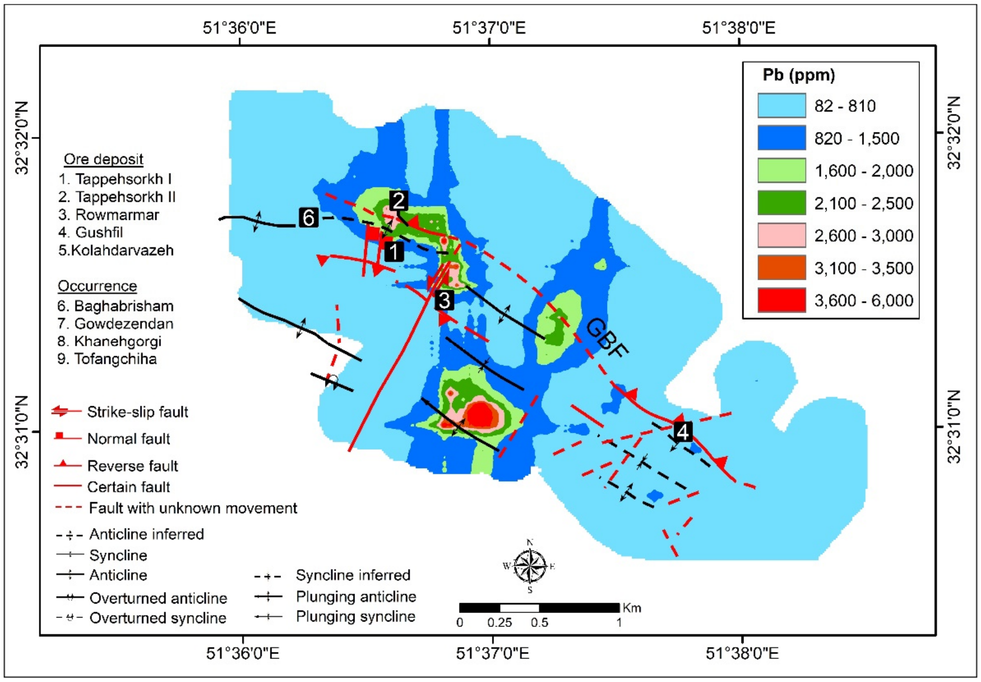

Figure 2. The geochemical distribution for Pb in the IMD

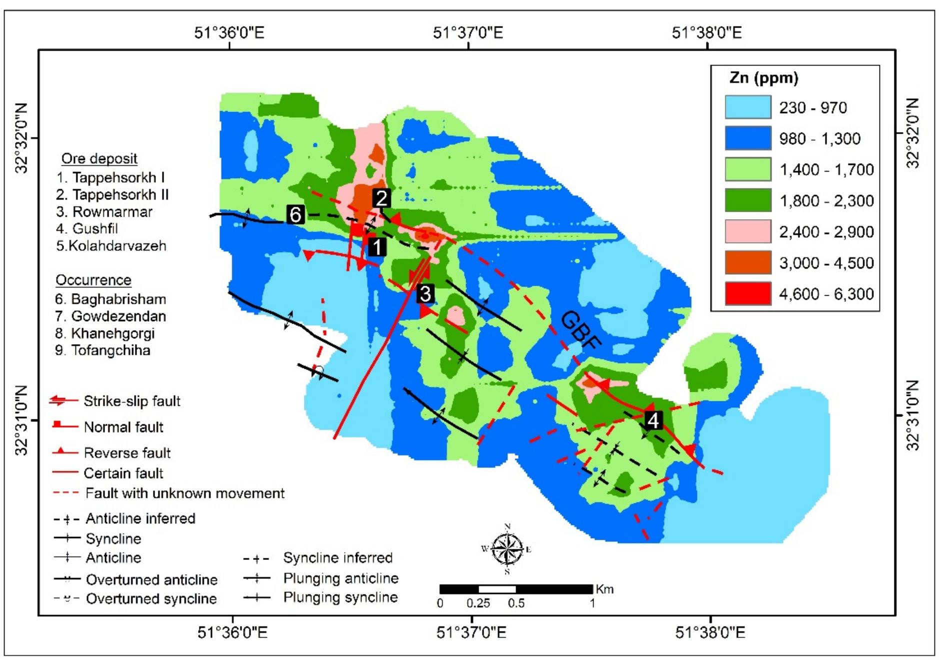

Figure 3. The geochemical distribution for Zn in the IMD

The results derived via hybrid machine learning methods correlated with the location of mining activities and borehole Pb-Zn mean values. The Pb anomalies in the IMD (Fig. 2) are situated on the sulfide orebodies from the Tappehsorkh to Rowmaramr mines. Another Pb anomaly corresponds with the western part of the Gushfil deposit, which was mined as Gushfil-e-Bala already. The Zn anomalies are correlated with the Tappehsorkh and Rowmarmar active mines (Fig. 3). Another anomaly of Pb is in the western part of the Gushfil deposit. In addition, major anomalies of the Pb and Zn are situated in the K3d carbonate unit as dolomitized orbitolina limestone and K1 siltstone and sandstone unit (Fig. 2).

Conclusion

Accurate ore grade estimation is essential in mineral exploration, and it directly affects the decision-making process. In this study, the combination of two machine learning algorithms, including KNNR and RFR, as a hybrid model was utilized to determine if the model accuracy will be enhanced. The advantage of the proposed model is that it does not require any complex assumptions or advanced mathematical and statistical knowledge. Even though the single regression models can solve the related problems, the present hybrid model can be considered a novel technique for the ore grade estimation with higher accuracy.

References

Afzal, P., Ahmadi, K., Rahbar, K., 2017. Application of fractal-wavelet analysis for separation of geochemical anomalies. Journal of African Earth Sciences, 128: 27-36.

Afzal, P., Khakzad, A., Moarefvand, P., Omran, N. R., Esfandiari, B., Alghalandis, Y. F., 2010. Geochemical anomaly separation by multifractal modeling in Kahang (Gor Gor) porphyry system, Central Iran. Journal of Geochemical Exploration, 104(1-2): 34-46.

Afzal, P., Zia Zarifi, A., Farhadi, S., Wetherelt, A., Yasrebi, A. B., 2012. Separation of Uranium Anomalies based on Geophysical Airborne Analysis by using Concentration-Area (C-A) Fractal Model,Mahneshan 1:50000 Sheet, NW IRAN. Journal of Mining and Metallurgy, 48 (1): 1-11.

Aghanabati, A., 2004. Geology of Iran.

Ahmadfaraj, M., Mirmohammadi, M., Afzal, P., Yasrebi, A. B., Carranza, E. J., 2019. Fractal modeling and fry analysis of the relationship between structures and Cu mineralization in Saveh region, Central Iran. Ore Geology Reviews, 107: 172-185.

Alavi, M., 1994. Tectonics of the Zagros orogenic belt of Iran: New data and interpretations. Tectonophysics, 229(3-4): 211-238.

Aliyari, F., Afzal, P., Lotfi, M., Shokri, S., Feizi, H., 2020. Delineation of geochemical haloes using the developed zonality index model by multivariate and fractal analysis in the Cu-Mo porphyry deposits. Applied Geochemistry, 121: 104694.

Biau, G., Scornet, E., Welbl, J., 2019. Neural Random Forests. Sankhya A, 81(2): 347-386.

Boveiri Konari, M., Rastad, E., 2018. Nature and origin of dolomitization associated with sulphide mineralization: New insights from the Tappehsorkh Zn-Pb (-Ag-Ba) deposit, Irankuh Mining District, Iran. Geological Journal, 53(1): 1-21.

Boveiri Konari, M., Rastad, E., Peter, J. M., 2017. A sub-seafloor hydrothermal syn-sedimentary to early diagenetic origin for the Gushfil Zn-Pb-(Ag-Ba) deposi. Mineralogy and Geochemistry Journal, 194 (1): 61-90.

Boveiri Konari, M., Rastad, E., Peter, J. M., Choulet, F., Kalender, L., Nakini, A., 2020. Sulfide ore facies, fluid inclusion and sulfur isotope characteristics of the Tappehsorkh Zn-Pb (± Ag-Ba) deposit, South Esfahan, Iran. Geochemistry, 80 (1):1256600.

Breiman, L., 2001. Random forests. Machine Learning, 45 (1): 5-32.

Burkov, A., 2019. The hundred-page machine learning book. Andriy Burkov.

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 (1): 35-50.

Chen, J., Lau, H. Y. K., 2016. Learning the inverse kinematics of tendon-driven soft manipulators with K-nearest Neighbors Regression and Gaussian Mixture Regression. 2016 2nd International Conference on Control, Automation and Robotics (ICCAR), 103-107.

Chen, R., Paschalidis, I., 2019. Selecting Optimal Decisions via Distributionally Robust Nearest-Neighbor Regression. 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada., 11.

Cheng, Q., Agterberg, F. P., Ballantyne, S. B., 1994. The separation of geochemical anomalies from background by fractal methods. Journal of Geochemical Exploration, 51 (2): 109-130.

Cunxiao, M., Qi, D., Min, H., Chune, W., Jianguo, C., 2021. Magnetic Anomaly Detection Based on Fast Convergence Wavelet Artificial Neural Network in the Aeromagnetic Field. Measurement, 109097.

Dramsch, J. S., 2020. 70 years of machine learning in geoscience in review. In Advances in Geophysics (61: 1-55). Elsevier.

Farshid Koohzadi, Afzal, P., Jahani, D., Pourkermani, M., 2021. Geochemical exploration for Li in regional scale utilizing Staged Factor Analysis (SFA) and Spectrum-Area (S-A) fractal model in north central Iran. Iranian Journal of Earth Sciences, 13 (4).

Hastie, T., Tibshirani, R., Friedman, J. H., 2009. The elements of statistical learning: Data mining, inference, and prediction (2nd ed). Springer.

Heidari, S. M., Afzal, P., Ghaderi, M., Sadeghi, B., 2021. Detection of mineralization stages using zonality and multifractal modeling based on geological and geochemical data in the Au-(Cu) intrusion-related Gouzal-Bolagh deposit, NW Iran. Ore Geology Reviews, 139, 104561.

Kombo, O. H., Kumaran, S., Sheikh, Y. H., Bovim, A., Jayavel, K., 2020. Long-Term Groundwater Level Prediction Model Based on Hybrid KNN-RF Technique. Hydrology, 7 (3): 59.

Lin, N., Chen, Y., Liu, H., Liu, H., 2021. A Comparative Study of Machine Learning Models with Hyperparameter Optimization Algorithm for Mapping Mineral Prospectivity. Minerals, 11 (2): 159.

Meigoony, M. S., Afzal, P., Gholinejad, M., Yasrebi, A. B., Sadeghi, B., 2014. Delineation of geochemical anomalies using factor analysis and multifractal modeling based on stream sediments data in Sarajeh 1:100,000 sheet, Central Iran. Arabian Journal of Geosciences, 7(12): 5333-5343.

Muller, A. C., Guido, S., 2016. Introduction to machine learning with Python: A guide for data scientists. O'Reilly Media, Inc.

Murphy, K. P., 2012. Machine learning: A probabilistic perspective. MIT Press.

Pourgholam, M. M., Afzal, P., Yasrebi, A. B., Gholinejad, M., Wetherelt, A., 2021. Detection of geochemical anomalies using a fractal-wavelet model in Ipack area, Central Iran. Journal of Geochemical Exploration, 220, 106675.

Rezaie, M., Afzal, P., 2016. The effect of estimation methods on fractal modeling for anomalies' detection in the Irankuh area, Central Iran. Geopersia, 6(1): 105-116.

Richter-Laskowska, M., Trybek, P., Bednarczyk, P., Wawrzkiewicz-Jałowiecka, A., 2021. Application of Machine-Learning Methods to Recognize mitoBK Channels from Different Cell Types Based on the Experimental Patch-Clamp Results. International Journal of Molecular Sciences, 840(22).

Saein, L. D., Afzal, P., 2017. Correlation between Mo mineralization and faults using geostatistical and fractal modeling in porphyry deposits of Kerman Magmatic Belt, SE Iran. Journal of Geochemical Exploration, 181: 333-343.

Schratz, P., Muenchow, J., Iturritxa, E., Richter, J., Brenning, A., 2019. Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. Ecological Modelling, 406: 109-120.

Shahsavar, S., Jafari Rad, A., Afzal, P., Nezafati, N., 2020. Selection of Optimum Fractal Model for Detection of Stream Sediments Anomalies. Geopersia, Online First.

Shamseddin Meigooni, M., Lotfi, M., Afzal, P., Nezafati, N., Razi, M. K., 2021. Application of multivariate geostatistical simulation and fractal analysis for detection of rare-earth element geochemical anomalies in the Esfordi phosphate mine, Central Iran. Geochemistry: Exploration, Environment, Analysis, 21(2): geochem2020-035.

Song, Y., Liang, J., Lu, J., Zhao, X., 2017. An efficient instance selection algorithm for k nearest neighbor regression. Neurocomputing, 251: 26-34.

Tin Kam Ho., 1995. Random decision forests. Proceedings of 3rd International Conference on Document Analysis and Recognition, 1: 278-282.

Wang, Y., Xia, S.-T., Tang, Q., Wu, J., Zhu, X., 2018. A Novel Consistent Random Forest Framework: Bernoulli Random Forests. IEEE Transactions on Neural Networks and Learning Systems, 29 (8): 3510-3523.

Wang, Z., Zuo, R., Jing, L., 2020. Fusion of Geochemical and Remote-Sensing Data for Lithological Mapping Using Random Forest Metric Learning. Mathematical Geosciences.

Zaremotlagh, S., Hezarkhani, A., 2017. The use of decision tree induction and artificial neural networks for recognizing the geochemical distribution patterns of LREE in the Choghart deposit, Central Iran. Journal of African Earth Sciences, 128: 37-46.

Zhang, C., Zuo, R., Xiong, Y., 2021. Detection of the multivariate geochemical anomalies associated with mineralization using a deep convolutional neural network and a pixel-pair feature method. Applied Geochemistry, 130: 104994.

Zuo, R., Carranza, E. J. M., 2011. Support vector machine: A tool for mapping mineral prospectivity. Computers & Geosciences, 37 (12): 1967-1975.

Zuo, R., Wang, J., Xiong, Y., Wang, Z., 2021. The processing methods of geochemical exploration data: Past, present, and future. Applied Geochemistry, 132: 105072.