

Research Article

Estimation of Cu-Mo anomaly values using Mahalanobis distance separation method and three commonly used data mining methods; Case study: Zafarghand

Zahra Varmazyari¹, Seyyed Saeed Ghannadpour^{1*}, Homayoon Katibeh¹

1- Department of Mining Engineering, Amirkabir University of Technology (Tehran Polytechnic)

Keywords: Estimation, Anomaly, Separation, Mahalanobis, Zafarghand.

1-Introduction

The importance of diagnosing anomalous values from the background is not obscured, so several methods have been introduced. These methods, from simple to complicated approaches, are divided into structural and nonstructural ones following a general objective based on decreasing the error of distinguishing anomaly from background (Hezarkhani and Ghannadpour, 2015; Ghannadpour and Hezarkhani, 2016a,b). Mahalanobis's distance method is an effective and multivariate method for separating the anomaly from the background.

The Mahalanobis Distance (MD) method is one of the multivariate techniques in which anomalous areas are separated based on Mahalanobis distances. The covariance matrix quantifies the shape and size of multivariate data. So, the Mahalanobis distance is the well-known distance measured to separate anomalies, which considers the covariance matrix. Explained and discussed techniques that are based on MD and applied in various fields of chemometrics, such as in multivariate calibration, pattern recognition, and process control; also, some applications in data clustering, especially MD, have substantially drawn attention (Ghannadpour et al. 2017).

2-Material and methods

The present study discusses the performance of combining the separation method with three data mining methods: K-Nearest Neighbor (KNN) (Li et al. 2017), Bayes simple classifier, and convolutional neural network. After separating copper and molybdenum anomalous values of 177 samples obtained from surface sampling operations in the Zafarghand area using the Mahalanobis distances method, three data mining methods are used to predict these values for each random sample.

The Zafarghand copper exploration area is located in southeastern Ardestan, in central Iran, about 110 kilometers northeast of Isfahan. The deposit is situated in the west of Zafarghand, as part of the 1:100,000 Ardestan geological sheet, ranging from 52°23'55"E to 52°26'30"E longitude and 33°10'30"N to 33°11'52"N latitude (Fig. 1) (Alaminia et al. 2017).

3-Results and discussions

The values of Cu-Mo and coordinates of sampling points were introduced to STATISTICA software as a dataset. After calculating the values of Mahalanobis, the values of the samples were calculated according to the ascending order of the Mahalanobis distances. Then, anomalous samples were determined for copper and molybdenum in the studied area. In the next step, based on the obtained results along with the parameters of Cu-Mo, the longitude and latitude of the samples, and the weight of each sample (based on the results of Mahalanobis distances of 1 for the field samples and 2 for anomalous samples) were studied in the data mining methods. After running the algorithm, the error rate and its performance were also investigated. According to the outputs, no sample in the KNN has been wrongly identified. It should be

* Corresponding author: s.ghannadpour@aut.ac.ir

DOI: 10.22055/AAG.2024.44558.2401

Received: 2023-08-23

Accepted: 2024-01-05

noted that the number of improperly predicted samples for convolutional neural network and Bayes methods are reported as 2 and 3, respectively.

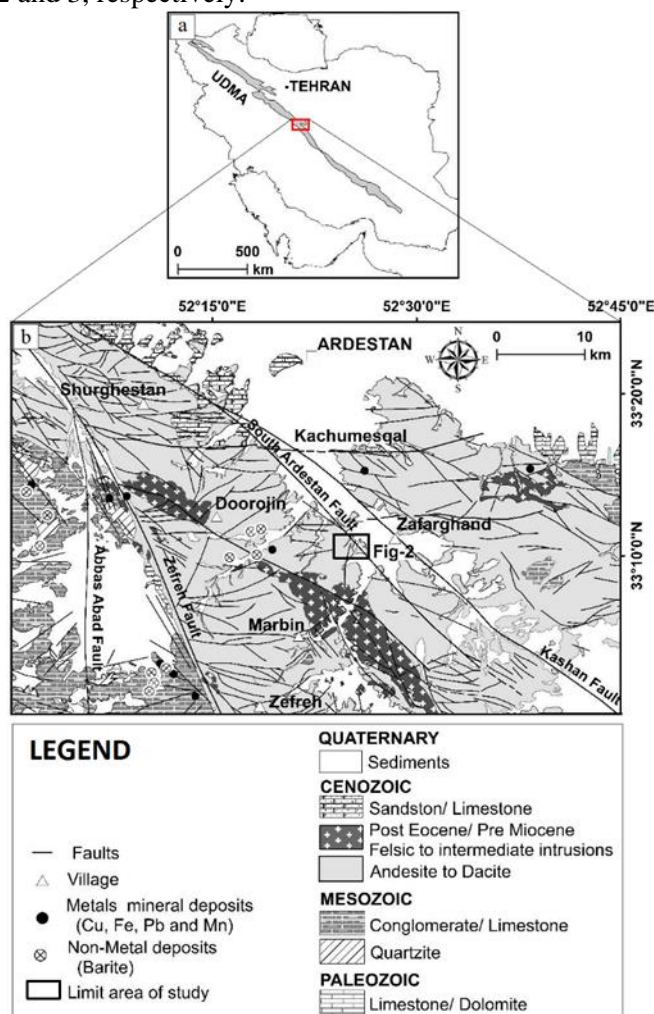


Fig. 1. Simplified geological map of the Zafarghand exploration area.

4-Conclusion

Considering the far more acceptable error rate for the network designed by the combination of the KNN method and Mahalanobis intervals, the said combination has been introduced to the decision-makers of this industry as a reliable and beneficial method to reach the most accurate predictions.

5-References

- Alamina, Z., Bagheri, H., Salehi, M., 2016. Geological, geophysical investigations and studies of fluids involved in Zafarghand exploration area (north-east of Isfahan province, Iran). *Economic Geology* 9(2), 295-312. <https://doi.org/10.22067/ECONG.V10I2.76528>.
- Ghannadpour, S.S., Hezarkhani, A. 2016a. Introducing 3D U-statistic method for separating anomaly from background in exploration geochemical data with associated software development. *Journal of Earth System Science* 125(2), 387–401. <https://doi.org/10.1007/s12040-016-0657-2>.
- Ghannadpour, S.S., Hezarkhani, A., 2016b. Exploration geochemistry data-application for anomaly separation based on discriminant function analysis in the Parkam porphyry system (Iran): *Geoscience Journal* 20(6), 837–850. <https://doi.org/10.1007/s12303-015-0064-8>.
- Ghannadpour, S.S., Hezarkhani, A., Roodpeyma, T., 2017. Combination of Separation Methods and Data Mining Techniques for Prediction of Anomalous Areas in Susanvar, Central Iran. *African Journal of Earth Sciences* 134, 516–525. <https://doi.org/10.1016/j.jafrearsci.2017.07.015>

- Hezarkhani, A., Ghannadpour, S.S., 2015. Exploration Information Analysis, first ed. Amirkabir University of Technology (Tehran Polytechnic) press, Tehran. (In Persian)
- Li, W., Kong, D., Wu, J., 2017. A novel hybrid model based on extreme learning machine, k-nearest neighbor regression and wavelet denoising applied to short-term electric load forecasting. *Energies* 10(5), 694. <https://doi.org/10.3390/en10050694>.

HOW TO CITE THIS ARTICLE:

Varmazyari, Z., Ghannadpour, S., Katibeh, H., 2024. Estimation of Cu-Mo anomaly values using Mahalanobis distance separation method and three commonly used data mining methods; Case study: Zafarghand. *Adv. Appl. Geol.* 14(2), 370-389.

DOI: 10.22055/AAG.2024.44558.2401

URL: https://aag.scu.ac.ir/article_19065.html

©2024 The author(s). This is an open access article distributed under the terms of the Creative Commons Attribution (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, as long as the original authors and source are cited. No permission is required from the authors or the publishers

تخمین مقادیر ناهنجاری مس-مولیبدن با استفاده از روش جدایش فواصل ماهالانوبیس و سه روش پر کاربرد داده کاوی؛ مطالعه موردی: ظفرقند

زهرا ورمزیاری

دانشکده مهندسی معدن دانشگاه صنعتی امیرکبیر

سید سعید قناده پور*

دانشکده مهندسی معدن دانشگاه صنعتی امیرکبیر

همایون کتیبه

دانشکده مهندسی معدن دانشگاه صنعتی امیرکبیر

*s.ghannadpour@aut.ac.ir

تاریخ دریافت: ۱۴۰۲/۰۶/۰۱ تاریخ پذیرش: ۱۴۰۲/۱۰/۱۵

چکیده

این پژوهش به منظور کاهش خطا در راستای صرف هزینه و انرژی به بررسی ترکیب روش‌های داده کاوی و جدایش ناهنجاری پرداخته است. اهمیت تشخیص مقادیر ناهنجار از زمینه بر هیچ یک پوشیده نیست، به این منظور روش‌های متعددی ابداع گشته است که از آن جمله می‌توان به روش جدایش فواصل ماهالانوبیس اشاره کرد که روشی مؤثر و چند متغیره در جدایش مقادیر ناهنجار از زمینه محسوب می‌شود. در مطالعه حاضر، به بررسی عملکرد ترکیب روش جدایش فوق با سه روش داده کاوی K- نزدیکترین همسایه، طبقه بند ساده بیز و شبکه عصبی کانولوشن پرداخته می‌شود، به این ترتیب که پس از جدایش مقادیر ناهنجار مس و مولیبدن در مورد ۱۷۷ نمونه حاصله از عملیات نمونه برداری سطحی در محدوده ظفرقند به کمک روش فواصل ماهالانوبیس، به منظور پیش‌بینی این مقادیر برای هر نمونه تصادفی، سه روش داده کاوی مذکور، مورد استفاده قرار گرفت. نتایج نشان می‌دهد که روش K- نزدیکترین همسایگی به مراتب قوی‌تر بوده، زیرا در شبکه طراحی شده توسط این روش، هیچ نمونه‌ای اشتباهاً شناسایی نشد که نشان دهنده دقت بالای شبکه طراحی شده است. لازم به ذکر است که تعداد نمونه‌های به اشتباه پیش‌بینی شده برای دو روش شبکه عصبی کانولوشن و بیز به ترتیب برابر با ۲ و ۳ عدد گزارش شده‌اند. با توجه به میزان خطای به مراتب قابل قبول‌تری برای شبکه طراحی شده توسط ترکیب روش K- نزدیکترین همسایگی و فواصل ماهالانوبیس، ترکیب مذکور به عنوان روشی قابل اطمینان و سودمند جهت رسیدن به صحیح‌ترین پیشگویی‌ها به تصمیم‌گیران این صنعت معرفی شده است.

واژه‌های کلیدی: تخمین، ناهنجاری، جدایش، ماهالانوبیس، ظفرقند.

۱- مقدمه

می‌توان پی برد و همچنین با فراهم آوردن مدارک معتبر نشان می‌دهند که آیا برنامه ما موفقیت آمیز بوده و به نوعی آگاهی دهنده در رابطه با محدودیت‌ها نیز می‌توانند باشند (Ghannadpour and Hezarkhani, 2015, Ghannadpour et al. 2015). قدرت روز افزون کامپیوترها و هزینه کم، با نیاز به تجزیه و تحلیل مجموعه اطلاعات زیاد، زمینه را جهت توسعه تکنیک‌های نو و بدیع بر اساس اکتشاف، راه حل‌های احتمالی فراهم می‌کند (Bax, 2000). تفکیک ناهنجاری‌های ژئوشیمیایی از زمینه همواره یکی از مشکلات اساسی در اکتشافات معدنی بوده است. امروزه ریاضیات به‌عنوان ابزاری قدرتمند در اختیار زمین‌شناسان قرار گرفته است که

مفهوم الگوریتم آزمون‌های موثری را برای رسیدن به عمق دانش و در ارتباط با هر موضوع داده شده، فراهم می‌کند؛ به این ترتیب مجموعه داده‌ها پس از کشف، مورد تجزیه و تحلیل قرار می‌گیرند. پس شناخت مناسب‌ترین روش برای آنالیز داده‌های در اختیارمان، چالشی بنیادین برای محققین می‌باشد، چرا که داده‌ها می‌توانند نشان دهند که آیا با هر تغییر قابل توجهی در متغیر وابسته می‌تواند امید به اثرگذاری باشد یا خیر. با استفاده از تجزیه داده‌ها، به روابط بین چندین فاکتور متعدد تأثیرگذار بر نتایج ارزیابی‌ها، پی برده می‌شود. با پردازش‌های کمی و کیفی، به تأثیرگذار بودن یا نبودن یک عملکرد و حتی علت آن

ژئوشیمیایی بر نقش اختلاط پوسته جوان و ماگما در کمان ارومیه-دختر با نگاه ویژه بر انکلاوهای میکروگرانولار مافیک و گرانیتوئیدهای ژنتیکی در مجموعه آذرین ظفرقند توسط Sarjoughian et al., (2018) تعیین گردید. در سال ۲۰۱۷ محدوده اکتشافی ظفرقند به عنوان مطالعه موردی به منظور شناسایی ذخایر معدنی عمیق و کور با استفاده از روش ضرایب فرکانس پیشنهادی جدید در حوزه فرکانسی داده‌های ژئوشیمیایی در نظر گرفته شد (Shahi et al., 2017).

۲- معرفی منطقه مورد مطالعه

۲-۱- زمین شناسی منطقه مورد مطالعه

منطقه اکتشافی مس ظفرقند در جنوب شرقی اردستان در ۱۱۰ کیلومتری شمال شرق اصفهان در ایران مرکزی واقع شده است. این کنسار در غرب ظفرقند به عنوان بخشی از برگه زمین‌شناسی ۱:۱۰۰۰۰۰ اردستان با جغرافیایی $52^{\circ}23'55''$ تا $52^{\circ}26'30''$ طول شرقی و $33^{\circ}11'52''$ تا $33^{\circ}10'30''$ عرض شمالی قرار دارد. منطقه مورد تحقیق بخشی از زون ساختاری ایران- مرکزی به‌شمار می‌رود و بطور دقیق‌تر در نوار آتشفشانی ارومیه-دختر قرار می‌گیرد. نوار تکتونو ماگمایی ارومیه-دختر یک کمان ماگمایی نوع آندی است که در اثر فرورانش پوسته اقیانوسی نئوتتیس به زیر لبه جنوبی ایران مرکزی در امتداد حاشیه قارهای ایران مرکزی، در طی کوهزایی آلپ شکل گرفته است. همچنین مطالعات انجام شده بر روی نوار ارومیه-دختر وجود زون فرورانش را در این منطقه تایید می‌کند (Barbarin and Bernard, 1990). بر اساس مطالعات مختلف انجام گرفته در نوار ارومیه-دختر در ائوسن فوران‌های اولیه، کالکوالکال و زیردریایی بوده و سنگ‌های ولکانیکی ارومیه-دختر طیف گسترده ترکیبی از بازالت، داسیت و کمر ریولیت همراه با توفها و آذرآواری‌ها را نشان می‌دهند. در ائوسن فوقانی، ترکیب گدازه‌ها به طرف آلکال و پراآلکال میل می‌کند. چنانکه در جنوب ناین سری پتاسیک شوشونیت آنالسیم‌دار و اطراف کاشان سری هیپرسدیک و در شمال شهر بابک سری پتاسیک شامل فنولیت لوسیت‌دار و بازانیت تشکیل گردیده‌اند. در اواخر ترشیری، گدازه‌های کالکوالکال به انواع ساب‌آلکال و آلکال تغییر و تحول یافته‌اند و در نتیجه گدازه‌های آلکال غنی از پتاسیم تشکیل شده است. در آتشفشانی‌های ائوسن ایران، اختلاط ماگماهای اسید و بازی یک پدیده فراگیر است که در تولید ماگمایی با ترکیب حدواسط گاه با ویژگی ژئوشیمیایی

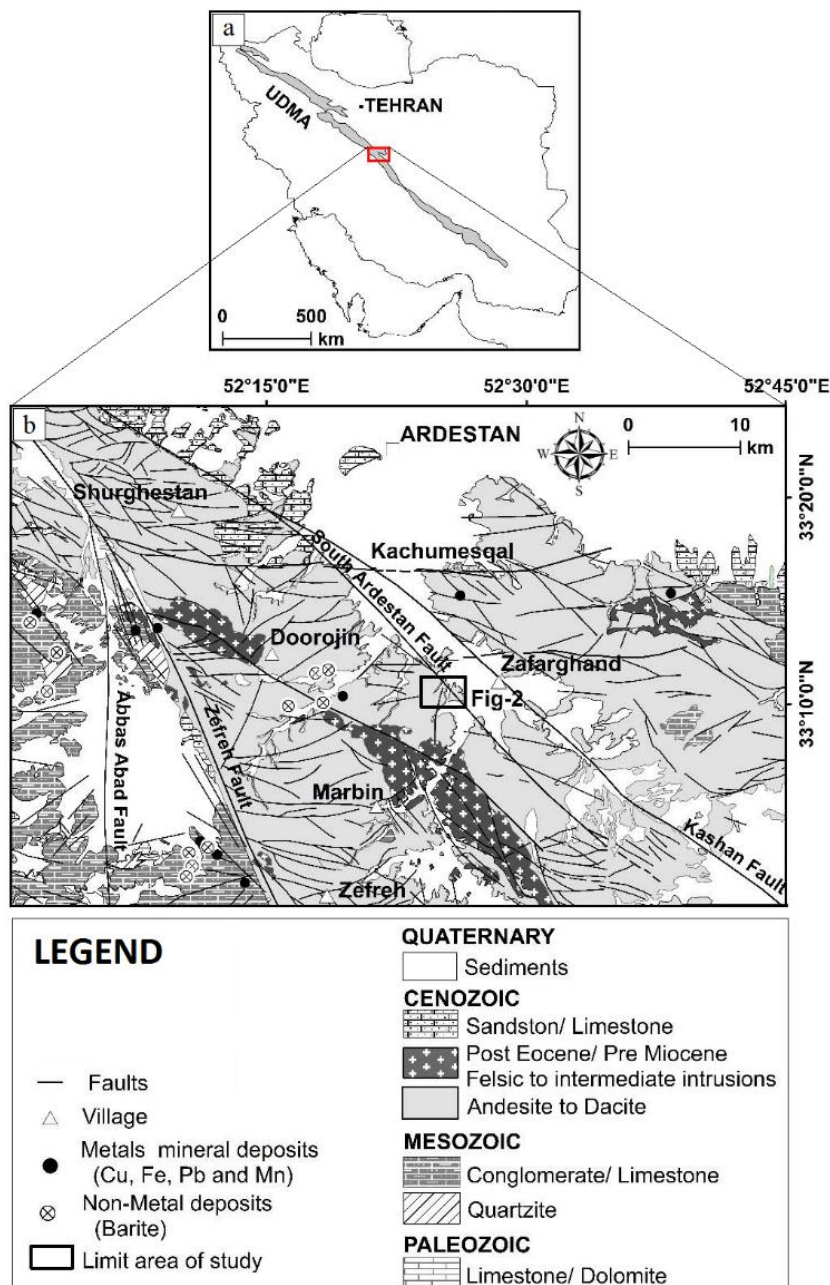
آنان را جهت توصیف فرآیندهای طبیعی یاری می‌دهد. تشخیص مقادیر ناهنجار در عملیات اکتشافی معدن و تعیین مناطق امیدبخش در برای رسیدن به کانی سازی اقتصادی در حوزه معدنکاری به خصوص اکتشافات ژئوشیمیایی بسیار حائز اهمیت است. بدین منظور جدایش ناهنجاری‌ها از مقادیر عادی و زمینه در دو شکل ساختاری و غیرساختاری صورت می‌گیرد. روش‌های آماری مختلفی برای جداسازی محدوده‌های ناهنجار توسعه یافته و توسط محققین ارائه شده است (Sinclair, 1991; Cheng, 1999; Ghannadpour and Hezarkhani, 2016a; b; Farhadi et al. 2022; Koohzadi et al. 2021; Shahbazi et al. 2021; Nabilou et al. 2022). روش جدایش بر اساس فواصل ماهالانویس از جمله روش‌های غیرساختاری محسوب می‌شود. در این روش مقادیر ناهنجار زمانی تعیین می‌شود که تمامی متغیرهای موردنظر، آن را تایید کرده باشد. هدف از این مطالعه، تخمین مقادیر ناهنجار عبار مس-مولیبدن در منطقه ظفرقند با سه روش داده کاوی K- نزدیکترین همسایه، طبقه بند ساده بیز و شبکه عصبی کانولوشن و بررسی عملکرد ترکیب روش‌های جدایش فوق می-باشد.

منطقه اکتشافی ظفرقند (در شمال شرقی اصفهان) در بخش میانی یک کمان ماگمایی نوع آندی واقع شده است (شکل ۱). کمر بند فلزایی تتیس از شرق اروپا (رومانی، اسلواکی و ترکیه) تا خاورمیانه (ایران، پاکستان و افغانستان) ادامه دارد. ایران به عنوان یکی از بزرگترین نواحی بر روی ۱۷۰۰ کیلومتر از این کمر بند واقع شده است و ذخایر بزرگی از مس پورفیری مانند سرچشمه را به خود اختصاص می‌دهد. بخشی از این کمر بند را کمان آتشفشانی-ماگمایی ارومیه-دختر شامل می‌شود. در سال‌های اخیر مطالعات بسیاری در مورد منطقه ظفرقند صورت گرفته که در ادامه به شرح مختصری از آن‌ها پرداخته خواهد شد.

در سال ۲۰۱۶ میلادی، کلریتی شدن بیوتیت به منظور بازسازی پارامترهای فیزیکی ژئوشیمیایی کانی‌سازی و دگرسانی مرتبط در سیستم مس پورفیری ظفرقند مورد تحلیل شیمی کانی‌ها و ایزوتوپ پایدار قرار گرفت (Aminroayaei Yamini et al., 2017). امین‌الرعایایی و همکاران در مطالعه دیگری به پتروگرافی، ژئوشیمی و دماسنجی دگرسانی در کانی‌سازی سینوژنیک مس در این ناحیه پرداختند (Aminroayaei Yamini et al., 2016). محدودیت‌ها و قیدهای ایزوتوپی و

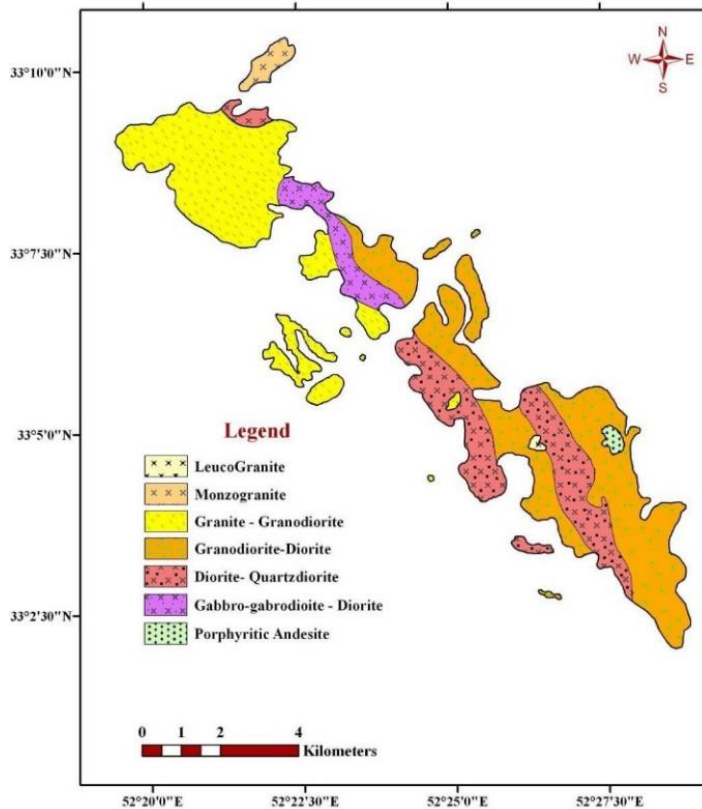
مس و مولیبدن در جنوب شرق اندیس اثبات شده است. با توجه به نقشه زمین شناسی اصلاح شده منطقه مورد مطالعه، واحدهای سنگی موجود در منطقه را می توان بر اساس سن زمین شناسی آن ها به صورت شکل ۲ تقسیم بندی کرد.

کلسیمی - قلیایی مؤثر می باشد. با توجه به مطالعات ژئوشیمیایی و بررسی هاله های تحت کانساری به فوق کانساری، مرکز اصلی کانه زایی در بخش شرقی و جنوب شرق منطقه قرار دارد. حضور بیشتر عناصر تحت کانساری مثل کبالت، نیکل،



شکل ۱- (a) نمایش موقعیت منطقه اکتشافی ظفرقند در بخش مرکزی کمان ارومیه-دختر و (b) بخشی از نقشه ساده شده زمین اردستان و شهرباب با مقیاس ۱:۱۰۰،۰۰۰ پس از اندکی تغییرات از Alaminia و همکاران (۲۰۱۷).

Fig. 1. (a) Showing the location of the Zafarghand exploration area in the central part of Urmia-Dokhtar Arc and (b) a part of the simplified map of Ardestan and Shahrab with a scale of 1:100,000 (after Alaminia et al., 2017 with some modifications).



شکل ۲- نقشه زمین شناسی اصلاح شده با توجه به لیتولوژی سنگ‌های منطقه مورد مطالعه.
 Fig. 2. Geological map modified according to the lithology of the rocks of the studied area.

سنگ‌های فلسیک (گرانودیوریت - گرانیت و مشتقات جانبی آن‌ها؛ ۴- دایک‌های آندزیتی سینپلوتونیک یا نسل دوم؛ ۵- تونالیت‌ها.

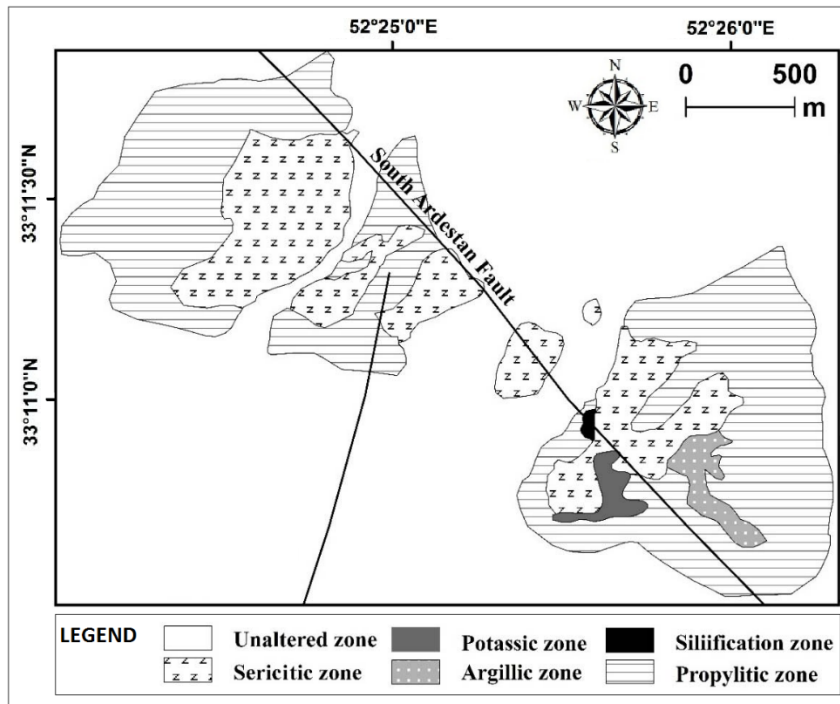
جایگزینی توده گرانیتوئیدی ظرفرند با دگرگونی مجاورتی کم وسعتی همراه بوده است که با توجه به ترکیب آندزیتی - داسیتی سنگ‌های میزبان، آثار دگرگونی مجاورتی در آن‌ها چندان بارز نیست. با این وصف، دوتریفیکاسیون یا شیشه‌زدایی و تشکیل فلندسپارهای نوظهور (در شمال روستای ماربین، اپیدوت‌زایی گسترده در سنگ‌های ایگنیمبریتی (با ترکیب کلی داسیتی) (در غرب تقی آباد) و تشکیل رگه‌هایی از گارنت‌های کلسیم‌دار (نوع آندرادیت -گروسولار) در سنگ‌های میزبان (جنوب مزرعه لامحمود) از شواهد بارز این نوع دگرگونی است.

بر اساس مطالعات و بررسی‌های صورت گرفته، انواع دگرسانی‌های منطقه شامل پتاسیک، فلیک، سیلیسی و آرژلیک و در نهایت پروپیلیتیک می‌باشند. این دگرسانی‌ها

توده گرانیتوئیدی ظرفرند در درون سنگ‌های آتشفشانی، آتشفشانی -رسوبی و آتشفشانی - تخریبی عمدتاً آندزیتی- داسیتی به سن ائوسن میانی تا بالایی نفوذ کرده است. با توجه به مطالعات صحرایی و میکروسکوپی، می‌توان ترکیب سنگ‌شناسی این توده را در هفت گروه یا واحد سنگی زیر رده‌بندی و معرفی کرد: ۱- دایک‌های آندزیتی نسل اول؛ ۲- گابرو - گابرودیوریت؛ ۳- دیوریت - کوارتزدیوریت؛ ۴- انکلاوهای میکروگرانولار مافیک؛ ۵- گرانودیوریت - گرانیت، ۶- تونالیت؛ ۷- دایک‌های آندزیتی نسل دوم (سینپلوتونیک). علیرغم دسته‌بندی سنگ‌های مورد مطالعه در هفت رده ذکر شده، بررسی‌های صحرایی نشان می‌دهد که می‌توان رده‌بندی کلیتری انجام داد؛ به طوری که در آن ترتیب جایگزینی واحدهای سنگی و ترکیب سنگ‌شناسی آن‌ها، به‌طور همزمان انعکاس پیدا کند. این رده‌بندی به‌صورت زیر است: ۱- دایک‌های آندزیتی پیش رس یا نسل اول؛ ۲- سنگ‌های مافیک - حد واسط (گابرو تا کوارتزدیوریت و مشتقات جانبی آن‌ها؛ ۳-

با فاصله گرفتن از فیلیک، دگرسانی‌های سیلیسی و آرژیلیک به صورت محدود جایگزین آنها شده است (شکل ۳). دگرسانی پروپیلیتیک در سنگ‌های محدوده بیرونی کانسار گسترش دارد (شکل ۳) (Alaminia et al., 2016). شایان ذکر است که کلریت در سطح وسیعی از منطقه جنوب اردستان به چشم می‌خورد.

اغلب در نزدیک توده‌های نفوذی و سنگ‌های میزبان اطراف گسل جنوب اردستان دیده می‌شود (شکل ۳). در گوشه جنوب شرقی محدوده، دگرسانی پتاسیک با رخنمون بسیار محدود در نزدیکی توده کوارتز دیوریت و دگرسانی فیلیک در نزدیکی توده‌های نفوذی دیوریت و کوارتز دیوریت پورفیری و نیز سنگ‌های آتشفشانی داسیتی و ریوداسیتی قابل مشاهده است.



شکل ۳- گسترش هاله‌های دگرسانی در محدوده اکتشافی ظرفرند. پس از اندکی تغییرات از Alaminia و همکاران (۲۰۱۷).

Fig. 3. Expansion of alteration halos in Zafarqand exploration range (after Alaminia et al., 2017 with some modifications).

(رابطه ۱)

$$D^2 = ([X] - [\bar{X}])[S]^{-1}([X] - [\bar{X}])^T$$

که در آن $[X]$ بردار داده‌ها برای نمونه‌های مورد نظر، $[\bar{X}]$ بردار میانگین کل داده‌ها و $[S]^{-1}$ ماتریس کوواریانس است. همانطور که مشاهده می‌شود فاصله مورد نظر معادل کسر بردار میانگین داده‌ها از بردار هر یک از داده‌ها تقسیم بر ماتریس $[S]^{-1}$ می‌باشد. توزیع فواصل ماهالانوبیس برای یک جامعه نرمال چند متغیره از یک توزیع کای-اسکور با درجه آزادی معادل تعداد متغیرها پیروی می‌کند. برای تشخیص مقادیر ناهنجار می‌بایست نمودار فواصل ماهالانوبیس بر حسب مقادیر χ^2 آنها رسم شود. در ادامه به بررسی و تخمین مقادیر ناهنجار مس بر اساس ترکیب این روش و روش‌های مذکور پرداخته شده است.

۳- مواد و روش‌ها

۳-۱- روش جدایش مقادیر ناهنجار بر اساس فواصل

ماهالانوبیس

این روش یک روش چند متغیره است که در آن بر اساس فواصل ماهالانوبیس ناهنجاری‌ها مشخص می‌شوند. در فضای n بعدی فاصله ماهالانوبیس (D^2) مشابه مقادیر استاندارد شده Z برای یک متغیر می‌باشد. در واقع فرمول محاسبه فواصل ماهالانوبیس شبیه استاندارد کردن داده‌های تک متغیره است. روش فواصل ماهالانوبیس بر اساس مقادیر تمام عناصر آنالیز شده در هر نمونه اقدام به جدایش نمونه‌های ناهنجار می‌نماید اما در انتها مشخص نمی‌شود که هر عنصر چه نقشی در ناهنجار شدن نمونه داشته است. این روش به صورت زیر بیان می‌شود (Filzmoser et al., 2005; Moeini and Torab, 2017):

۲-۳- ساختار مدل شبیه سازی با استفاده از الگوریتم K - نزدیکترین همسایه

نزدیکترین همسایه یک الگوریتم یادگیری می باشد که در روش بازشناسی الگو طی چندین دهه مطالعه شده است (Yang and Liu, 1999)، مدل KNN یکی از کارآمدترین و ساده ترین الگوریتم های یادگیری ماشین است. این تکنیک از زمان کاور و هارت در مدل های رگرسیون و طبقه بندی مختلف استفاده شده است. نزدیکترین همسایگی یک روش تشخیص الگوی آماری بدون متغیر است که برای الگوی مورد نظر، K الگوی مشابه به نام نزدیکترین همسایه ها را می یابد. هدف اصلی این روش به دست آوردن بهترین برآوردگرها، با استفاده از همسایگی بین متغیر مستقل پیش بینی کننده ها و محاسبه این متغیر در یک مجموعه داده است (Li et al., 2017). در این روش ابتدا توسط بردار مشخصه F یک ساختار وابستگی برای متغیرهای مستقل و وابسته تعریف می شود. به بیان دیگر بردار مشخصه نشان می دهد که متغیر مورد پیش بینی، Rt به چه پارامترهایی بستگی دارد. با داشتن بردار مشخصه (داده های آزمون مدل)، K نزدیکترین بردارهای مشخصه در داده های آموزش مدل با فاصله اقلیدسی وزن دار شده، یافته می شوند. مقادیر ناهنجار مس-مولییدن منطقه مورد نظر (متغیر در حال پیش بینی) از K نزدیکترین همسایه ها با استفاده از رگرسیون وزن دهی شده که از تابع کرنل استفاده می کند، برآورد می شود. بطور کلی در این روش یک داده آموزشی برای دسته بندی وجود دارد، الگوریتم K همسایه نزدیک در میان داده های آموزشی پیش دسته بندی شده، بر اساس یک معیار شباهت پیدا کرده و دسته های این K همسایه نزدیک برای پیش بینی دسته داده آزمایشی به وسیله امتیازدهی داده های هر دسته منتخب، استفاده می شود. اگر بیشتر از یک همسایه به دسته های مشابه تعلق داشته باشد، مجموع امتیاز آن ها به عنوان وزن آن دسته استفاده می شود و دسته با بالاترین امتیاز به داده مورد آزمایش انتساب می یابد، که اگر از یک مقدار آستانه تجاوز کند، بیشتر از یک دسته می تواند به سند آزمایشی انتساب یابد. یک مشکل در روش KNN، تعیین مقدار K می باشد و برای تعیین آن باید یک سری از آزمایشات با مقادیر مختلف K انجام شود، تا بهترین مقدار برای K را تعیین کند. مقدار بالا و پایین K میتواند به ترتیب منجر به overfitting و noise شود (Deng, 2020).

عیب دیگر KNN پیچیدگی زمانی محاسباتی مورد نیاز برای پیمایش همه داده های آموزشی می باشد.

۳-۳- ساختار مدل با روش طبقه بندی ساده بیز

در روش بسیار کاربردی یادگیری بیز روش یادگیرنده ساده بیزی ست که عموماً روش طبقه بندی ساده بیز نامیده می شود. در برخی زمینه ها نشان داده شده است که کارایی آن قابل قیاس با کارایی روش هایی مانند شبکه عصبی ست. طبقه بندی ساده بیزی برای مسائلی که هر نمونه x در آن توسط مجموعه ای از مقادیر صفات و تابع هدف f(x) از مجموعه ای مانند v انتخاب می شود، کاربرد دارد. روش بیزی برای طبقه بندی نمونه جدید این است که محتمل ترین طبقه یا مقدار هدف vmap را با داشتن مقادیر صفات {a1, a2, ..., an} که توصیف کننده نمونه جدید است شناسایی کند.

(رابطه ۲)

$$v_{MAP} = \operatorname{argmax} P(v_j | a_1, a_2, \dots, a_n), v_j \in V$$

با استفاده از قضیه بیز می توان عیارت بالا را به صورت زیر

بازنویسی کرد،

(رابطه ۳)

$$v_{MAP} = \operatorname{argmax} \frac{P(a_1, a_2, \dots, a_n | v_j)}{P(a_1, a_2, \dots, a_n)}, v_j \in V$$

(رابطه ۴)

$$v_{MAP} = \operatorname{argmax} P(a_1, a_2, \dots, a_n | v_j) P(v_j), v_j \in V$$

حال با استفاده از داده های آموزشی سعی می کنیم دو جمله معادله بالا را تخمین بزنیم. محاسبه از روی داده های آموزشی به این صورت که میزان تکرار v_j در داده ها چقدر است، آسان است. اما محاسبه جملات مختلف $P(a_1, a_2, \dots, a_n | v_j)$ به این صورت قابل قبول نخواهد بود مگر اینکه حجم بسیار زیادی از داده های آموزشی در اختیار داشته باشیم. مشکل اینجاست که تعداد این جملات برابر تعداد نمونه های ممکن ضرب در تعداد مقادیر تابع هدف است. بنابراین باید هر نمونه را چندین بار مشاهده کنیم تا تخمین مناسبی از آن به دست آید. فرض روش طبقه بندی ساده بیز بر اساس این ساده سازی است که مقادیر صفات با داشتن مقادیر تابع هدف از یکدیگر مستقل شرطی باشند. به عبارت دیگر، این فرض بیانگر این است که به شرط مشاهده خروجی تابع هدف احتمال مشاهده صفات a_1, a_2, \dots, a_n برابر ضرب احتمالات هر صفت به طور

هر شبکه عصبی از تعدادی نرون و وزن‌های ارتباطی که نرون‌ها را به هم وصل می‌کنند، تشکیل شده است. از اجتماع چند نرون که به طور موازی عمل کنند یک لایه به وجود می‌آید. شبکه‌های عصبی اغلب شامل یک لایه ورودی، یک یا چند لایه میانی و یک لایه خروجی‌اند. لایه یا لایه‌های میانی، اطلاعات دریافت شده از لایه ورودی را پردازش می‌کنند و در اختیار لایه خروجی قرار می‌دهند، شکل ۴-الف ساختار شبکه عصبی با یک لایه میانی و شکل ۴-ب مدل ریاضی یک نرون برای پردازش اطلاعات در شبکه عصبی را نشان می‌دهد که روابط ریاضی به صورت روابط زیر قابل ارایه است.

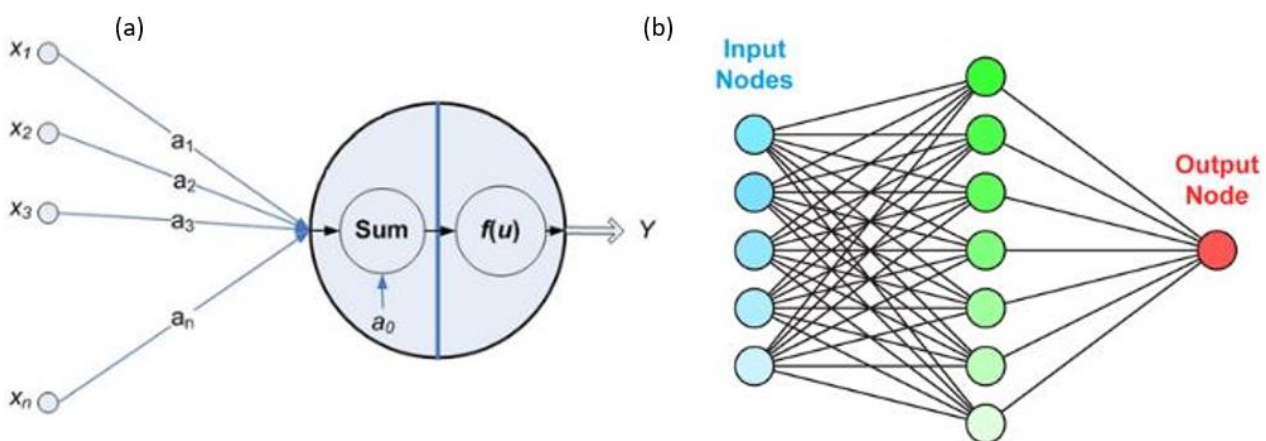
(رابطه ۶)

$$U = \sum_{i=1}^n x_i a_i + a_0$$

(رابطه ۷)

$$Y = f(u)$$

متغیرهای ورودی (X_1, \dots, X_n) در وزن‌های متناظرشان (a_1, \dots, a_n) ضرب و به همراه مقدار بایاس (a_0) در متغیری جدید (U) ذخیره می‌شوند. متغیر جدید (U) از یک تابع فعال‌سازی (f) عبور کرده و مقدار خروجی نرون (Y) را مشخص می‌کند (Kavzoglu, 2009; Beucher et al., 2013).



شکل ۴- (a) ساختار شبکه عصبی با یک لایه پنهان و (b) ساختار یک نرون مورد استفاده در شبکه عصبی (Beucher et al., 2013).

Fig. 4. (a) The structure of a neural network with a hidden layer and (b) The structure of a neuron used in a neural network (Beucher et al., 2013).

جداگانه باشد. اگر این را جایگزین معادله بالا کنیم روش طبقه‌بندی ساده بیز را نتیجه می‌دهد،
(رابطه ۵)

$$v_{NB} = \operatorname{argmax}_{v_j} P(v_j) \prod P(a_i | v_j), v_j \in V$$

که v_{NB} خروجی طبقه‌بندی ساده بیز برای تابع هدف است. توجه کنید که تعداد جملات $P(a_i | v_j)$ که در این روش باید محاسبه شوند برابر تعداد صفات ضرب در تعداد دسته‌های خروجی برای تابع هدف است که این مقدار از تعداد جملات $P(a_1, a_2, \dots, a_n | v_j)$ بسیار کمتر است (Ghannadpour and Hezarkhani, 2017).

۳-۴- شبکه عصبی کانولوشن

امروزه شبکه‌های عصبی مصنوعی به عنوان زیر مجموعه‌ای از تکنیک‌های هوش مصنوعی، با عملکردی مانند مجموعه نرون‌های مغز انسان انواع روش‌های قدرتمند برای حل مشکلات در بسیاری مسایل از قبیل پیش‌بینی، بهینه‌سازی، کلاسه‌بندی، تجزیه و تحلیل داده، شناسایی الگو، کنترل و موارد دیگر را فراهم کرده است از مهمترین ویژگی‌های شبکه‌های عصبی مصنوعی، وابسته نبودن به فرضیه‌های اولیه در مورد داده‌های ورودی است. داده‌های ورودی می‌توانند هرگونه توزیع آماری دلخواهی داشته باشند. این امتیاز ویژه آن‌ها در مقابل روش‌های آماری است و این امکان را به آنها می‌دهد که به طور یکسان از انواع مختلف داده‌های ورودی با هر توزیع دلخواه استفاده کنند.

پس انتشار می‌یابد. شبکه عصبی کانولوشن به مقدار زیادی داده اول لایه جهت آموزش نیاز ندارد و آموزش چنین مجموعه داده‌هایی در مقیاس بزرگ بسیار زمان بر خواهد بود (Liu et al., 2019).

۴- پردازش داده‌ها و نتایج

در این بخش اقدام به پردازش آنالیز داده‌های شیمیایی نموده سپس به جدایش مقادیر ناهنجار مس-مولیبدن به کمک روش فواصل ماهالانویس می‌شود. پس از آن با روش‌های داده کاوی به پیش بینی نمونه‌های ناهنجار در منطقه مورد مطالعه پرداخته می‌شود. در این مطالعه ۱۷۷ نمونه سنگی از منطقه مورد مطالعه جمع آوری شد، نمونه‌های برداشت شده در آزمایشگاه با روش طیف سنجی پلاسمای جفت شده القایی مورد آنالیز ۴۳ عنصر قرار گرفت. عناصر آنالیز شده به همراه حد تشخیص شان در جدول ۱ آمده است.

برای بررسی داده‌ها، اکثراً لازم است که دامنه تغییرات داده‌ها مشابه یکدیگر باشند که در اصطلاح، این عمل را نرمال‌سازی داده‌ها می‌نامند. نرمال‌سازی داده‌های مس-مولیبدن در این مطالعه توسط رابطه زیر انجام گرفته است.

$$P_n = \text{Log}(P_i) \quad (\text{رابطه ۸})$$

که در آن: P_n داده نرمال شده، P_i داده حقیقی هستند.

پس از استفاده از رابطه فوق مشخص شد داده‌های مس-مولیبدن دارای توزیع نرمال بوده (شکل ۵ و ۶). جدول ۱ و ۲ به ترتیب مشخصات آماری قبل از لگاریتم و پس از آن آمده است.

شبکه عصبی کانولوشن (به اختصار CNN یا ConvNet) نوعی از شبکه عصبی مصنوعی است که در آن نورون‌ها، به نواحی روی هم افتاده در یک ناحیه دیداری واکنش نشان می‌دهند. این نوع شبکه‌ها از فرایندهای بیولوژیکی الهام گرفته شده‌اند و گونه‌هایی از شبکه‌های پرسپترون چند لایه‌اند که با طراحی خاصی از حداقل میزان پیش پردازش بهره می‌برند. یادگیری عمیق و شبکه عصبی کانولوشن عمیق به عنوان یکی از پرکاربردترین الگوریتم‌های آن، بصورت گسترده در اکتشاف استفاده می‌شود. طرح اصلی یادگیری عمیق و اجرای آن در علوم مختلف توسط پروفسور هینتون در سال 2006 مطرح شد. در شبکه‌های عصبی کانولوشن از یک واحد خطی اصلاح شده برای حذف مقادیر منفی از تابع فعال‌سازی و جلوگیری از مشکلات Over fitting استفاده می‌گردد. این فرآیند همواره به دنبال یک کانولوشن می‌باشد. عملیات حداکثر پولینگ برای ساده‌سازی خروجی با انجام نمونه‌گیری غیرخطی و کاهش تعداد پارامترهایی که شبکه برای یادگیری نیاز دارد، استفاده می‌گردد. پس از آن لایه کاملاً متصل برداری با ابعاد k (تعداد کلاس‌هایی که شبکه قادر به کلاس‌بندی آنها می‌باشد) را استخراج می‌کند. هر نرون از سطح ویژگی استخراج شده در لایه کانولوشن به صورت محلی به ورودی آن متصل می‌باشد و مجموع وزنی با وزن اتصال موجود و مقدار جبرانی به ورودی محلی اضافه می‌گردد. در فرآیند آموزش، الگوریتم گرادینان نزولی عموماً به عنوان تابع زیان انتخاب می‌شود. مقدار باقی مانده به وسیله گرادینان نزولی برای آپدیت مقادیر آموزشی وزن‌ها (w) و بایاس (b) در هر لایه CNN به صورت لایه به لایه

جدول ۱- مشخصات آماری عناصر قبل از لگاریتم‌گیری

Table 1. Statistical characteristics of the elements before the logarithm

Element	Average value (ppm)	Variance value (ppm)	Maximum value (ppm)
Copper	91.6	28578	1040
Molybdenum	6.08	138	81

جدول ۲- مشخصات آماری عناصر پس از لگاریتم‌گیری

Table 2. Statistical characteristics of the elements after the logarithm

Element	Average value (ppm)	Variance value (ppm)	Maximum value (ppm)
Copper	1.44	0.49	3
Molybdenum	0.36	0.27	1.9

همچنین در این مقاله برای ارزیابی عملکرد مدل، از شاخص آماری رابطه ۹ استفاده شده است.

$$MSE = \frac{1}{N} \sum_{i=1}^N (y - y') \quad (\text{رابطه ۹})$$

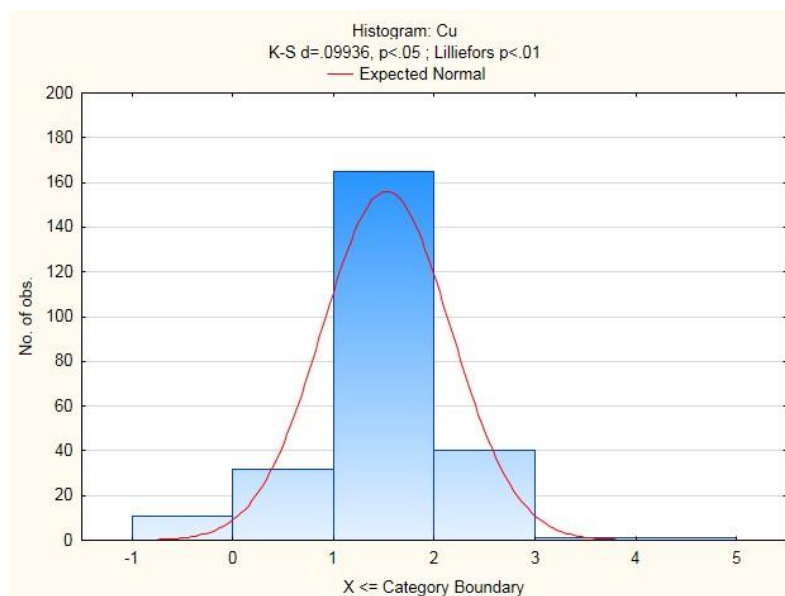
که در این روابط y مقدار اندازه‌گیری شده و y' مقدار پیش‌بینی شده می‌باشد. لازم به توضیح است که معیار میانگین مربعات خطا (MSE) بیانگر خطای بین مقادیر اندازه‌گیری شده و پیش‌بینی شده می‌باشد و هر چه مقدار آن کمتر باشد عملکرد مدل از اطمینان بیشتری برخوردار است.

جدول ۳- عناصر آنالیز شده به همراه حد تشخیص

Table 3. Analyzed elements with detection limit

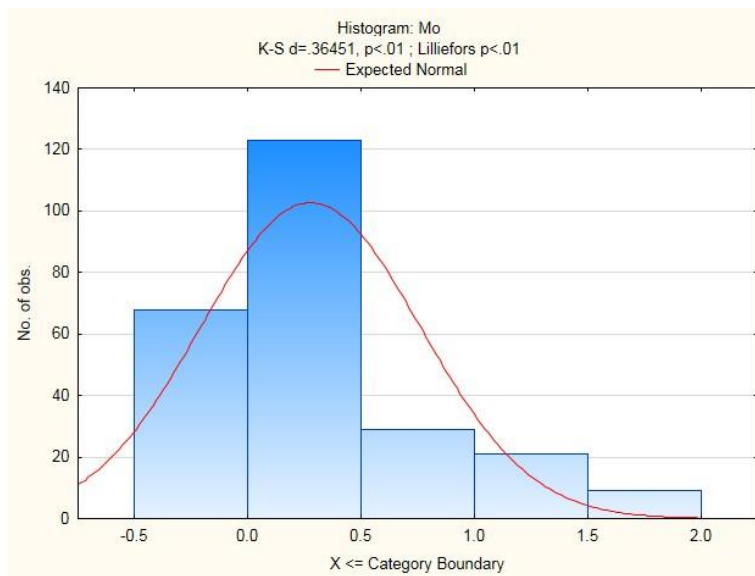
Element	Detection limit*	Element	Detection limit*	Element	Detection limit*
Au	1	Cr	1	Sn	0/5
Al	10	Cs	0/5	Sr	2
Ca	100	Cu	1	Te	0/2
Fe	100	La	1	Th	0/5
K	100	Li	1	Ti	10
Mg	100	Mn	5	Tl	0/2
Na	100	Mo	0/5	U	0/5
Ag	0/1	Nb	1	V	2
As	0/5	Ni	1	W	0/5
Ba	2	P	10	Y	0/5
Be	0/2	Pb	1	Yb	0/2
Bi	0/2	Rb	1	Zn	1
Cd	0/1	S	50	Zr	5
Ce	1	Sb	0/5		
Co	1	Sc	0/5		

*واحد حد تشخیص همه عناصر به جز طلا بر حسب میلی‌گرم بر کیلوگرم می‌باشد و برای عنصر طلا میکروگرم بر کیلوگرم می‌باشد.

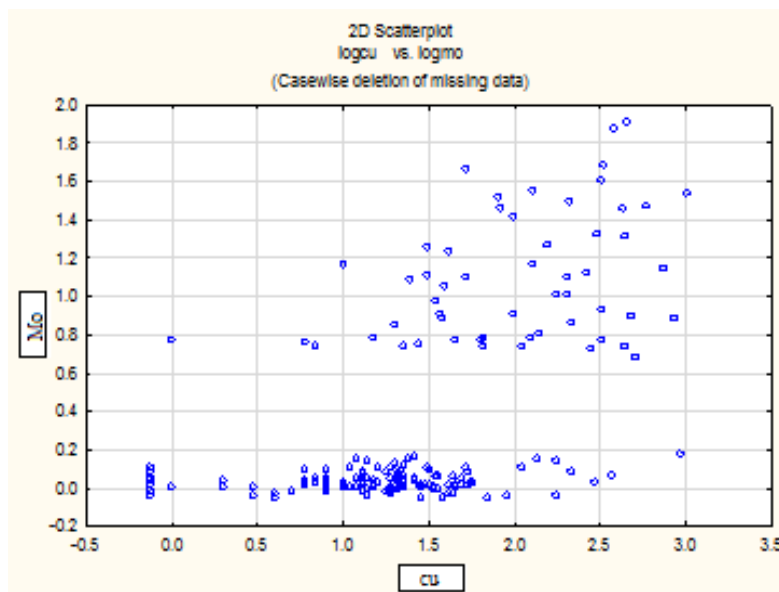


شکل ۵- تابع توزیع نرمال مقادیر عیار مس

Fig. 5. Normal distribution function of copper grade values



شکل ۶- تابع توزیع نرمال مقادیر عیار مولیبدن
Fig. 6. Normal distribution function within molybdenum grades



شکل ۷- ترسیمی از پراکندگی مس-مولیبدن در مقیاس لگاریتمی
Fig. 7. Cu-Mo scatterplot in logarithmic scale

است. سپس مقادیر عیار مس - مولیبدن و مختصات نقاط نمونه برداری به صورت یک مجموعه داده، به نرم افزار استاتیسیتیکا معرفی شد بعد از آن نرم افزار به کمک رابطه ۱ مقدار D_2 را برای هر نمونه به صورت تک ستونی محاسبه کرد با توجه به تعداد زیاد نمونه‌ها و محدودیت تعداد صفحات مقاله مقادیر D_2 به شکل جدول گزارش نشده و به جای آن از

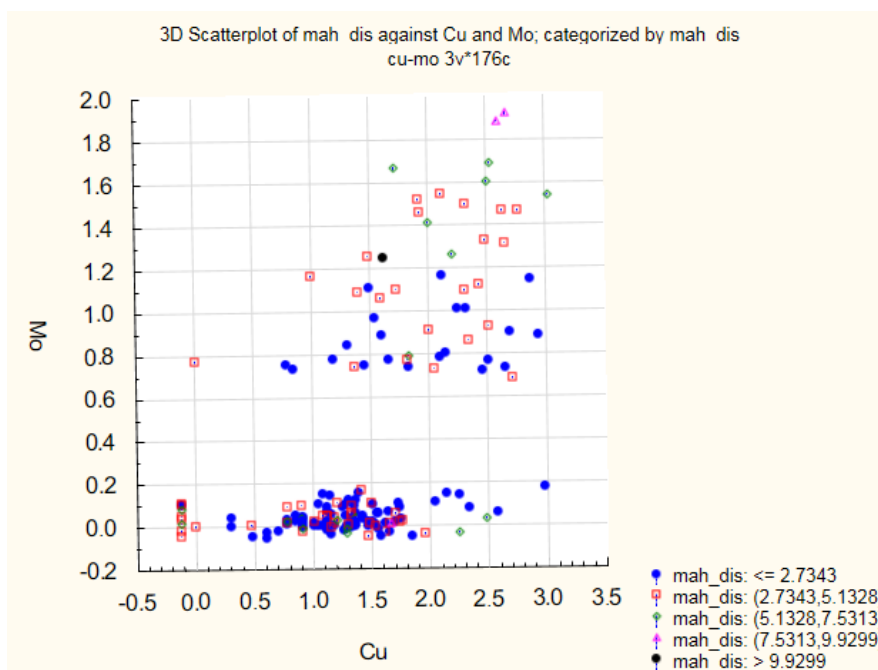
۴-۱-جدایش مقادیر ناهنجار بر اساس فواصل ماهالانوبیس

پس از پردازش اولیه مقادیر مس و مولیبدن در این بخش جدایش مقادیر ناهنجار انجام می‌شود. در شکل ۷ نمایی از پراکندگی ۱۷۷ نمونه مورد بررسی را بر اساس دو متغیر (متغیر اول عیار مس و متغیر دوم مولیبدن) در مقیاس لگاریتمی آمده

آمده است. همانطور که در نمودار مشاهده می‌شود در فاصله ۶٫۵ یک انفصال ضعیف و در فاصله ۶٫۹ یک جدایش قوی‌تر ایجاد شده. لذا می‌توان نمونه‌هایی با فاصله ماهالانوبیس فراتر از ۶٫۵ را به عنوان نمونه‌های ناهنجار ممکن و نمونه‌هایی با فاصله فراتر از ۶٫۹ را به عنوان نمونه‌های ناهنجار احتمالی معرفی کرد. نمونه‌های ناهنجار تعیین شده از نظر دو متغیر مس و مولیبدن توسط روش فوق‌الذکر در شکل ۱۰ آمده است.

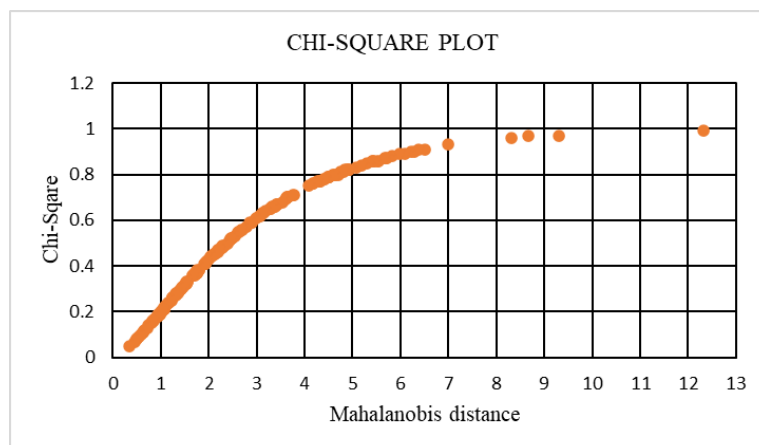
نموداری جهت نمایش پراکندگی و مقادیر D2 برای هر نمونه استفاده شده است (شکل ۸).

پس از محاسبه مقادیر ماهالانوبیس اقدام به محاسبه مقادیر احتمال تجمعی نمونه‌ها با توجه به ترتیب صعودی فواصل ماهالانوبیس شد. حال با در دست داشتن مقادیر احتمال تجمعی برای نمونه‌ها و همچنین دو متغیره بودن نوع محاسبات مقادیر χ^2 نمونه‌های مورد نظر تعیین شد. نمودار کای-اسکور بر اساس مقادیر D2 و χ^2 رسم شد که نتیجه آن در شکل ۹



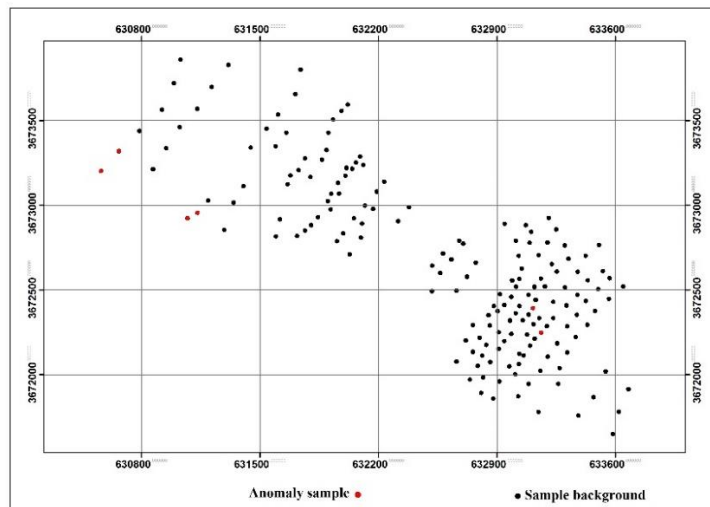
شکل ۸- نمایش مقادیر فواصل ماهالانوبیس برای هر نمونه

Fig. 8. Display of Mahalanobis distances for each sample



شکل ۹- نمودار کای - اسکور مس و مولیبدن محدوده ظفرقند

Fig. 9. Chi-square diagram of copper and molybdenum in Zafarghand area



شکل ۱۰- نمایش نمونه‌های ناهنجار به کمک روش فواصل ماهالانوبیس
Fig. 10. Displaying anomaly samples using the Mahalanobis distances

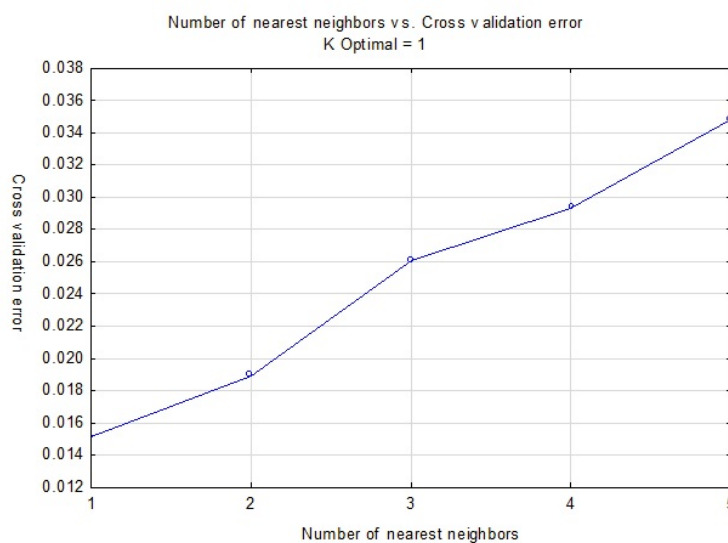
برای تعیین بهترین مقدار K الگوریتم این روش در نرم افزار استاتستیکا (STATISTICA) اجرا شد. برای مقادیر K از مقدار ۱ تا ۱۰ نرم افزار به تعداد ۱۰ بار اجرا شد و هر بار مقدار خطا اندازه‌گیری شد. با توجه به مقدار خطا محاسبه شده مقدار K بهینه برابر با $K=1$ تعیین شد (شکل ۱۱). برای محاسبه خطا از معیار میزان خطا براساس اعتبارسنجی متقاطع (در ابتدا مدل Cross validation شبکه مورد نظر ساخته شده سپس میزان خطا محاسبه شد) استفاده شد.

حال جهت اعمال روش‌های داده کاوی بر روی نتایج حاصله مقدار ۱ به عنوان وزن نمونه‌های زمینه و مقدار ۲ به عنوان وزن نمونه‌های ناهنجار برای هر نمونه در نظر گرفته شد و علاوه بر مقادیر طول و عرض جغرافیایی و عیار دو عنصر مس و مولیبدن به عنوان ورودی، مقادیر وزن هر نمونه نیز به منظور فرآیند پیش بینی به نرم افزار معرفی می‌شود.

۴-۲- پیش‌بینی مقادیر ناهنجار با استفاده از الگوریتم

KNN

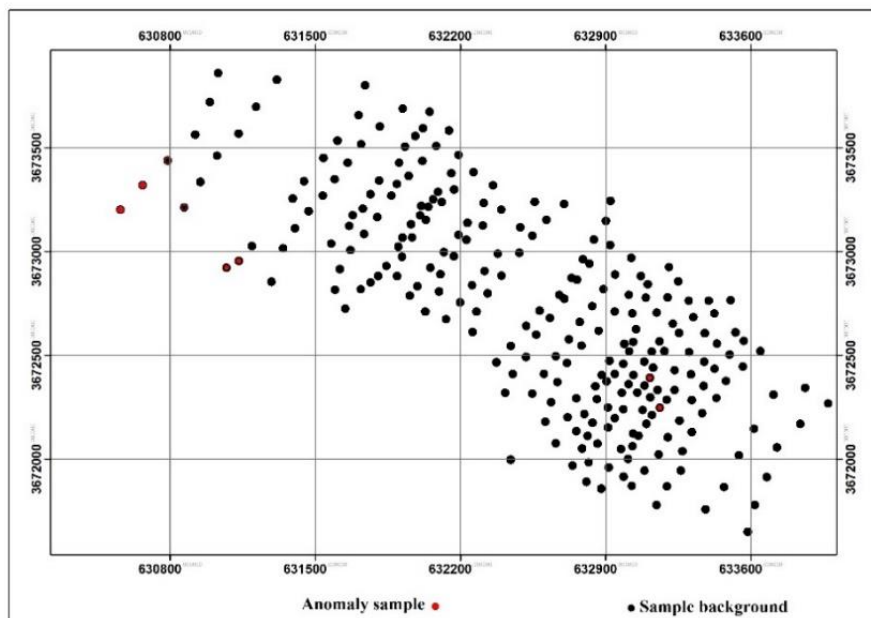
به منظور استفاده از روش KNN بر روی داده‌ها و همچنین



شکل ۱۱- نمودار مقدار خطای میانگین برای مقادیر مختلف k
Fig. 11. Plot of mean error value for different values of K

توجه به نوع نمونه های مورد بررسی، نمونه های ناهنجار بر اساس روش مذکور پیش بینی شد. خطای این روش براساس معیار MSE مقدار صفر محاسبه شد به این صورت که از ۱۷۷ نمونه مورد بررسی هیچ نمونه ای اشتباه قضاوت نشده است، که این نشان می دهد که ترکیب این دو روش برای پیش بینی مقادیر ناهنجار از دقت بالایی برخوردار است. به منظور نمایش عملکرد ترکیب این دو روش مجدد از داده های محدود زلفرند به عنوان داده هایی جدید برای شبکه مذکور استفاده شده که نتایج آن در شکل ۱۲ قابل مشاهده است.

مجموعه داده های یادگیری خود به سه دسته داده های آموزش- اعتبارسنجی و آزمایش تقسیم بندی می شوند. در این تقسیم بندی برای دستیابی به پاسخ قابل قبول، تعداد داده های مورد استفاده در مجموعه داده های فاز آموزش بیشتر است. همچنین حد امکان، تقسیم بندی داده ها به صورت تصادفی باشد. در این مطالعه ۷۰ درصد داده ها برای آموزش مابقی آن برای اعتبارسنجی و تست شبکه مورد نظر صورت گرفت. در این روش مجموعه داده شامل طول و عرض جغرافیایی، عیار مس-مولیبدن و همچنین وزن هر یک از داده ها (ارزش گذاری بر اساس فواصل ماهالانوبیس) جهت آموزش معرفی شد. سپس با



شکل ۱۲- نمایش نمونه های ناهنجار به کمک ترکیب دو روش KNN و فواصل ماهالانوبیس

Fig. 12. Displaying anomaly samples using the combination of KNN and Mahalanobis distances

نمونه تنها ۳ نمونه به اشتباه محاسبه شد، که این مقدار قابل قبولی است. در شکل ۱۳ نتیجه حاصل از اجرای این روش در مورد ۱۷۷ نمونه محدود زلفرند را به عنوان خروجی شبکه طراحی شده و عملکرد آن آمده است.

۴-۴- پیش بینی مقادیر ناهنجار با استفاده از شبکه عصبی کانولوشن CNN

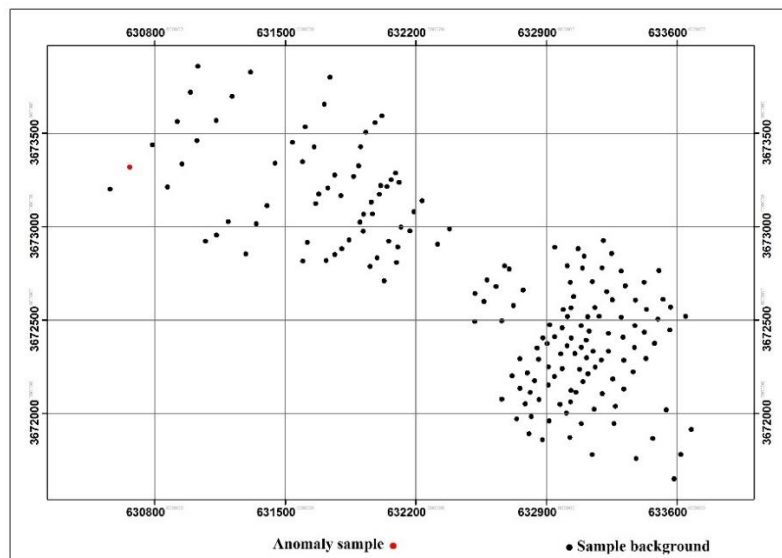
مهمترین قسمت شبکه عصبی، بخش آموزش است که با تغییر تعداد لایه های پنهان، نرون، نوع تابع فعال سازی و تعداد تکرارها بر مبنای سعی و خطا می توان مدل بهینه شبکه با کمترین خطا را به دست آورد. شبکه مورد استفاده دارای پارامتر ورودی (طول و عرض جغرافیایی، عیار مس و مولیبدن و وزن هر نمونه) و یک

۴-۳- پیش بینی مقادیر ناهنجار با استفاده از شبکه ساده بیز

در این قسمت به منظور استفاده از روش طبقه بندی ساده بیزی، از الگوریتم آن در نرم افزار استاتستیکا استفاده شد. در این روش نیز یک مجموعه داده شامل طول و عرض جغرافیایی، عیار مس-مولیبدن و همچنین وزن هر یک از نمونه ها معرفی شد. سپس فرآیند پیش بینی مقادیر ناهنجار با الگوریتم مذکور روی داده انجام شد. نتیجه حاصل، طراحی شبکه ای برای ارزیابی نمونه های ناهنجار بر اساس مقادیر وزن هر یک از نمونه ها می باشد. خطای شبکه طراحی شده در مورد تعیین مقادیر ناهنجار در این روش ۰,۰۱ بدست آمد به این صورت که از ۱۷۷

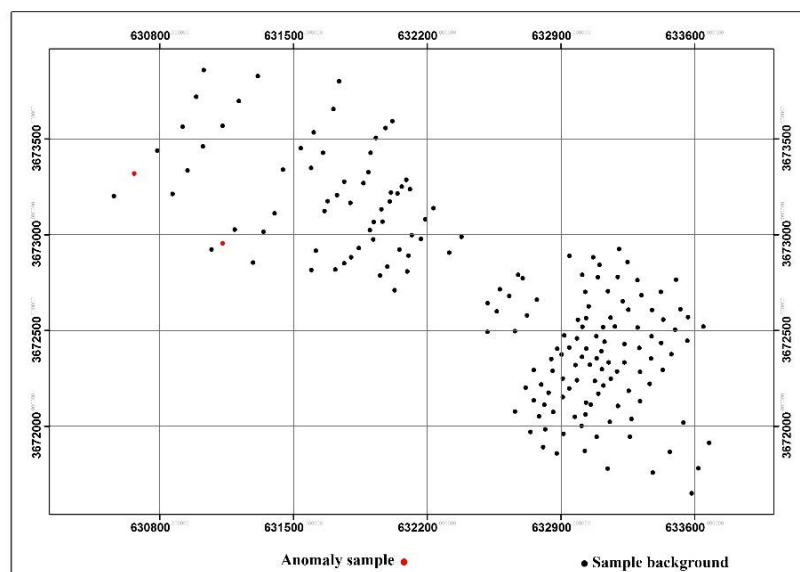
مصنوعی استفاده شد. نتایج حاصل از این روش در جدول ۴ آمده است. خطای بدست آمده در این روش ۰,۰۱ بدست آمد و در مورد محاسبه مقادیر ناهنجار با این روش، تنها ۲ نمونه از ۱۷۷ نمونه اشتباه بررسی شده است. به منظور عملکرد و کارایی ترکیب ۲ روش فواصل ماحالانوبیس و شبکه عصبی مجدداً نمونه‌های ناهنجار تعیین شده توسط این روش‌های فوق‌الذکر به عنوان خروجی در شکل ۱۴ نشان داده شده است.

لایه میانی با ۴ نرون و یک لایه خروجی با ۲ نرون می‌باشد. به منظور استفاده از شبکه عصبی کانولوشن بر روی داده‌ها نیز از الگوریتم آن در نرم افزار استاتیسیتیکا استفاده شد. از توابع تانژانت هیپربولیک (هدلولی) و نمایی نرمال شده به عنوان توابع فعال‌سازی و از الگوریتم BFGS (روشی در محاسبات عددی بهینه‌سازی است. این روش تقریبی، روش بهینه‌سازی نیوتون است). برای به دست آوردن وزن‌های بهینه شبکه عصبی



شکل ۱۳- نمایش نمونه‌های ناهنجار به کمک ترکیب دو روش شبکه ساده بیز و فواصل ماحالانوبیس

Fig. 13. Displaying anomaly samples using combining two methods simple Bayes network and Mahalanobis



شکل ۱۴- نمایش نمونه‌های ناهنجار به کمک ترکیب دو روش شبکه عصبی کانولوشن (CNN) و فواصل ماحالانوبیس

Fig. 14. Displaying anomaly samples using a combination of two methods, convolutional neural network (CNN) and Mahalanobis distance

جدول ۴- خلاصه ای از نتیجه دست آمده از روش شبکه عصبی کانولوشن

Table 4. Summary of the results obtained from the convolutional neural network method

net.name	training perf	test perf	validation perf	ALGORITHM	hidden activation	Output activation
MLP 4-4-2	98.3870968	100.0000	100.0000	BFGS 7	Tanh	SoftMax

۵- نتیجه گیری

کاوی مورد مطالعه قرار گرفت. پس از اجرای الگوریتم مربوط به هر روش میزان خطا و عملکرد آن نیز بررسی شد. با توجه به خروجی‌های بدست آمده، K- نزدیکترین همسایگی که هیچ نمونه خطا نداشته نسبت به روش‌های دیگر قوی‌تر عمل کرده (روش شبکه عصبی کانولوشن ۲ نمونه و شبکه ساده بیز ۳ نمونه از ۱۷۷ نمونه خطا داشته‌اند). به این ترتیب با مقایسه میزان خطای شبکه‌های طراحی شده، می‌توان برای پیش‌بینی مقادیر ناهنجار در مبحث جدایش ناهنجاری از زمینه، ترکیب روش‌های فواصل ماهالانوبیس و K- نزدیکترین همسایگی را معرفی کرد. لازم به ذکر است که با نوشتن برنامه رایانه‌ای مختص به روش ترکیبی معرفی شده با استفاده از نرم افزارهای برنامه نویسی به خوبی این ترکیب قابل استفاده برای کارهای مشابه نیز می‌باشد.

منطقه مورد پژوهش بخشی از زون ساختاری ایران- مرکزی به‌شمار می‌رود و بطور دقیق‌تر در نوار آتشفشانی ارومیه-دختر قرار می‌گیرد. به منظور جدایش مقادیر ناهنجاری از زمینه از نظر دو عنصر مس و مولیبدن منطقه مورد مطالعه (ظفرقند - اصفهان) از روش چند متغیره فواصل ماهالانوبیس استفاده شد. به این صورت با رسم نمودار کای-اسکور برای دو عنصر مس و مولیبدن محدوده مورد مطالعه، نمونه‌های ناهنجار مشخص شدند. پس از آن براساس نتایج حاصل به همراه پارامترهای عیار مس-مولیبدن، طول و عرض جغرافیایی نمونه‌ها و وزن هر نمونه (براساس نتایج فواصل ماهالانوبیس مقدار ۱ برای نمونه‌های زمینه و مقدار ۲ برای نمونه ناهنجار)، در اختیار روش‌های داده

مراجع

- Alaminia, Z., Bagheri, H., Salehi, M., 2016. Geological, geophysical investigations and studies of fluids involved in Zafarghand exploration area (north-east of Isfahan province, Iran). *Economic Geology* 9(2), 295-312. <https://doi.org/10.22067/ECONG.V10I2.76528>.
- Aminoroayaei Yamini, M., Tutti, F., Haschke, M., Ahmadian, J., Murata, M., 2017. Synorogenic copper mineralization during the Alpine-Himalayan orogeny in the Zafarghand copper exploration district, Central Iran: petrography, geochemistry and alteration thermometry. *Geological Journal* 52(2), 263-281. <https://doi.org/10.1002/gj.2755>.
- Aminroayaei Yamini, M., Tutti, F., Aminoroayaei Yamini, M.R., Ahmadian, J., Wan, B., 2017. Examination of chloritization of biotite as a tool for reconstructing the physicochemical parameters of mineralization and associated alteration in the Zafarghand porphyry copper system, Ardestan, Central Iran: mineral-chemistry and stable isotope analyses. *Mineralogy and Petrology* 111, 747-759. <https://doi.org/10.1007/s00710-016-0486-7>.
- Bax, E., 2000. Validation of nearest neighbor classifiers: *IEEE transactions information theory* 58, 2746-2752. <https://doi.org/10.1109/18.887892>.
- Beucher, A., Österholm, P., Martinkauppi, A., Eden, P., 2013. Artificial neural network for acid sulfate soil mapping: Application to the Sirppujoki River catchment area, south western Finland *Journal of Geochemical Exploration* 125, 46-55. <https://doi.org/10.1016/j.gexplo.2012.11.002>.
- Barbarin, B., 1990. Granitoids: main petrogenetic classifications in relation to origin and tectonic setting. *Geological Journal* 25(3-4), 227-238. <https://doi.org/10.1002/GJ.3350250306>.
- Cheng, Q., 1999. Spatial and scaling modelling for geochemical anomaly separation. *Journal of Geochemical exploration* 65(3), 175-194. [https://doi.org/10.1016/S0375-6742\(99\)00028-X](https://doi.org/10.1016/S0375-6742(99)00028-X).
- Deng, B., 2020. Machine learning on density and elastic property of oxide glasses driven by large dataset. *Journal of non-crystalline solids* 529, 119768. <https://doi.org/10.1016/j.jnoncrysol.2019.119768>.
- Filzmoser, P., Garrett, G., Reimann, C., 2005. Multivariate outlier detection in exploration geochemistry. *Computers & Geosciences* 31(5), 579-587. <https://doi.org/10.1016/j.cageo.2004.11.013>.

- Farhadi, S., Afzal, P., Boveiri Konari, M., Daneshvar Saein, L., Sadeghi, B. 2022. Combination of Machine Learning Algorithms with Concentration-Area Fractal Method for Soil Geochemical Anomaly Detection in Sediment-Hosted Irankuh Pb-Zn Deposit, Central Iran. *Minerals* 12(6), 689. <https://doi.org/10.3390/min12060689>.
- Ghannadpour, S.S., Hezarkhani, A., 2015. Investigation of Cu, Mo, Pb, and Zn geochemical behavior and geological interpretations for Parkam porphyry copper system, Kerman, Iran. *Arabian Journal of Geosciences* 8(9), 7273-7284. <https://doi.org/10.1007/s12517-014-1732-0>.
- Ghannadpour, S.S., Hezarkhani, A., Sabetmobarhan, A., 2015. Some statistical analyses of Cu and Mo variates and geological interpretations for Parkam porphyry copper system, Kerman, Iran. *Arabian Journal of Geosciences* 8(1), 345-355. <https://doi.org/10.1007/s12517-013-1096-x>.
- Ghannadpour, S.S., Hezarkhani, A. 2016a. Introducing 3D U-statistic method for separating anomaly from background in exploration geochemical data with associated software development. *Journal of Earth System Science* 125(2), 387-401. <https://doi.org/10.1007/s12040-016-0657-2>.
- Ghannadpour, S.S., Hezarkhani, A., 2016b. Exploration geochemistry data-application for anomaly separation based on discriminant function analysis in the Parkam porphyry system (Iran): *Geoscience Journal* 20(6), 837-850. <https://doi.org/10.1007/s12303-015-0064-8>.
- Ghannadpour, S.S., Hezarkhani, A., Roodpeyma, T., 2017. Combination of Separation Methods and Data Mining Techniques for Prediction of Anomalous Areas in Susavar, Central Iran. *African Journal of Earth Sciences* 134, 516-525. <https://doi.org/10.1016/j.jafrearsci.2017.07.015>.
- Kavzoglu, T., 2009. Increasing the accuracy of neural network classification using refined training data. *Environmental Modelling & Software* 24(7), 850-858. <https://doi.org/10.1016/j.envsoft.2008.11.012>.
- Koozhadi, F., 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, 299-307. <https://doi.org/10.30495/ijes.2021.685397>.
- Liu, X., Han, F., Ghazali, K.H., Mohamed, I.I., Zhao, Y., 2019. A review of convolutional neural networks in remote sensing image. In *Proceedings of the 2019 8th International Conference on Software and Computer Applications*, p. 263-267. <https://doi.org/10.1145/3316615.3316712>.
- Li, W., Kong, D., Wu, J., 2017. A novel hybrid model based on extreme learning machine, k-nearest neighbor regression and wavelet denoising applied to short-term electric load forecasting. *Energies* 10(5), 694. <https://doi.org/10.3390/en10050694>.
- 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. <https://doi.org/10.1016/j.gexplo.2017.05.008>.
- Nabilou, M., Afzal, P., Arian, M., Adib, A., Kazemi Mehrnia, A., Jami, M., Kheyrollahi, H., Akhavan Aghdam, M.R., Ameri, A., Daneshvar Saein, L., 2022. Determination of relationship between Rare Earth Elements (REEs) mineralization and major faults using fractal modeling in Gazestan deposit, central Iran. *Bollettino di Geofisica Teorica ed Applicata* 63(3), 495-518. <https://doi.org/10.4430/bgo00388>.
- Sinclair, A. J., 1991. A fundamental approach to threshold estimation in exploration geochemistry: probability plots revisited. *Journal of geochemical exploration* 41(1-2), 1-22. [https://doi.org/10.1016/0375-6742\(91\)90071-2](https://doi.org/10.1016/0375-6742(91)90071-2).
- Sarjoughian, F., Lentz, D., Kananian, A., Ao, S., Xiao, W., 2018. Geochemical and isotopic constraints on the role of juvenile crust and magma mixing in the UDMA magmatism, Iran: evidence from mafic microgranular enclaves and cogenetic granitoids in the Zafarghand igneous complex. *International Journal of Earth Sciences* 107, 1127-1151. <https://doi.org/10.1007/s00531-017-1548-8>.
- Shahbazi, S., Ghaderi, M., Afzal, P., 2021. Prognosis of gold mineralization phases by multifractal modeling in the Zehabad epithermal deposit, NW Iran. *Iranian Journal of Earth Sciences* 13, 31-40. <https://doi.org/10.30495/ijes.2021.678957>.
- Shahi, H., Ghavami, R., Rouhani, A.K., 2016. Detection of deep and blind mineral deposits using new proposed frequency coefficients method in frequency domain of geochemical data. *Journal of Geochemical Exploration* 162, 29-39. <https://doi.org/10.1016/j.gexplo.2015.12.006>.

Yang, Y., Liu, X., 1999. A re-examination of text categorization methods. In Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, pp. 42-49. <https://doi.org/10.1145/312624.312647>.