Vol. 17, No. 4, 2025, pp. 1–36 ISSN (Print): 2008-7306 ISSN (Online): 2423-5865

Journal of Economic Geology

https://econg.um.ac.ir

RESEARCH ARTICLE 10.22067/econg.2025.1157 OPEN ACCESS

Application of Deep Learning-Based Multi-Class Anomaly Detection in Geochemical Exploration with Bayesian Hyperparameter Tuning

Puyan Javandel [1 *], Nader Fathianpour [2], Seyed Hassan Tabatabaei [3]

[1] Ph.D. candidate, Faculty of Mining, Isfahan University of Technology, Isfahan, Iran
[2] Associate Professor, Faculty of Mining, Isfahan University of Technology, Isfahan, Iran
[3] Professor, Faculty of Mining, Isfahan University of Technology, Isfahan, Iran

ARTICLE INFO

Article History

Received: 29 September 2025
Revised: 15 December 2025
Accepted: 17 December 2025

Keywords

Geochemical exploration
Deep learning
Autoencoder
DBSCAN
Isolation Forest
Bayesian optimization

***Corresponding author**

Puyan Javandel
✉ puyanjavan.pj@aut.ac.ir

ABSTRACT

In this study, 1,238 stream-sediment samples from the Pariz and Chahargonbad areas of southwestern Kerman Province were analyzed to identify geochemical anomalies. The dataset was enhanced using principal component analysis (PCA), independent component analysis (ICA) and polynomial feature expansions, followed by normalization to capture both linear and non-linear variability. Four anomaly–background separation models were evaluated: Isolation Forest, DBSCAN, Autoencoder and Variational Autoencoder (VAE). Model hyper-parameters (e.g. the number of layers and neurons per layer) were tuned via Bayesian optimization in Python using the Optuna library. During tuning, the scoring metrics were reconstruction error for the Autoencoder; likelihood-based reconstruction error for the VAE; the anomaly score for Isolation Forest; and the mean distance of each sample to the cluster core for DBSCAN. For spatial presentation, samples were assigned to upstream drainage basins, and each sample's result was propagated to its upstream area; the outcomes were then compared with known mineralization points across the region. Systematic hyper-parameter tuning improved the detection of geochemical signatures and highlighted zones with mineral potential. Comparative analysis showed that robust feature engineering combined with Bayesian optimization significantly enhances anomaly-detection accuracy. These findings underscore the practical value of advanced machine-learning techniques in geochemical exploration and provide a framework for more targeted field investigations in mineral-rich regions.

How to cite this article

Javandel, P., Fathianpour, N. and Tabatabaei, S.H., 2025. Application of Deep Learning-Based Multi-Class Anomaly Detection in Geochemical Exploration with Bayesian Hyperparameter Tuning. Journal of Economic Geology, 17(4): 1–36. (in Persian with English abstract) https://doi.org/10.22067/econg.2025.1157

EXTENDED ABSTRACT

Introduction

Efficient recognition of geochemical anomalies underpins mineral exploration and environmental assessment. Rising data dimensionality and nonlinearity challenge classical uni/ multivariate thresholds and geostatistics, motivating machine-learning (ML) and deep-learning (DL) approaches that capture complex, multi-element signals and integrate ancillary data [1]. Autoencoder-family models have proven effective for weak, multivariate anomalies using reconstruction-error criteria, and adversarial variants further enrich latent representations. Yet performance hinges on feature engineering and hyperparameter selection; prior work often relies on defaults or ad-hoc tuning. This study targets those gaps by: (i) expanding geochemical features (PCA, ICA, polynomial terms), (ii) optimizing IF, DBSCAN, AE and VAE via Bayesian search, and (iii) validating anomaly maps against known mineralization in a well-endowed metallogenic belt (Sarcheshmeh–Meydouk district).

Methodology and Approaches

Geological setting and data: The area lies within the Urumieh–Dokhtar magmatic arc, with Oligocene–Miocene calc-alkaline intrusions hosting world-class porphyry Cu deposits (e.g., Sarcheshmeh, Meydouk). A dataset of 1,238 stream-sediment samples was compiled for multi-element analysis.

Preprocessing and feature engineering: Outliers were screened (interquartile-range criteria), element values log-transformed and min-max normalized. Correlation structure and hierarchical clustering guided feature design. Five PCA and five ICA components were appended, and second-order polynomial terms were generated for Cu-correlated elements to capture nonlinear interactions

Models and objective functions

- **Isolation Forest (IF):** tuned (estimators, max_ samples, contamination) to maximize average decision-function score (higher = more normal), sharpening separation from anomalies
- **DBSCAN:** tuned (ε, min_samples, k) under a composite objective balancing outlier fraction and neighborhood-distance compactness to form geologically coherent clusters while isolating true outliers.
- **AE (deep):** encoder–decoder with monotonically contracting hidden sizes; Optuna minimized a composite of mean reconstruction error and its standard deviation to favor accurate and stable reconstructions
- **VAE:** latent Gaussian prior (μ, σ) trained with MSE reconstruction + β·KL divergence; layers, hidden width, latent size, batch size, and output activation were optimized for minimum validation loss

Bayesian optimization (Optuna) adaptively explored hyperparameter spaces, improving efficiency over grid/random search and providing consistent, data-specific configurations. Model scores were aggregated to catchment outlets (SCB) to generate anomaly probability classes for mapping and validation against 39 known mineralized points.

Results and Discussion

Optimized configurations and comparative performance

Bayesian tuning yielded clear performance gains across all detectors.

- **Isolation Forest:** high-anomaly area 21%, capturing 10/39 occurrences.
- **DBSCAN:** 47% area, 24/39 occurrences.
- **Autoencoder (AE):** 36% area, 24/39 occurrences - the best balance of precision (compact footprint) and recall of known mineralization.
- **Variational Autoencoder (VAE):** 42% area, 24/39 occurrences.

Spatially, the AE geochemical anomaly map outlines coherent belts coincident with known porphyry systems and prospective intrusive centers. Feature augmentation (PCA+ICA+polynomial terms) consistently lifted signal-to-noise and stabilized training, especially for deep models.

Model-specific behavior

IF favored compact delineations but under-captured weak, diffuse halos. DBSCAN effectively separated dense background clusters from dispersed outliers but inflated high-anomaly areal extent at tuned ε/min_samples. AE/VAE reduced over-generalization while preserving mineralization hits; AE slightly outperformed VAE in footprint control under the chosen β and architecture

Conclusions

1) Robust feature engineering coupled with Bayesian hyperparameter optimization substantially improves unsupervised geochemical anomaly detection.

2) Among tested models, the Autoencoder provides the most favorable precision–recall trade-off for the Pariz–Chahargunbad dataset, capturing 24/39 known occurrences within a 36% high-anomaly footprint.

3) VAE and DBSCAN achieve comparable recall but at larger footprints (42–47%), whereas Isolation Forest is conservative but misses a portion of known targets.

4) SCB-based spatial propagation yields interpretable maps aligned with drainage-mediated signal transport.

.

دوره ۱۷، شماره ۴، ۱۴۰۴، صفحه ۱ تا ۳۶ شاپا چاپی: ۷۳۰۶-۲۰۰۸ شاپا الکترونیکی: ۵۸۶۵-۲۴۲۳

زمین‌شناسی اقتصادی

https://econg.um.ac.ir

مقاله پژوهشی 10.22067/econg.2025.1157 OPEN ACCESS

بهینه‌سازی فرایند تشخیص ناهنجاری‌های چند دسته‌ای در اکتشافات زمین‌شیمیایی با استفاده از یادگیری عمیق و تنظیم بهینه هایپرپارامترها به روش بیزی

پویان جواندل [۱*]، نادر فتحیان‌پور [۲]، سید حسن طباطبایی [۳]

[۱] دانشجوی دکتری، گروه اکتشاف معدن، دانشکده معدن، دانشگاه صنعتی اصفهان، اصفهان، ایران

[۲] دانشیار، گروه اکتشاف معدن، دانشکده معدن، دانشگاه صنعتی اصفهان، اصفهان، ایران

[۳] استاد، گروه اکتشاف معدن، دانشکده معدن، دانشگاه صنعتی اصفهان، اصفهان، ایران

اطلاعات مقاله

تاریخ دریافت: ۱۴۰۴/۰۷/۰۷

تاریخ بازنگری: ۱۴۰۴/۰۹/۲۴

تاریخ پذیرش: ۱۴۰۴/۰۹/۲۶

واژه‌های کلیدی

اکتشافات زمین‌شیمیایی

یادگیری عمیق

خودرمزنگار

دی‌بی اسکن

جنگل ایزوله

بهینه‌سازی بیزی

نویسنده مسئول

پویان جواندل

puyanjavan.pj@aut.ac.ir

چکیده

در این پژوهش، ۱۲۳۸ نمونه رسوب از پاریز و چارگنبد در جنوب‌غربی کرمان برای شناسایی آنومالی‌های زمین‌شیمیایی بررسی شد. داده‌ها با تحلیل مؤلفه اصلی، تحلیل مؤلفه مستقل و تبدیل‌های چندجمله‌ای غنی‌سازی و برای ثبت تغییرات خطی و غیرخطی نرمال‌سازی شد. چهار مدل جدایش آنومالی از زمینه به‌کار رفت: جنگل ایزوله، دی‌بی اسکن، خودرمزنگار و خودرمزنگار واریانسی. ابرپارامترهای هر مدل (مانند تعداد لایه‌ها و نورون‌ها) با بهینه‌سازی بیزی در پایتون و با کتابخانه آپچونا تنظیم‌شد. برای امتیازدهی طی فرایند تنظیم، در خودرمزنگار از خطای بازسازی، در خودرمزنگار واریانسی از معیار احتمال/بازسازی، در جنگل ایزوله از امتیاز آنومالی و در دی‌بی اسکن از میانگین فاصله هر نمونه تا مرکز/هسته خوشه استفاده شد تا تنظیم به درستی انجام‌شود. برای نمایش مکانی نتایج از حوضه‌های آبریز بالادست استفاده شد و پاسخ هر نمونه به بالادست تعمیم‌یافت. سپس نتایج با نقاط کانی‌زایی شناخته‌شده سطح منطقه مقایسه شد. تنظیم نظام‌مند هایپرپارامترها موجب بهبود آشکارسازی نشانه‌های زمین‌شیمیایی و برجسته‌سازی نواحی دارای پتانسیل معدنی شد. تحلیل مقایسه‌ای نشان داد که مهندسی ویژگی نیرومند همراه با بهینه‌سازی بیزی دقت تشخیص ناهنجاری را به طور معناداری افزایش می‌دهد. این یافته‌ها ارزش عملی روش‌های پیشرفته یادگیری ماشین در اکتشاف زمین‌شیمیایی را برجسته کرده و چهارچوبی برای پژوهش‌های میدانی هدفمندتر در مناطق غنی از مواد معدنی ارائه می‌دهد.

استناد به این مقاله

جواندل، پویان؛ فتحیان‌پور، نادر و طباطبایی، سید حسن، ۱۴۰۴. بهینه‌سازی فرایند تشخیص ناهنجاری‌های چند دسته‌ای در اکتشافات زمین‌شیمیایی با استفاده از یادگیری عمیق و تنظیم بهینه هایپرپارامترها به روش بیزی. زمین‌شناسی اقتصادی، ۱۷(۴): ۱-۳۶. https://doi.org/10.22067/econg.2025.1157

مقدمه

تشخیص دقیق و کارآمد ناهنجاری‌های زمین‌شیمیایی سنگ بنای اکتشاف معدنی، نظارت بر محیط زیست و راهبردهای مدیریت منابع گسترده است. با مشخص‌کردن مکان‌هایی با غلظت‌های غیرعادی عناصر خاص، اکتشاف گران می‌توانند به سرعت مناطق امیدبخش را هدف قرار داده و در نتیجه هزینه و زمان بررسی‌های میدانی را کاهش دهند. علاوه بر این، آشکارسازی آنومالی‌ها در داده‌های زمین‌شیمیایی به طور فزاینده‌ای برای ارزیابی‌های زیست‌محیطی اهمیت پیداکرده است؛ جایی که شناسایی نقاط احتمالی آلودگی می‌تواند به تلاش‌های اصلاحی و تصمیم‌گیری‌های سیاستی کمک کند (Zuo, 2011). پس شناسایی آنومالی‌های زمین‌شیمیایی، جنبه‌ای اساسی در اکتشاف معدنی است که به دانشمندان علوم زمین این امکان را می‌دهد که مناطق مستعدی را شناسایی کنند که ممکن است دارای ذخایر معدنی با ارزش اقتصادی باشند. روش‌های سنتی اکتشاف زمین‌شیمیایی بر روش‌های آماری و زمین‌آماری متکی بوده‌اند؛ با این حال، پیچیدگی فزاینده سامانه‌های کانی‌سازی و ماهیت چندبعدی داده‌های زمین‌شیمیایی نوین، نیاز به روش‌های محاسباتی پیشرفته‌تر دارد (Xiong and Zuo, 2020; Yang et al., 2021). پیشرفت‌های اخیر در یادگیری ماشین و یادگیری عمیق جایگزین‌های قدرتمندی برای شناسایی آنومالی‌های زمین‌شیمیایی ارائه داده‌اند که توانایی مدل‌سازی روابط غیرخطی، مدیریت داده‌های با ابعاد بالا و ادغام چندین مجموعه داده زمین‌شیمیایی، ژئوفیزیکی و سنجش از دور را دارند (Zhang et al., 2019; Xiong and Zuo, 2020; Zhang et al., 2024). این روش‌ها در حوزه‌های مختلف زمین‌شیمی کاربردی نویدبخش بوده‌اند؛ به ویژه در شناسایی ناهنجاری‌های چند عنصری ضعیف که ممکن است توسط روش‌های آماری متداول نادیده گرفته شوند (Ueki et al., 2018; Horrocks, 2019; Li et al., 2020). با این حال، چالش‌های کلیدی همچنان باقی است؛ به ویژه در مهندسی ویژگی‌ها، تنظیم هایپرپارامتر و اعتبارسنجی در برابر کانی‌سازی‌های شناخته‌شده (Xiong and Zuo, 2016). این بررسی بر مناطق پاریز و چهارگنبد در جنوب‌غربی استان کرمان ایران متمرکز است؛ محلی که از آن ۱۲۳۸ نمونه رسوب برای ارزیابی آنومالی‌های زمین‌شیمیایی جمع‌آوری شده است. در این پژوهش، چهار روش پیشرفته تشخیص آنومالی (جنگل ایزوله، دی‌بی اسکن، خودرمزنگار ساده و خودرمزنگار واریانسی) استفاده شد و در نهایت عملکرد مدل‌های ساخته‌شده با استفاده از نقاط کانی‌زایی و نشانه‌های کانی‌سازی در سطح زمین مورد مقایسه قرار گرفتند.

از نظر تاریخی، تشخیص ناهنجاری‌های زمین‌شیمیایی به روش‌های آستانه آماری تک متغیره مانند قانون میانگین ± ۲* برابر انحراف معیار، نمودارهای احتمال و نمودارهای جعبه‌ای متکی بوده است (Xiong and Zuo, 2020). برای درک بهتر روابط پیچیده زمین‌شیمیایی از روش‌های پیشرفته‌تر چند متغیره، از جمله تحلیل مؤلفه‌های اصلی به کارگرفته شده‌اند؛ اما این روش‌ها اغلب فرض می‌کنند که روابط خطی وجود دارد و ممکن است نتوانند الگوهای زمین‌شیمیایی پیچیده را به طور مؤثری مدل‌سازی کنند (Yang et al., 2021). روش‌های زمین آماری، مانند کریجینگ، مدل‌های فراکتالی/ چندفراکتالی و تحلیل‌های محلی تکینگی، برای یکپارچه‌سازی اطلاعات مکانی در تشخیص آنومالی‌ها استفاده شده است. این روش‌ها بینش‌های ارزشمندی را ارائه می‌دهند؛ اما در برخورد با مجموعه داده‌های بزرگ مقیاس و با ابعاد بالا دچار مشکل می‌شوند و اغلب به انتخاب مؤلفه‌ها نیاز دارند (Xiong and Zuo, 2016).

رویکردهای یادگیری ماشین به دلیل توانایی آنها در یادگیری روابط پیچیده و غیرخطی از مجموعه داده‌های بزرگ، توجه زیادی را در اکتشافات زمین‌شیمیایی به خود جلب کرده‌اند. روش‌های یادگیری نظارت‌شده، مانند ماشین‌های بردار پشتیبان، درختان تصمیم و جنگل‌های تصادفی، به طور گسترده‌ای برای مدل‌سازی پتانسیل معدنی و جداسازی آنومال از زمینه مورد استفاده قرار گرفته‌اند (Xiong and Zuo, 2020). با این حال،

DOI: 10.22067/econg.2025.1157

عملکرد آنها اغلب به دلیل نیاز به داده‌های آموزشی برچسب‌خورده که در مراحل اولیه اکتشاف کمیاب است، محدود می‌شود (Yang et al., 2021). روش‌های یادگیری بدون نظارت به طور فزاینده‌ای برای شناسایی آنومالی‌های زمین‌شیمیایی به کارگرفته شده‌اند. این روش‌ها به ویژه در مدیریت مجموعه داده‌های بزرگ و با ابعاد بالا که ناهنجاری‌های برچسب‌گذاری‌شده در دسترس نیستند، بسیار مؤثر هستند (Xiong and Zuo, 2016).

روشهای یادگیری عمیق که زیر مجموعه‌ای از روشهای یادگیری ماشین هستند، به‌ویژه خودرمزنگارها وخودرمزنگارهای واریانسی، عملکرد برتری در شناسایی ناهنجاری‌های جزئی در مجموعه داده‌های زمین‌شیمیایی نشان داده‌اند که خودرمزنگارها برای بازسازی داده‌های ورودی آموزش داده می‌شوند و آنومالی‌ها بر اساس خطاهای بازسازی شناسایی می‌شوند (Yang et al., 2021). ژیانگ و ژو (Xiong and Zuo, 2016) از یک شبکه خودرمزگذار عمیق برای شناسایی آنومالی‌های زمین‌شیمیایی بر اساس خطاهای بازسازی استفاده کردند و به طور موفقیت‌آمیزی ناهنجاری‌های مرتبط با ذخایر آهن نوع اسکارن در جنوب‌غربی فوجیان چین را شناسایی کردند (Xiong and Zuo, 2016). چن و همکاران (Chen et al., 2019a) دوباره از یک شبکه خودرمزنگار عمیق برای جداسازی واحدهای آنومالی استفاده کردند؛ با این تفاوت که در این پژوهش ناهمگنی مکانی و فضایی نمونه‌ها در نظرگرفته می‌شود و سپس با استفاده از چندین خودرمزنگار ثبات جداسازی بالاتر می‌رود (Chen et al., 2019a). در پژوهش بعدی تنها از یک خودرمزنگار ساده استفاده نشده و در ابتدا از یک شبکه عمیق برای کاهش ابعاد غیر خطی استفاده شد و بعد نتایج به صورت باینری برای ماشین بردار پشتیبان استفاده شده است (Xiong and Zuo, 2020). در سال ۲۰۲۰، خودرمزنگار واریانسی برای جداسازی آنومالی از زمینه بر روی داده‌های زمین‌شیمی مورد استفاده قرارگرفت که در این شبکه از احتمال ساخت دوباره برای جداسازی آنومالی‌ها استفاده شد (Luo et al., 2020). یک نسخه از شبکه‌های متخاصم در سال ۲۰۲۱ توسط ژانگ و ژو (Zhang and Zuo, 2021) مورد استفاده قرارگرفت که در این شبکه از ترکیب شبکه‌های متخاصم با خودرمزنگار واریانسی استفاده شد. در همین سال، لو وهمکاران (Luo et al., 2021)، یک نسخه دیگر از شبکه‌های متخاصم را به نام GANomaly برای جداسازی آنومالی مورد استفاده قرار دادند که نمونه‌های آنومال بر اساس تفاوت در فضای لتنت شناسایی می‌شوند. ژانک و زو (Zhang and Zuo, 2024) برای جداسازی آنومالی، یک شبکه متخاصم خودنگار توسعه دادند که دانش و اطلاعات زمین‌شناسی را به شبکه برای بهبود جداسازی آنومالی اضافه می‌کند. با وجود موفقیت یادگیری عمیق در اکتشاف زمین‌شیمیایی، یک چالش کلیدی باقی‌مانده است: تنظیم ابرپارامترها. عملکرد مدل به شدت به مؤلفه‌هایی مانند تعداد لایه‌ها در یک خودرمزنگار، عامل آلودگی در جنگل ایزوله و اپسیلون در دی‌بی اسکن حساس است (Li et al., 2020). بیشتر پژوهش‌های قبلی بر انتخاب ابرپارامترهای مبتنی بر قوانین تجربی متکی هستند که در این پژوهش به نتایج ناسازگار منجر می‌شود. بهینه‌سازی بیزی رویکردی کارآمد برای بررسی نظام‌مند فضاهای ابرپارامتر ارائه می‌دهد. بر خلاف روش‌های جست‌وجوی شبکه‌ای یا تصادفی سنتی ، بهینه‌سازی بیزی یک مدل احتمالاتی از تابع هدف می‌سازد و به طور هوشمندانه ابرپارامترها را برای به بهینه‌سازی عملکرد انتخاب می‌کند. با استفاده از بهینه‌سازی بیزی ابرپارامترها، شکاف عمده‌ای در تشخیص آنومالی‌های زمین‌شیمیایی برطرف می‌شود که فقدان بهینه‌سازی نظام‌مند برای روش‌های مبتنی بر یادگیری عمیق است. بهینه‌سازی بیزی در این پژوهش به صورت نظام‌مند فضای هایپرپارامترها را جست‌وجو کرد و به یافتن تنظیم‌های بهینه با تعداد آزمون کمتر نسبت به روش‌های سنتی منجر شد. این روش با بهره‌گیری از مدل احتمالاتی گاوسی در هر گام، نقاط امیدبخش را برای ارزیابی بعدی انتخاب می‌کند و بنابراین کارآمدتر از جست‌وجوی شبکه‌ای یا تصادفی عمل می‌کند.

DOI: 10.22067/econg.2025.1157

با این حال، باید توجه داشت که هر گام بهینه‌سازی بیزی مستلزم برازش مدل کمکی است و در مسائل دارای فضای پیچیده هایپرپارامتر، هزینه محاسباتی آن قابل توجه است. همچنین بهینه‌سازی بیزی تضمین مطلقی برای یافتن بهینه سراسری ندارد و ممکن است در صورت تخمین نادرست تابع هدف، پیرامون بهینه‌های محلی متمرکز شود (مشکلی که در روش‌های تکاملی نیز مطرح است). با در نظرداشتن این ملاحظه‌ها، بهینه‌سازی بیزی را به دلیل ماهیت هدایت‌شده و استفاده مؤثر از مشاهده‌های گذشته انتخاب کردیم و مزیت آن در افزایش دقت مدل‌ها با وجود صرف زمان بیشتر، در نتایج نمایان‌شد. در سال‌های اخیر، رویکردهای یادگیری ماشین پیشرفته در نقشه‌برداری زمین‌شناسی و اکتشاف کانی‌زایی پررنگ‌تر شده است. فرهادی و همکاران (Farhadi et al., 2024) نشان داده‌اند که به‌کارگیری روش‌های تلفیقی و ترکیبی می‌تواند دقت طبقه‌بندی سنگ‌شناسی را به طور معناداری بهبود دهد و پورغلام و همکاران (Pourgholam et al., 2025) با یادگیری عمیق و آستانه‌گذاری فرکتال- اهداف مگنتیت–آپاتیت را در مقیاس ناحیه‌ای تفکیک کرده‌اند. در این پژوهش، با تمرکز بر تشخیص ناهنجاری نظارت نشده و مهندسی ویژگی مبتنی بر PCA/ICA، مسیری مکمل و کم‌نیاز به برچسب‌گذاری ارائه می‌کنیم که با ادغام لایه‌های چندمنبع نیز سازگار است؛ در حالی که بررسی‌های قبلی پتانسیل یادگیری عمیق و یادگیری ماشین را برای تشخیص آنومالی‌های زمین‌شیمیایی نشان داده‌اند. چندین شکاف کلیدی همچنان باقی‌مانده است که این شکاف‌ها شامل:

۱- مهندسی ویژگی: تجزیه و تحلیل‌های زمین‌شیمیایی سنتی اغلب بر غلظت‌های خام و نرمال شده عناصر متکی هستند و مزایای بالقوه تبدیل‌های پیشرفته ویژگی مثل ویژگی‌های چند جمله‌ای) را نادیده می‌گیرند.

۲- تنظیم ابرپارامترها: بیشتر بررسی‌ها از ابرپارامترهای پیش‌فرض استفاده می‌کنند که ممکن است برای مجموعه داده‌های خاص بهینه نباشد و به عملکرد غیربهینه منجر شود.

۳- ارزیابی مقایسه‌ای: بررسی‌های اندکی به طور نظام‌مند عملکرد مدل های یادگیری عمیق مختلف (مانند AE، VAE، DBSCAN، Isolation Forest) را با استفاده از یک چهارچوب بهینه‌سازی یکپارچه مقایسه می‌کنند.

در این پژوهش به جای آزمون همه حالت‌ها (جست‌وجوی شبکه‌ای) یا امتحان تصادفی چند حالت، از بهینه‌سازی بیزی استفاده کردیم. این روش از نتیجه آزمون‌های قبلی یاد می‌گیرد و حدس می‌زند که در مرحله بعد کدام تنظیم‌ها احتمالاً بهترند؛ بنابراین به آزمایش‌های کمتری نیاز است و زمان/ هزینه کاهش می‌یابد. از طرف دیگر، تابع هدف را طوری تعریف کردیم که فقط دقت مدل بالا نرود؛ بلکه از «بزرگ شدن بی دلیل محدوده‌های آنومالی» هم جلوگیری شود. این پژوهش با توجه به اهداف گسترش نظام‌مند ویژگی‌های زمین‌شیمیایی، بهینه‌سازی ابرپارامترهای AE، VAE، DBSCAN و Isolation Forest با استفاده از بهینه‌سازی بیزی، تولید نقشه‌های آنومالی زمین‌شیمیایی با استفاده از روش حوضه آبریز نمونه برای نمایش آنومالی‌ها و ارزیابی نقشه‌ها با نقاط کانی‌زایی شناخته شده در محل، در بخش‌های زیر ارائه می‌شود:

توصیف منطقه مورد بررسی و داده‌های مورد استفاده، بحث در مورد مهندسی ویژگی، مدل‌های تشخیص آنومالی و بهینه‌سازی بیزی، ارائه نتایج و ارزیابی عملکرد مدل، یافته‌ها، ارائه نکته‌های کلیدی و جهت‌گیری‌های پژوهشی آینده و در آخر نتیجه‌گیری و جمع‌بندی مقاله

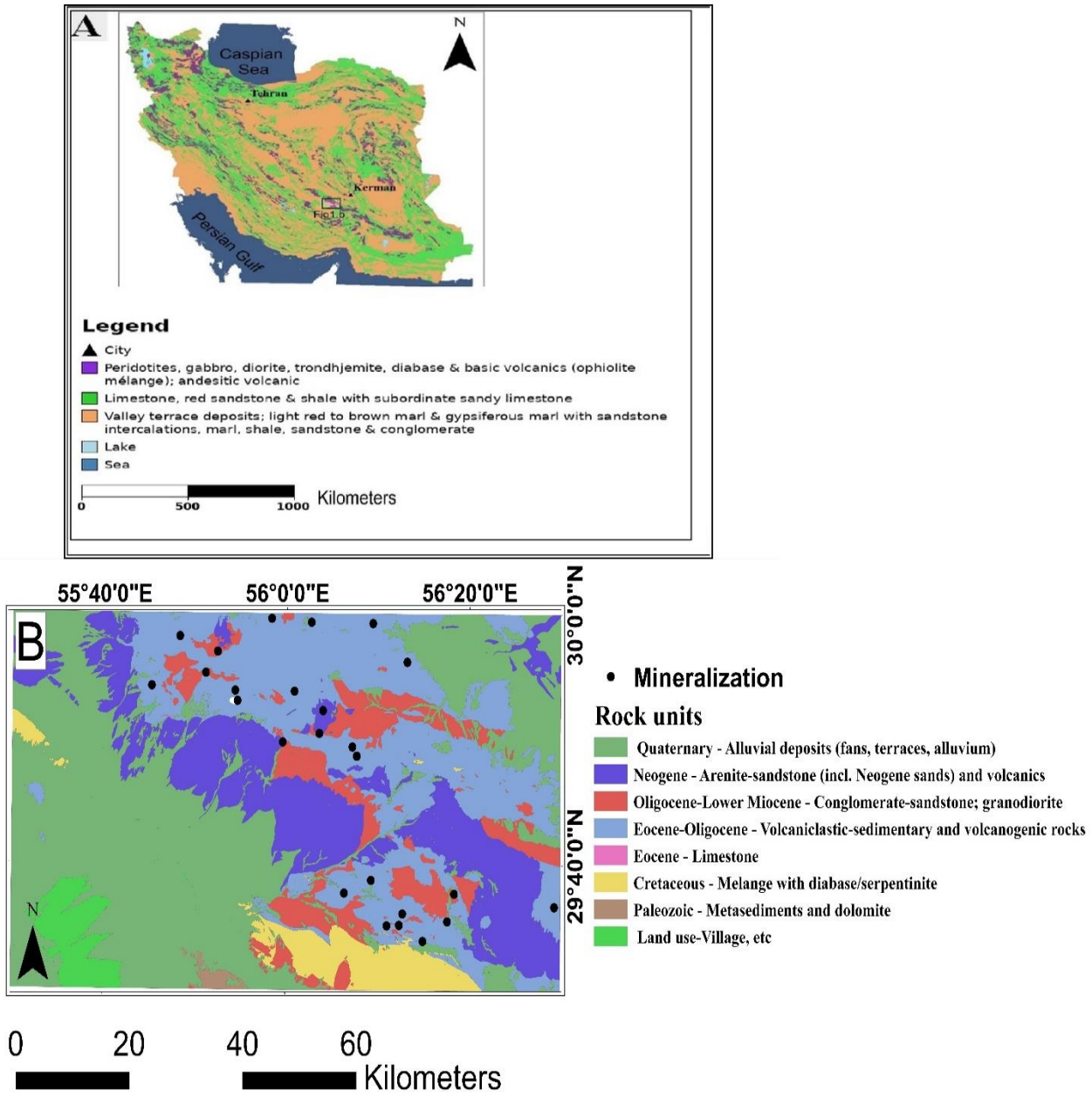
مواد و روش‌ها

زمین‌شناسی منطقه و داده‌ها

منطقه مورد بررسی ترکیبی از دو برگه زمین‌شناسی یکصدهزار پاریز و چهارگنبد است که نقشه زمین‌شناسی خلاصه‌شده این منطقه در شکل ۱-A و B نشان‌داده شده است. منطقه مورد بررسی در استان کرمان ایران، به دلیل سازندهای زمین‌شناسی مهم و ذخایر غنی معدنی، به ویژه رخدادهای پورفیری مس مشهور است. این

DOI: 10.22067/econg.2025.1157

منطقه شامل سایت‌های معدنی قابل‌توجهی مانند سرچشمه، میدوک و چندین معدن دیگر است که به زمین‌شناسی اقتصادی آن کمک می‌کنند. این ناحیه بخشی از کمربند آتشفشانی ایران مرکزی است که به خاطر فعالیت‌های گسترده ماگمایی و متالوژنی شناخته‌شده است. چهارچوب زمین‌شناسی منطقه با تعامل پیچیده‌ای از فرایندهای زمین‌ساختی، ماگمایی و رسوبی مشخص می‌شود.



شکل ۱. A: موقعیت منطقه مورد بررسی در کشور ایران و B: نقشه زمین‌شناسی ساده شده از منطقه پاریز و چهارگنبد

Fig. 1. A: Position of study area in Urmiyeh Dokhtar magmatic arc, and B: Simplified geological map of the Pariz and Chargonbad area

منطقه مورد بررسی در داخل کمان ماگمایی ارومیه- دختر قرار دارد که یک ویژگی تکتونوماگمایی برجسته در ایران است. این کمان به دلیل فرورانش پوسته اقیانوسی نئوتتیس در زیر پوسته ایران مرکزی در طی دوره‌های کرتاسه پسین تا میوسن تشکیل‌شده است. مشخصات زمین‌ساختی نقشی مهم در جای‌گیری سامانه‌های پورفیری مس در منطقه ایفا کرده است (Berberian and King, 1981). فعالیت ماگمایی در منطقه پاریز اغلب با نفوذهای کالک‌آلکالن تا شوشونیتی مرتبط است که درون سنگ‌های رسوبی و آتشفشانی قدیمی‌تر نفوذ کرده‌اند. این نفوذها معمولاً دارای سن الیگوسن تا میوسن هستند. ذخایر پورفیری مس در این منطقه از جمله سرچشمه و میدوک، با این رخدادهای نفوذی مرتبط هستند. تکامل ماگمایی با چندین مرحله نفوذ، دگرسانی و کانی‌سازی مشخص می‌شود (Richards, 2003). چینه‌شناسی منطقه شامل انواع مختلفی از سنگ‌هاست که از سنگ‌های دگرگونی پرکامبرین تا واحدهای آتشفشانی و رسوبی دوره سوم را در بر می‌گیرد. واحدهای لیتولوژیکی غالب عبارتند از: ۱- سنگ‌های پی‌سنگ پرکامبرین شامل گنیس‌ها، شیست‌ها و آمفیبولیت‌ها که مجموعه پی‌سنگ منطقه را تشکیل می‌دهند. ۲- سنگ‌های رسوبی پالئوزوئیک تا مزوزوئیک شامل سنگ‌آهک‌ها، ماسه‌سنگ‌ها و شیل‌ها که به صورت محلی توسط توده‌های آذرین نفوذ کرده‌اند. ۳- سنگ‌های آتشفشانی و پلوتونیک دوره سوم اغلب شامل سنگ‌های آتشفشانی آندزیتی تا داسیتی و نفوذهای گرانودیوریتی تا دیوریتی که کانی‌سازی پورفیری مس را در خود جای داده‌اند. پس در گستره دو برگه ۱:۱۰۰۰۰۰ پاریز و چهارگنبد واحدهای عمده عبارتند از:

آتشفشانی– پیروکلاستیک ائوسن (آندزیت، لاتیت، داسیت و توف‌های مرتبط) که سنگ میزبان غالب هستند؛

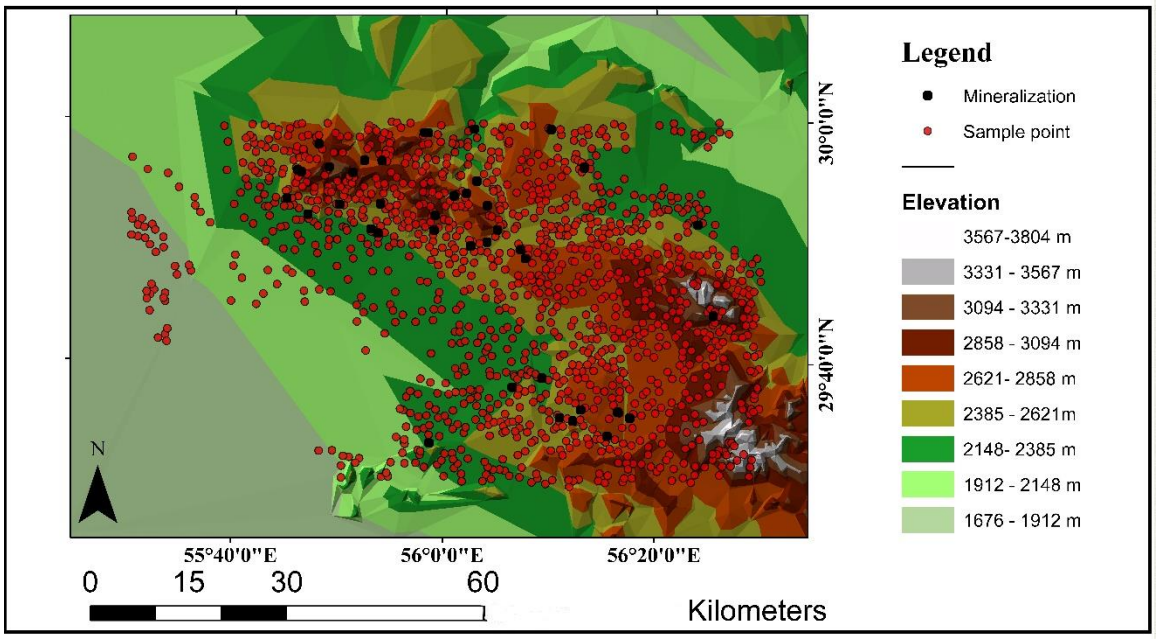
نفوذی‌های میوسن با ترکیب‌های گرانودیوریتی، دیوریتی و کوارتزمونزونیتی به صورت استوک‌ها و دایک‌های حدواسط؛

رخساره‌های رسوبی/ کربناته محلی که در تماس با نفوذی‌ها، زمینه اسکارن‌های Cu (±Fe, ±Zn) را مهیا می‌کنند.

الگوی دگرسانی گزارش‌شده در حواشی توده‌ها شامل زون پتاسیک در هسته، فیلیک در زون میانی و پروپیلیتیک در حاشیه است. آرژیلیک پیشرفته به صورت محدود مشاهده می‌شود، الگویی که با سامانه‌های پورفیری منطقه همخوانی دارد(Hezarkhani, 2006). از نظر هندسه زهکشی، آبراهه‌های رده ۲ تا ۴ در دامنه‌های جنوبی- شرقی نسبت به شکستگی‌های مایل NE–SW هم‌گرایی نشان می‌دهند که این موضوع در نمایش فضایی نتایج لحاظ شده است.

این منطقه میزبان چندین ذخیره پورفیری مس در سطح جهانی است که سرچشمه و میدوک از مهم‌ترین آنها هستند. سرچشمه یکی از بزرگ‌ترین و مهم‌ترین ذخایر پورفیری مس در جهان است. این ذخیره با یک سامانه گسترده از نفوذهای پورفیری کوارتز مونزونیتی مشخص می‌شود که تحت دگرسانی شدید گرمابی و کانی‌سازی مس قرار گرفته‌اند. مناطق دگرسانی شامل انواع پتاسیک، فیلیک، آرژیلیک و پروپیلیتیک هستند که هر کدام با مجموعه‌های کانی‌شناسی و محتوای فلزی متفاوتی همراه هستند (Hezarkhani, 2006). میدوک یکی دیگر از ذخایر بزرگ پورفیری مس در منطقه پاریز است. مشابه سرچشمه، این ذخیره دارای مجموعه‌ای از نفوذهای گرانودیوریتی تا دیوریتی با دگرسانی گرمابی قابل توجه و کانی‌سازی مس است. در این ذخیره یک الگوی زون‌بندی برای دگرسانی‌ها و کانی‌سازی وجود دارد؛ به طوری‌که زون پتاسیک در مرکز قرارگرفته است و در اطراف آن به ترتیب دگرسانی‌های فیلیک و پروپیلیتیک وجود دارد (Taghipour et al, 2008). موقعیت منطقه مورد بررسی و نقشه زمین‌شناسی ساده‌شده در شکل ۱ نشان‌داده شده است. در مجموع ۱۲۳۸ نمونه رسوب‌های آبراهه‌ای توسط سازمان زمین‌شناسی و اکتشافات معدنی کشور از این منطقه جمع‌آوری شد. این نمونه‌ها در این پژوهش برای جداسازی آنومالی‌ها از زمینه مس و اکتشاف ذخایر پورفیری مس استفاده شدند. شکل ۲، موقعیت نقاط نمونه‌برداری را نسبت به شرایط آبراهه‌ای و وضعیت توپوگرافی منطقه نشان می‌دهد که منبع شرایط آبراهه‌ای، تصویرهای رقومی ارتفاعی از منطقه است.

DOI: 10.22067/econg.2025.1157



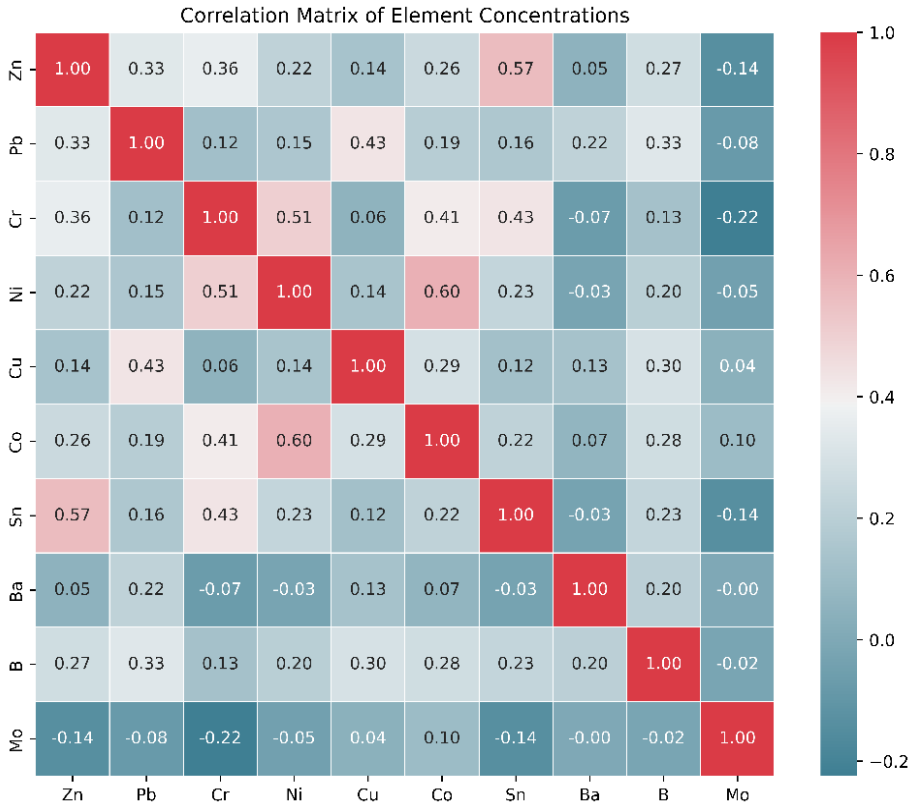
شکل ۲. موقعیت نقاط نمونه‌برداری نسبت به شرایط توپوگرافی در منطقه پاریز و چهارگنبد

Fig. 2. Location of sampling points in relation to topographical conditions in the region of Pariz and Chargonbad

مرحله پیش‌پردازش داده‌ها شامل جایگزینی مقادیر نامشخص و شناسایی نقاط پرت با استفاده از روش دامنه بین چارکی انجام‌شد (Verma, 1997). پس از انجام این مراحل برای نرمال‌سازی مقادیر عنصری، ابتدا یک تبدیل لگاریتمی اعمال‌شد و سپس نرمال‌سازی کمینه- بیشینه بر روی مقادیر عنصری انجام‌شد (Praveena et al., 2011). سپس، ضریب همبستگی که در محدوده -۱ تا ۱ قرار دارد و مقدار ۱ نشان‌دهنده بالاترین همبستگی است، برای عناصر محاسبه‌شد. این مقادیر در شکل ۳ نشان‌داده شده‌اند و همچنین در شکل ۴ نمودار دندوگرام مربوط به عناصر با روش وارد نشان‌داده شده است.از این دو تصویر می‌توان رابطه پیچیده مس که عنصر موردنظر است را با دیگر عناصر و همچنین رابطه غیرخطی پیچیده، نهان و الگوهای پنهان بین عناصر را در داخل داده‌ها نتیجه‌گرفت.

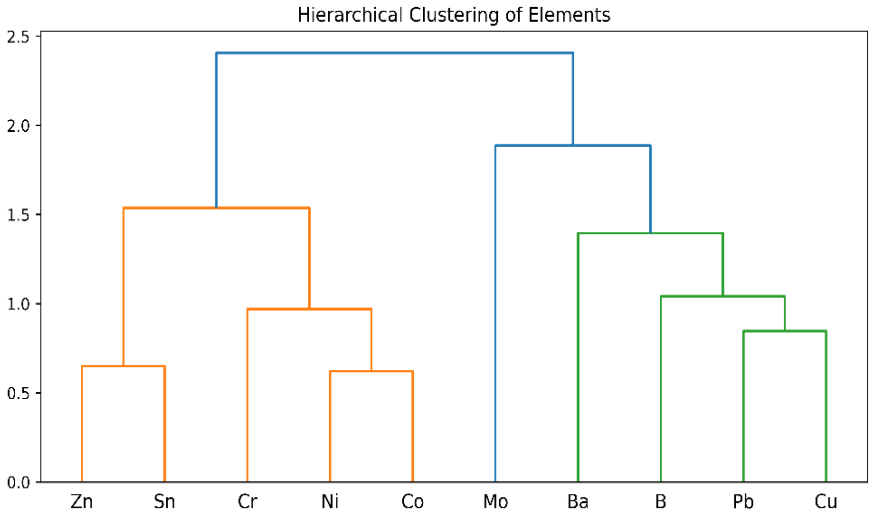
ادغام و به کارگیری روش‌هایی مانند ویژگی‌های چندجمله‌ای، تحلیل مؤلفه‌های مستقل و تحلیل مؤلفه‌های اصلی در تجزیه و تحلیل داده‌های زمین‌شیمیایی، امکانات تحلیلی و تفسیری متعددی را فراهم می‌کند. منابعی مانند یاسوکاوا و همکاران (Yasukawa et al., 2016)، ایواموری و همکاران (Iwamori et al., 2017) و وین و همکاران (Wein et al., 2020) بر کارایی تحلیل مؤلفه‌های مستقل در شناسایی نشانگرهای زمین‌شیمیایی پنهان، جداسازی ویژگی‌های مستقل و مدیریت خوشه‌بندی در مجموعه داده‌های با ابعاد بالا تأکید می‌کنند. ادغام تحلیل مؤلفه مستقل به پژوهشگران امکان می‌دهد تا الگوهای نهفته را کشف کنند، تجسم داده‌ها را بهبود بخشند و ساختارهای نهان درون مجموعه داده‌های زمین‌شیمیایی را تشخیص دهند. علاوه بر این، پژوهش‌های دیگری مزایای استفاده از تحلیل مؤلفه‌های مستقل و اصلی را در حوزه‌های استخراج ویژگی، جداسازی سیگنال و تفکیک داده‌ها مورد بررسی قرار داده‌اند که مشخص شد این روش‌ها به پژوهشگران امکان می‌دهد سیگنال‌های مخلوط را تفکیک‌کنند، ویژگی‌های متمایز را شناسایی‌کنند و قابلیت تفسیر داده‌های زمین‌شیمیایی را بهبود بخشند (Li et al., 2011; Adankon et al., 2012; Jin et al., 2019).

DOI: 10.22067/econg.2025.1157



شکل ۳. ضریب همبستگی محاسبه‌شده برای عناصر عناصر منطقه پاریز و چهارگنبد

Fig. 3. The calculated correlation coefficient for the elements Pariz and Chargonbad area



شکل ٤. دندوگرام رسم‌شده مربوط به عناصر به روش وارد در منطقه پاریز و چهارگنبد

Fig. 4. The Hierarchical clustering for elements of Pariz and Chargonbad area

DOI: 10.22067/econg.2025.1157

با به کارگیری تجزیه و تحلیل مؤلفه‌های مستقل، پژوهشگران قادر می‌شوند منابع داده‌ای پیچیده را بازگشایی کنند، مؤلفه‌های اساسی زیرین را آشکار سازند و درک زمین‌شیمی رسوب‌ها را پیشرفت دهند (Li et al., 2011; Adankon et al., 2012; Jin et al., 2019). به همین دلیل، در این پژوهش تحلیل مؤلفه‌های اصلی نیز بر روی عناصر یادشده در شکل ۳ اعمال‌شد و پنج مؤلفه اصلی به جدول اولیه اضافه‌شد. در ادامه، تحلیل مؤلفه‌های مستقل دوباره بر روی همان عناصر یادشده انجام‌شد و پنج مؤلفه به داده‌ها اضافه شد. پس از این دو تحلیل، ویژگی‌های چند جمله‌ای با درجه ۲ و بدون در نظرگرفتن بایاس، بر روی عناصری که با مس همبستگی دارند، اعمال‌شد و نتایج به جدول اولیه اضافه‌شد. با توجه به حضور فرایندهای اختلاط، جذب سطحی در رسوب‌های آبراهه‌ای، روابط بین برخی عناصر نسبت به غلظت‌های منفرد غیرخطی است. از این رو، برای عناصر هم‌بسته با Cu، ویژگی‌های چند جمله‌ای درجه ۲ ساخته‌شد تا برهم‌کنش‌ها بهتر مدل شود.ویژگی‌های مهندسی‌شده در داده (شامل مؤلفه‌های اصلی، مؤلفه‌های مستقل و ویژگی‌های چند جمله‌ای درجه ۲ برای عناصر همبسته با مس) هر کدام بر اساس یک منطق زمین‌شیمیایی انتخاب شده‌اند. مؤلفه‌های اصلی و مستقل با فشرده‌سازی اطلاعات چندعنصری، امکان نمایان‌شدن الگوهای پنهان زمین‌شیمیایی را فراهم می‌کنند (مثلاً مؤلفه‌های اصلی شماره ۱، می‌تواند بیانگر روند غنی‌شدگی کلی مرتبط با کانی‌سازی باشد و مؤلفه‌های مستقل فرایندهای مستقل زمین‌شناختی نظیر تأثیر لیتولوژی‌ها یا دگرسانی‌ها را تفکیک‌کنند). افزودن ویژگی‌های چند جمله‌ای (مانند Cu^2، Cu×Mo و ...) نیز به این دلیل انجام‌شد که بسیاری از روابط زمین‌شیمیایی در محیط‌های رسوبی– گرمابی غیرخطی هستند (برای نمونه، افزایش هم‌زمان Cu و Mo می‌تواند نشان‌دهنده بخش مرکزی‌تر یک سامانه پورفیری باشد و تنها Cu یا Mo به تنهایی این تمایز را به خوبی نشان ندهند). بنابراین، این مجموعه ویژگی‌ها کمک می‌کند تا مدل بتواند هم الگوهای خطی (از طریق عناصر اصلی) و هم پدیده‌های پیچیده‌تر (از طریق برهم‌کنش‌های درجه ۲) را بیاموزد و آنومالی‌های ناشی از فرایندهای طبیعی مختلف را بهتر تفکیک‌کند. انتظار می‌رود، مؤلفه‌هایی که بیشترین واریانس یا اطلاعات مستقل را توضیح می‌دهند (مثل PCAهای اولیه) بیشترین تأثیر را در جداسازی آنومالی داشته باشند؛ در حالی‌که ویژگی‌های چند جمله‌ای خاص مانند Cu×Mo، ممکن است سیگنال‌های ظریف مرتبط با کانی‌سازی مشترک عناصر را تقویت‌کنند. در جدول ۱ مشخصات آماری عناصر مورد استفاده آمده است.

جدول ۱. مؤلفه‌های آماری عناصر هدف در رسوب‌های آبراهه‌ای منطقه پاریز و چهارگنبد

Table 1. Statistical parameters of the target elements in stream-sediment samples of Pariz and Chargonbad area

Element	min	q1	median	q3	max	mean	std	skew	kurtosis
B	2.00	21.00	30.00	45.00	241.00	36.05	25.57	2.49	10.64
Ba	15.00	308.00	418.50	531.00	4650.00	431.60	223.05	5.94	103.50
Co	2.00	18.00	25.00	35.00	247.00	27.87	15.82	3.66	37.07
Cr	7.00	58.00	83.50	128.00	1200.00	113.53	102.38	3.61	21.34
Cu	2.00	46.00	61.00	82.00	1200.00	77.92	90.33	7.87	78.92
Mo	0.10	1.00	2.00	2.00	52.00	2.03	2.56	10.57	161.20
Ni	2.00	35.00	50.00	80.00	1200.00	64.64	59.17	7.14	113.36
Pb	2.00	13.00	19.00	32.00	389.00	27.28	33.51	6.05	49.63
Sn	5.00	5.00	5.00	5.00	171.00	10.40	14.71	5.06	36.94
Zn	2.00	68.25	94.00	131.00	752.00	110.87	73.68	2.79	13.40

DOI: 10.22067/econg.2025.1157

مدل‌های مورد استفاده برای جداسازی آنومالی

در این پژوهش از چهار مدل جنگل ایزوله، دی‌بی اسکن، خودرمزنگار، خودرمزنگار واریانسی برای جداسازی آنومالی از زمینه زمین‌شیمیایی در منطقه مورد بررسی استفاده شد که در ادامه در مورد کلیات و مراحل اجرای این مدل‌ها و نحوه انتخاب بهترین هایپرپارامترها برای این مدل‌ها توضیحاتی ارائه خواهد شد.

روش جنگل ایزوله

روش جنگل ایزوله بر اساس این ایده است که نمونه‌های آنومال خیلی آسان‌تر از بقیه نمونه‌ها جدا می‌شوند. نمونه‌های آنومال معمولاً دورتر از بقیه نمونه‌ها و در نقاطی کم چگالی‌تر قرار گرفته‌اند. الگوریتم با ساخت چندین درخت تصمیم‌گیری (که به آنها درخت‌های ایزوله گفته می‌شود) عمل می‌کند. این درخت‌ها با انتخاب تصادفی ویژگی‌ها ساخته می‌شوند (Hariri et al., 2021; Bulut et al., 2024). سپس با تقسیم‌بندی بازگشتی مجموعه داده، نمونه‌ها را جدا می‌کند. مشاهده کلیدی این است که نمونه‌های آنومال آسان‌تر ایزوله می‌شوند (به تقسیم‌های کمتری نیاز دارند)؛ زیرا آنها دور از خوشه‌های متراکم نمونه‌ها قرار دارند. مراحل انجام این الگوریتم را به صورت زیر است (Hariri et al., 2021; Bulut et al., 2024):

در مرحله اول، الگوریتم جنگل ایزوله به طور تصادفی یک ویژگی را انتخاب می‌کند و سپس به طور تصادفی یک مقدار تقسیم‌بندی بین کمینه و بیشینه مقادیر آن ویژگی انتخاب می‌کند. این فرایند به صورت بازگشتی ادامه می‌یابد و بخش‌بندی‌های تصادفی از داده‌ها ایجاد می‌کند.

در مرحله دوم، هر درخت ایزوله با تقسیم بازگشتی داده‌ها به دو بخش بر اساس یک ویژگی تصادفی و نقطه تقسیم تصادفی ایجاد می‌شود. این تقسیم‌بندی ادامه می‌یابد تا هر نقطه در بخش خودش ایزوله شود یا تا زمانی که یک شرط توقف برقرار شود (مثلاً یک عمق بیشینه برای درخت تعیین‌شده باشد) (Hariri et al., 2021; Bulut et al., 2024).

در مرحله سوم، طول مسیر نمونه تعداد تقسیم‌هایی است که برای ایزوله‌کردن آن نمونه نیاز است؛ به طور مثال، نمونه‌های آنومال خارج از ردیف در مناطق کم چگال‌تر قرار دارند؛ به همین دلیل با تعداد تقسیم‌های کمتری ایزوله می‌شوند و در نتیجه طول مسیر کوتاه‌تری را دارا هستند. ولی نمونه‌های نرمال توسط نمونه‌های دیگری احاطه شده‌اند؛ در نتیجه به تقسیم‌های بیشتری نیازمند بوده و طول مسیر آنها بیشتر است (Hariri et al., 2021; Bulut et al., 2024).

در مرحله چهارم، الگوریتم یک امتیاز آنومالی را بر اساس طول مسیر متوسط نمونه در میان همه درخت‌های ایزوله محاسبه می‌کند که امتیاز آنومالی نمونه $s(p)$ به صورت رابطه ۱ محاسبه می‌شود؛ ولی نکته مهم در این پژوهش، طول مسیر متوسط برای هر نمونه است که هرچه این عدد بزرگ‌تر باشد، نشان‌دهنده نمونه نرمال و هرچه کوچک‌تر باشد، نشان‌دهنده نمونه آنومال است (Hariri et al., 2021; Bulut et al., 2024).

رابطه ۱:

$$S(p) = 2^{\frac{-E(h(p))}{c(n)}}$$

که $E(h(p))$ عبارت است از طول مسیر میانگین در نمونه برای همه درخت‌های ایزوله و $c(n)$ عبارت است از (رابطه ۲):

رابطه ۲:

$$c(n) \approx 2H(n-1) - \frac{2(n-1)}{n}$$

و $H(n)$ عدد هارمونیک بوده و توسط فرمول زیر (رابطه ۳) قابل دستیابی است:

رابطه ۳:

$$H(n) \approx \ln(n) + 0.577215$$

پس $s(p)$ یا امتیاز آنومالی نزدیک به یک نشان‌دهنده یک نمونه آنومال و امتیاز آنومال نزدیک به صفر نشان‌دهنده یک نمونه نرمال است؛ ولی با توجه به رابطه این امتیاز یا رابطه شماره ۴، هرچه $E(h(p))$ بزرگ‌تر باشد، امتیاز آنومالی کوچک‌تر و نمونه نرمال است (Hariri et al., 2021; Bulut et al., 2024). این روش نقاطی را که زودتر ایزوله می‌شوند، به عنوان ناهنجار تشخیص

DOI: 10.22067/econg.2025.1157

می‌دهد، مقیاس‌پذیر است و معمولاً آنومالی را به صورت فشرده و هدفمند نشـان می‌دهد. با این حال، به مؤلفه نرخ ناهنجاری؛ سـهم مورد انتظار نقاط آنومال) و اندازه نمونه هر درخت حساس است. اگر این نرخ نامناسـب انتخاب شـود، ممکن اسـت هاله‌های کم ضعیف یا گسترده به خوبی بازیابی نشوند.

دی‌بی اسکن

الگوریتم دی‌بی اسـکن و یا خوشـه‌بندی مبتنی بر تراکم فضـایی برنامه‌های کاربردی با نویز به عنوان ابزاری قدرتمند برای تشخیص آنومالی‌ها ظاهرشـده اسـت؛ زیرا به طور ذاتی توانایی شـناسـایی داده‌های پرت حین خوشه‌بندی داده‌ها را دارد. بر خلاف روش‌های مبتنی بر مرکز دسته مانند k-means که برای تشخیص ناهنجاری‌ها نیاز به پردازش پس از اجرا دارند، دی‌بی اسـکن به طور ذاتی نقاط نویز را در طول فرایند خوشـه‌بندی با اختصـاص‌دادن آنها به یک شناسه خوشه خاص (۱-) جدا می‌کند (Thu and Armitage, 2008; Andriyani and Puspitarani, 2022). این عملکرد دوگانه، آن را برای کاربردهایی که نیاز به خوشه‌بندی و شناسایی داده‌های پرت هم‌زمان دارند، ارزشمند می‌سازد. این روش توسط ایسـتر و همکاران (Ester et al., 1996) گسـترش‌یافته اسـت. دی‌بی اسـکن بر اسـاس اصـل تراکم عمل می‌کند و این امر آن را برای شـناسـایی آنومالی‌ها در داده‌های مکانی و زمانی مؤثر می‌سازد. با تفکیک بین نقاط هسته متراکم و نقاط نویزی پراکنده، دی‌بی اسـکن بدون نیاز به داشـتن اطلاعات قبلی درباره تعداد یا شـکل‌های خوشـه‌ها، یک چهارچوب قوی برای شـناسـایی بی‌نظمی‌ها فراهم می‌کند (Ester et al., 1996).

الگوریتم دی‌بی اسـکن نقاط داده را به سـه دسـته تقسـیم می‌کند: نقاط هسـته، نقاط مرزی و نقاط نویز. این طبقه‌بندی به دو مؤلفه حیاتی وابسته است: epsilon (ε): شعاع تعریف‌کننده همسایگی یک نقطه و minPts: کمترین تعداد نقاط مورد نیاز برای تشـکیل یک ناحیه متراکم.

نقاط هسته: یک نقطه p به عنوان نقطه هسته تعیین می‌شود؛ اگر کمینه به تعداد minPts نقطه (شـامل خودش) در همسـایگی ε آن قرار داشته باشند. نقاط هسته به عنوان پایه‌ای برای تشکیل خوشه عمل می‌کنند و به عنوان دانه‌هایی هسـتند که خوشـه‌ها از آنها گسترش می‌یابند (Ester et al., 1996).

نقاط مرزی: این نقاط در محدوده ε- همسایگی یک نقطه هسته قرار دارند؛ اما به اندازه کافی همسایه ندارند تا به عنوان نقاط هسته واجد شـرایط باشـند. نقاط مرزی به خوشـه ها تعلق دارند؛ اما نمی توانند عضویت در خوشه را به سایر نقاط انتقال دهند (Ester et al., 1996).

نقاط نویز: نقاطی که نه مرکز هسـتند و نه مرزی، به عنوان نویز برچسب‌گذاری می‌شوند. این نقاط پرت در مناطق کم‌چگالی قرار دارند و از تمام خوشه‌ها جدا هستند (Ester et al., 1996).

نحوه کار دی‌بی اسکن از طریق جست‌وجوی تکراری همسایگی و گسترش خوشه ها انجام می شود: الگوریتم با انتخاب یک نقطه دیده نشـده و محاسـبه همسـایگی در فاصـله اپسـیلون آن شـروع می‌شـود. اگر نقطه در همسـایگی خود کمینه minPts نقطه داشته باشد، نقطه به عنوان هسته علامت‌گذاری شده و یک خوشه جدید آغاز می‌شـود. در غیر این صـورت، به طور موقت به عنوان نویز برچسـب‌گذاری می‌شـود (Ester et al., 1996). برای هر نقطه هسته، الگوریتم به طور مرتب نقاط در فاصله اپسیلون از خود را بررسی می کند و نقاط واجد شرایط را به خوشه اضافه می کند. این فرایند تا زمانی که هیچ نقطه دیگری نتواند ادغام شـود، ادامه می یابد. نقاطی که در ابتدا به عنوان نویز برچسب گذاری شده اند، اگر در محدوده ε از یک نقطه هسته‌ای از خوشه دیگر قرار گیرند، ممکن است بعد به عنوان نقاط مرزی طبقه‌بندی شوند. این رویکرد تکراری به دی‌بی اسـکن اجازه می‌دهد تا به طور پویا خود را با سـاختار ذاتی داده‌ها تطبیق دهد و تراکم نقاط را بر فرضـیه‌های هندسـی اولویت دهد. ویژگی جداسـازی آنومالی از زمینه این الگوریتم از این ویژگی جداسازی نقاط نویز آن نشأت می‌گیرد (Ester et al., 1996).

DOI: 10.22067/econg.2025.1157

از آنجا که نقاط نویز در مناطق کم تراکم قرار دارند، به طور ذاتی با بیشتر نقاط داده که در خوشه‌های متراکم تجمع می‌کنند، متفاوت هستند. این ویژگی با تعریف آنومالی‌ها به عنوان مشاهداتی که به طور قابل توجهی از نرمال منحرف می شوند، همسو است. کارایی دی‌بی اسکن در تشخیص آنومالی‌ها به انتخاب مؤلفه‌های مناسب بستگی دارد؛ در حالی که ε و minPts وابسته به داده هستند (Ester et al., 1996). دی‌بی اسکن خوشه‌ها را بر اساس چگالی نقاط پیدا می‌کند و نقاط پراکنده را «نویز» می‌نامد.دو مؤلفه کلیدی آن eps (شعاع همسایگی) و کمینه همسایه (minPts) است و به هر دو حساس است. در نواحی با چگالی پس‌زمینه ناهمگون، انتخاب نادرست eps می‌تواند به گسترش بیش از اندازه پهنه‌ها یا بر عکس تکه‌تکه‌شدن خوشه‌ها منجرشود. این روش برای آنومالی‌های پیوسته و کشیده (مثلاً دنبال‌شونده در امتداد آبراهه‌ها) مفید است، به‌شرط آنکه مقیاس داده‌ها استاندارد و مؤلفه‌ها به درستی تنظیم شوند.

خودرمزنگار عمیق

خودرمزنگارها، یک دسته از شبکه‌های عصبی بدون نظارت هستند که به دلیل کارایی بالای آنها در شناسایی ناهنجاری‌ها در حوزه‌های مختلف از جمله تحلیل داده‌های زمین‌شیمیایی در اکتشاف مواد معدنی توجه زیادی را به خود جلب کرده‌اند. توانایی آنها در مدل‌سازی توزیع‌های پیچیده و با ابعاد بالا، آنها را به ویژه برای شناسایی آنومالی‌های ضعیف زمین‌شیمیایی که نشان‌دهنده کانی‌زایی هستند، مناسب می‌سازد. یک خودرمزنگار از دو مؤلفه اصلی تشکیل‌شده است: یک رمزگذار و یک رمزگشا. رمزگذار داده‌های ورودی را به یک نمایش فشرده در فضای نهفته تبدیل می‌کند و ویژگی‌های اساسی داده را استخراج می‌کند. سپس رمزگشا داده‌های اصلی را از این نمایش فشرده بازسازی می‌کند. این شبکه به گونه‌ای آموزش داده می‌شود که اختلاف بین ورودی و خروجی بازسازی‌شده را به کمینه کند که معمولاً از طریق یک تابع هزینه مانند میانگین مربع خطا انجام می‌شود (Feng et al., 2022). از نظر ریاضی می‌توان به این شکل بیان‌کرد که اگر $x\in R^n$ به عنوان داده ورودی باشد، تابع رمزگذار f ورودی x را به صورت رابطه ۴، به فضای پنهان $z\in R^m$ تصویر می‌کند؛ در صورتی‌که m کوچک‌تر از n است. در رابطه ۴، W_e وزن‌های خودرمزنگار و b_e بایاس خودرمزنگار و σ نشان‌دهنده تابع فعال‌ساز است (Chen et al., 2019b):

رابطه ۴: $z=f(x)=\sigma(W_e x+b_e)$

سپس تابع رمزگشای g مقادیر فضای پنهان را به صورت رابطه ۵ به مقادیر ساخته شده $\hat{x}$ تصویر می‌کند (Chen et al., 2019b):

رابطه ۵: $\hat{x}=g(z)=\sigma(W_d z+b_d)$

در حالی‌که، W_d و b_d وزن‌ها و بایاس شبکه رمزگشا هستند که هدف در این مدل کم‌کردن خطای بازسازی داده‌های ورودی $L(x,\hat{x})$ است و اغلب به صورت رابطه ۶ تعریف می‌شود (Chen et al., 2019b):

رابطه ۶: $L(x,\hat{x})=\|x-\hat{x}\|^2$

در زمینه اکتشاف زمین‌شیمیایی، شناسایی آنومالی‌ها برای تعیین ذخایر معدنی از اهمیت زیادی برخوردار است. داده‌های زمین‌شیمیایی اغلب به دلیل فرایندهای زمین‌شناسی مختلف، توزیع‌های پیچیده و چندمتغیره‌ای را نشان می‌دهند. خودرمزنگارها می‌توانند این توزیع‌ها را با یادگیری الگوهای پایه‌ای زمینه زمین‌شیمیایی مدل‌سازی کنند. نمونه‌هایی که از این زمینه یادگرفته‌شده انحراف دارند، دارای خطای بازسازی بالاتری خواهند بود و به عنوان نمونه‌های آنومال در نظرگرفته می‌شوند . ژیانگ و ژو (Xiong and Zuo, 2016)، کاربرد شبکه‌های خودرمزنگار عمیق را در شناسایی آنومالی‌های چندمتغیره زمین‌شیمیایی نشان دادند. آنها یک خودرمزنگار را بر روی داده‌های زمین‌شیمیایی آموزش دادند تا توزیع زمینه را بازسازی کند. نمونه‌هایی که دارای خطای بازسازی بالاتری بودند، به عنوان آنومالی شناسایی شدند و با مناطق کانی‌سازی‌شده شناخته‌شده، همبستگی داشتند. این رویکرد به طور مؤثری زمینه زمین‌شیمیایی را از آنومالی‌های مرتبط با کانی‌سازی چندفلزی آهن تفکیک‌کرد

DOI: 10.22067/econg.2025.1157

(Xiong and Zuo, 2016). بر همین اساس، چن و همکاران (Chen et al., 2019b) رویکردی مبتنی بر خودرمزنگارهای کانوولوشنی را برای استخراج الگوهای ساختاری فضایی داده‌های زمین‌شیمیایی معرفی کردند. با بهره‌گیری از لایه‌های کانولوشنی، خودرمزنگار توانست وابستگی‌های فضایی را به‌طور مؤثری یاد بگیرد و شناسایی آنومالی‌های زمین‌شیمیایی را بهبود بخشد. این روش در جنوب‌غربی استان فوجیان چین مورد ارزیابی قرار گرفت، جایی که آنومالی‌های شناسایی‌شده همبستگی فضایی قوی با ذخایر شناخته‌شده آهن نشان دادند (Chen et al., 2019b)، مهم‌ترین معیار برای شناسایی آنومالی‌ها با استفاده از خودرمزنگارها، خطای بازسازی است. برای یک نمونه داده x، پس از به دست‌آوردن خروجی بازسازی‌شده x^، خطای بازسازی RE به صورت رابطه ۷ محاسبه می‌شود (Chen et al., 2019b):

رابطه ۷: $RE=\|x-x^\|^2$

بعد از به دست آوردن خطای بازسازی، برای مشخص کردن حد آستانه برای جدا کردن نمونه‌های آنومال می‌توان از توزیع خطاهای بارسازی و یا روش‌های طبقه‌بندی دیگر استفاده کرد. در این پژوهش، خودرمزنگار با فرض اینکه بیشتر نمونه‌ها متعلق به دسته زمینه هستند، آموزش خودرمزنگار بر روی تمام داده‌ها صورت گرفت. خودرمزنگار با فشرده‌سازی و بازسازی داده‌ها، خطای بازسازی را به عنوان امتیاز ناهنجاری به‌کار می‌گیرد. هرچه بازسازی بدتر باشد، احتمال آنومالی بیشتر است. به سبب توانایی در یادگیری روابط غیرخطی، برای داده‌های زمین‌شیمی چندمتغیره مناسب است و معمولاً بین پوشش رخدادها و کنترل مساحت آنومالی، توازن خوبی ایجاد می‌کند. محدودیت اصلی آن نیاز به تنظیم معماری شبکه و اعمال قوانینی مانند توقف زودهنگام برای جلوگیری از بیش‌برازش است.

خودرمزنگار واریانسی

خودرمزنگارهای واریانسی یک مدل توسعه‌یافته از خودرمزنگارهای سنتی هستند که یک چهارچوب مولد برای یادگیری نمایش فضای پنهان داده‌ها معرفی می‌کنند. بر خلاف خودرمزنگارهای استاندارد که نگاشت‌های قطعی را یاد می‌گیرند، خودرمزنگارهای واریانسی فرض می‌کنند که فضای نهان داده‌ها از یک توزیع احتمالی پیروی می‌کنند که به آنها امکان می‌دهد ساختارهای پیچیده داده‌ها را به طور موثرتری مدل کنند. خودرمزنگارهای واریانسی که توسط کینگما و ولینگ (Kingma and Welling, 2013) توسعه‌یافتند، به طور گسترده در تشخیص آنومالی‌های تولید تصویر و یادگیری ویژگی های بدون نظارت مورد استفاده قرار گرفته‌اند. در اکتشافات زمین‌شیمیایی، آنها روشی قوی برای شناسایی آنومالی‌های زمین‌شیمیایی با مدل سازی الگوهای فضایی و ترکیبی در مناطق معدنی ارائه می دهند. یک خودرمزنگار واریانسی از دو جزء اصلی تشکیل‌شده است: یک رمزگذار (شبکه استنتاج) و یک رمزگشا (شبکه مولد). تفاوت اساسی با رمزگذارهای متعارف این است که به جای رمزگذاری ورودی x به یک نقطه به نام z در فضای محدود فرض می‌کنند که z از یک توزیع احتمالی پیروی می‌کند. شبکه خودرمزنگار واریانسی نیز مانند شبکه خودرمزنگار از دو شبکه رمزگذار و رمزگشا تشکیل‌شده است و گفته‌شد که شبکه رمزگذار به جای اینکه داده ورودی x را به یک نقطه در فضای z تصویر کند، آن را به صورت یک توزیع احتمالاتی در نظر می‌گیرد. بنابراین در رابطه ۸ داریم (Rezende et al., 2014):

رابطه ۸: $z\sim N(\mu(x),\sigma^2(x))$

که $\mu(x)$ میانگین توزیع یادگرفته شده، $\sigma^2(x)$ واریانس آن و N نشان‌دهنده توزیع نرمال است و یا می‌توان با تعریف متغیر تصادفی ϵ عبارت شماره ۸ را به صورت رابطه ۹ نوشت که این متغیر تصادفی از توزیع نرمال نمونه‌برداری شده است (Rezende et al., 2014):

رابطه ۹: $z=\mu(x)+\sigma(x)\cdot\epsilon, \quad \epsilon\sim N(0,1)$

DOI: 10.22067/econg.2025.1157

قسمت رمزگشا نیز داده ورودی x را از فضای محدود به صورت رابطه ۱۰ بازسازی می‌کند که در آن f شبکه رمزگشا است (Rezende et al., 2014):

رابطه ۱۰: $\hat{x}=f(z)$

برای آموزش خودرمزنگار واریانسی و کم‌کردن مقدار تابع هزینه، در این پژوهش از یک تابع هزینه دو قسمتی استفاده شد. در قسمت اول که مسئولیت دارد، مطمئن شود تا داده‌های بازسازی‌شده $\hat{x}$ بیشترین شباهت را به داده‌های ورودی x دارند، از تابع میانگین مربعات خطا استفاده شد (Rezende et al., 2014) (رابطه ۱۱):

رابطه ۱۱: $L_{recon}=\|x-\hat{x}\|$

و قسمت دوم تابع وظیفه این را دارد که مطمئن شود توزیع یادگرفته شده نزدیک به توزیع نرمال است و به عنوان تابع زیر (رابطه ۱۲) شناخته می‌شود:

رابطه ۱۲: $L_{KL}=D_{KL}(N(\mu,\sigma^2)\|N(0,1))$

و در نهایت، تابع هزینه کلی به صورت رابطه ۱۳ است که β در آن یک نرخ یا ثابت برای برابرکردن خطای بازسازی با فضای محدود است:

رابطه ۱۳: $L=L_{recon}+\beta L_{KL}$

اکتشاف زمین‌شیمیایی شامل تجزیه و تحلیل غلظت‌های عنصری برای تشخیص الگوهای غیرعادی نشان‌دهنده کانی‌سازی است. خودرمزنگارهای واریانسی با یادگیری توزیع زمین‌شیمی زمینه و مشخص‌کردن نمونه‌هایی که به طور قابل توجهی انحراف دارند، به شناسایی آنومالی‌های زمین‌شیمیایی کمک می‌کنند. در این پژوهش همانند خودرمزنگار معمولی با فرض اینکه بیشتر نمونه‌ها متعلق به دسته زمینه هستند، آموزش خودرمزنگار واریانسی بر روی تمام داده‌ها انجام‌شد. VAE نسخه احتمالاتی AE است و با افزودن ترم KL، ساختار نهفته را منظم و برآوردی هموار از امتیاز ناهنجاری ارائه می‌کند. این هموارسازی در داده‌های پرنویز مفید است؛ اما ممکن است لبه‌های آنومالی را کمی نرم‌تر نشان دهد. همچنین تعداد ابرپارامترها بیشتر است و به تنظیم دقیق‌تری نیاز دارد و زمانی مناسب است که پایداری و یکنواختی امتیاز در اولویت باشد و امکان تنظیم دقیق فراهم باشد.

انتخاب بهترین مقدار مؤلفه‌ها

مدل‌های یادگیری ماشین اغلب دارای ابرپارامترهایی (مانند نرخ یادگیری، عمق درخت، تعداد لایه‌ها) هستند که به طور قابل توجهی بر عملکرد مدل تأثیر می‌گذارند. بر خلاف مؤلفه‌های مدل که در طول آموزش یادگرفته می‌شوند، ابرپارامترها باید از قبل با توجه به داده‌ها تعریف‌شده و برای هر مسئله تنظیم‌شوند. تنظیم و انتخاب درست و مناسب می‌تواند به طور چشمگیری دقت یا تعمیم‌پذیری را بهبود بخشد؛ به همین دلیل بهینه‌سازی ابرپارامترها به مرحله‌ای حیاتی در استفاده از روش‌های یادگیری ماشین تبدیل شده است (Snoek et al., 2012). روش‌هایی همچون جست‌وجوی شبکه‌ای ، جست‌وجوی تصادفی و غیره برای پیداکردن مقادیر ابرپارمترها وجود دارد؛ ولی هرکدام با نقص‌هایی روبرو هستند. به طور مثال، جست‌وجوی شبکه‌ای تمام ترکیب‌ها ممکن ابرپارامترها را مورد بررسی قرار داده؛ لذا حجم پردازشی زیادی نیاز دارند و یا در جست‌وجوی تصادفی که ترکیب‌هایی به طور تصادفی از مقادیر پیشنهادی برای ابرپارامترها انتخاب می‌کند، تمام ترکیب‌هایی که بعد از چند مرحله پیشنهاد می‌شوند جهت‌دار نبوده و فقط به صورت تصادفی انتخاب می‌شوند (Li et al., 2016). این روش‌ها نیازمند اجرای آموزش‌های متعدد مدل هستند که وقتی هر آموزش مدل گران باشد (مثلاً آموزش یک شبکه عمیق برای دوره‌های زیاد)، هزینه به طور سرسام‌آوری زیاد می‌شود. همچنین انتخاب مقادیر هایبرپارامترها برای آموزش بعدی بر اساس نتایج گذشته تنظیم نمی‌شوند. ما به راهبردهایی نیاز داریم که مجموعه ابرپارامتر بعدی را هوشمندانه‌تر از جست‌وجوی کور انتخاب کنند. اینجاست که بهینه‌سازی بیزی وارد می‌شود و رویکردی اصولی و کارآمد برای تنظیم ابرپارامترها ارائه می‌دهد. بهینه‌سازی بیزی یک راهبرد مبتنی بر مدل‌سازی ترتیبی و تکراری برای بهینه‌سازی توابعی است که ارزیابی آنها پرهزینه است. ایده اصلی در این استراتژی این است که تابع هدف ناشناخته را به عنوان

DOI: 10.22067/econg.2025.1157

تابعی تصادفی در نظرگرفته و یک مدل احتمالی بر روی آن قرار دهیم که اغلب مدل گاوسی است (Sarikhani et al., 2022). فرایند گاوسی یک توزیع روی تابع هدف را تعریف می‌کند. با توجه به میانگین و کوواریانس پیشین، هر نقطه x در فضای جست‌وجو دارای یک توزیع پیش‌بینی مرتبط است. این مدل‌سازی احتمالاتی به این معناست که BO با جمع‌آوری داده‌های بیشتر (مشاهدات از هدف)، یک باور یا نظریه از اینکه کدام نواحی فضای ابرپارامتر امیدوارکننده هستند را حفظ می‌کند (Chen and Okle, 2024). در مقایسه با روش‌های ساده، بهینه‌سازی بیزی با استفاده از مشاهدات گذشته برای هدایت انتخاب‌های آینده، به طور کارآمد ابرپارامترها را انتخاب می‌کند. در این بررسی، از بهینه‌سازی بیزی از طریق یک کتابخانه برنامه‌نویسی زبان پایتون به نام آپچونا برای تنظیم ابرپارامترها برای الگوریتم‌های یادگیری بدون نظارت جنگل ایزوله، دی‌بی اسکن، خودرمزنگار و خودرمزنگار واریانسی استفاده کردیم. این کتابخانه یک کتابخانه متن‌باز پایتون برای تنظیم خودکار هایپرپارامترهاست که به صورت داخلی از نوعی بهینه‌سازی بیزی استفاده می‌کند و بدان معناست که همان‌طور که تنظیم پیشرفت می‌کند، کتابخانه مدل‌های احتمالی تابع هدف را ایجاد می‌کند و از آنها برای انتخاب مجموعه بعدی هایپرپارامترها استفاده می‌کند. ما این کتابخانه را به دلیل انعطاف‌پذیری، سهولت استفاده و ویژگی‌های پیشرفته آن نسبت به سایر کتابخانه‌های بهینه‌سازی بیزی انتخاب کردیم. استفاده از بهینه‌سازی بیزی در تنظیم ابرپارامترهای ما، بهبودهای قابل توجهی در عملکرد نسبت به تنظیمات پیش‌فرض و تلاش‌های تنظیم قبلی به همراه داشت. همه الگوریتم‌های مورد استفاده از تنظیم ابرپارامترهای حیاتی برای تطبیق بهتر با مجموعه داده ما بهره‌مند شدند.چندین رویکرد دیگر برای تنظیم ابرپارامترها وجود دارد که هر کدام مزایا و معایب خود را دارند. به طور مثال، روش شبکه‌ای که هر ترکیبی از بازه تعریف شده را برای هر هایپرپارامتر اجرا می‌کند، به صرف هزینه و منابع پردازشی بسیاری منجر می‌شود (Ghawi and Pfeffer, 2019) و یا روش جست‌وجوی تصادفی که ترکیب تصادفی از بازه انتخابی برای هایپرپارامترها را انجام می‌دهد و هیچ‌گونه سیاستی برای انتخاب ترکیب بعدی در نظر نمی‌گیرد (Rijn and Hutter, 2018). روش‌های دیگر الگوریتم‌های تکاملی مانند الگوریتم ژنتیک هستند که قدرت الگوریتم‌های تکاملی در توانایی آنها برای کاوش فضاهای جست‌وجوی پیچیده، با ابعاد بالا یا گسسته مؤثر است (Thavasimani and Srinath, 2022). در این پژوهش، تنظیم ابرپارامترها با بهینه‌سازی بیزی انجام‌شد تا تعداد آزمون‌های کم‌فایده کم شود و جست‌وجو روی گزینه‌های امیدبخش متمرکز شود. تابع هدف به صورت ترکیبی تعریف شد: الف- امتیاز اعتبارسنجی مدل بیشینه شود؛ ب- اگر مساحت دسته آنومالی بالا بیش از حد بزرگ شود، امتیاز جریمه بر روی آن مدل اعمال شود. این کار باعث‌شد نقشه‌ها هم از نظر کشف رخدادها قوی باشند و هم از نظر منطقه تحت پوشش توسط دسته آنومالی کوچک بمانند. بازه جست‌وجو برای هر روش معقول تعیین‌شد و ترکیب‌های ناسازگار حذف شد.

در این بررسی، بهینه‌سازی بیزی از طریق کتابخانه آپچونا برای تنظیم ابرپارامترهای چندین مدل تشخیص ناهنجاری به صورت زیر پیاده‌سازی شد:

جنگل ایزوله

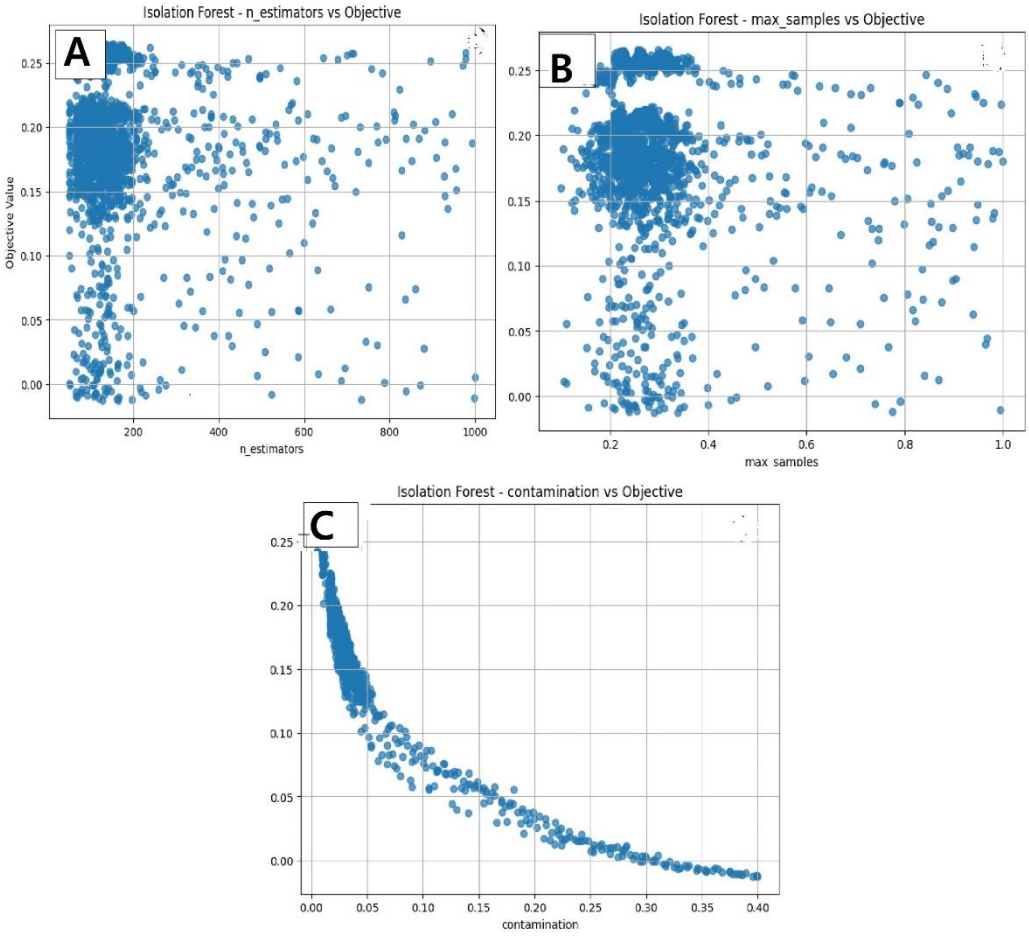
ما ابرپارامترها (تعداد تخمین گرها ، درصد نمونه‌های آنومال، تعداد بیشینه نمونه‌ها) را با بیشینه‌سازی نمره تابع تصمیم گیری مدل بهینه‌کردیم. این نمره میزان آنومال یا عادی‌بودن هر نمونه را اندازه‌گیری می‌کند که مقادیر بالاتر نشان‌دهنده عادی‌بودن یا نرمال‌بودن بیشتر است. بهینه‌سازی این معیار به طور مؤثر توانایی الگوریتم را در تشخیص نمونه‌های عادی از آنومالی‌ها بهبود می‌بخشد. با توجه به نسبت کم آنومالی‌های مورد انتظار در مجموعه داده ما، بیشینه‌سازی میانگین نمره تصمیم‌گیری، مرز مشخص‌تری بین آنومالی‌ها و داده‌های عادی ایجاد می‌کند که در نهایت عملکرد تشخیص ناهنجاری را بهبود می‌بخشد. در شکل ۵-

DOI: 10.22067/econg.2025.1157

A، B و C، نمودارهای مربوط به بهینه‌سازی‌های ابرپارامترهای یادشده نمایش‌داده شده است و همچنین مقدار بهینه انتخاب‌شده برای هایپرپارامترهای مدل جنگل ایزوله نیز در جدول ۲ نشان داده شده است.

دی‌بی اسکن

در این بررسی، به طور نظام‌مند ابرپارامترهای الگوریتم خوشه‌بندی دی‌بی اسکن را برای شناسایی آنومالی‌ها در داده‌های زمین‌شیمیایی با استفاده از بهینه‌سازی بیزی از طریق کتابخانه آپچونا بهینه‌سازی کردیم.



شکل ۵. نتیجه مدل‌سازی‌های مدل جنگل ایزوله انجام‌شده بر روی نمونه‌های منطقه پاریز و چهارگنبد با مقادیر مختلف ابرپارامترها نسبت به تابع هدف و یا میانگین تابع تصمیم‌گیری که هرچه مقدار میانگین این تابع برای یک مدل بیشتر باشد با توجه به فرض ما که اغلب نمونه‌ها نرمال هستند، مدل عملکرد بهتری دارد. A: نمودار مربوط به تأثیر تعداد درخت‌های تصمیم بر روی مقدار تابع هدف، B: نمودار مربوط به تأثیر بیشینه تعداد نمونه‌ها روی مقدار تابع هدف و C: نمودار مربوط به تأثیر مقدار تعداد نمونه‌های آنومال روی مقدار تابع هدف

Fig. 5. Results of isolated forest modeling for Pariz and Chargonbad area performed with different hyperparameter values relative to the objective function or mean decision function, where the higher the mean value of this function for a model, considering our assumption that most samples are normal, the better the model performs. A: Graph showing the effect of number of decision trees on objective function value, B: Graph showing the effect of max-samples on objective function value, and C: Graph showing the effect of Contamination value on objective function value

DOI: 10.22067/econg.2025.1157

جدول ۲. مقادیر انتخاب‌شده برای مدل جنگل ایزوله بر اساس جست‌وجوی بایزین در منطقه پاریز و چهارگنبد

Table 2. The selected values for the Isolation Forest model parameters are based on Bayesian search in Pariz and Chargonbad area

Parameter	Value
n_estimator	125
Max_samples	0.29
Contamination	0.01

دی‌بی اسکن به دلیل توانایی خود در مدیریت خوشه‌های با تراکم‌های مختلف و شناسایی مؤثر نقاط پرت فضایی، به طور گسترده در اکتشافات زمین‌شیمیایی مورد استفاده قرار می‌گیرد. با این حال، عملکرد آن به طور مؤثری به ابرپارامترهایی مانند شعاع همسایگی (eps) و کمینه نقاط مورد نیاز برای تشکیل خوشه بستگی دارد.

برای اطمینان از عملکرد بهینه در یک زمینه بدون نظارت (بدون داده‌های آنومالی برچسب‌گذاری شده)، یک معیار ارزیابی ترکیبی جدید تعریف کردیم که دو معیار مکمل را ادغام می‌کند:

۱. نسبت نقاط پرت: نسبت نقاطی که توسط دی‌بی اسکن به عنوان نویز (آنومالی) شناسایی شده‌اند، یک کسر مناسب منعکس‌کننده انتظارات واقع‌بینانه زمین‌شناسی از داده‌های آنومال است.

۲. امتیاز تنظیم‌شده بر اساس فاصله: این معیار فاصله اقلیدسی هر نقطه داده را نسبت به نزدیک‌ترین همسایه k-ام در نظر می‌گیرد که برای نقاط غیر هسته‌ای با افزودن فاصله به نزدیک‌ترین نقطه هسته بیشتر تنظیم می‌شود. کمینه‌سازی این معیار به تشکیل خوشه‌های فشرده و معنادار زمین‌شناسی کمک می‌کند و در عین حال ناهنجاری‌های واقعی را به وضوح جدا می‌کند.

این دو معیار به یک تابع هدف واحد ترکیب شدند تا تعادل بین فشردگی خوشه‌ای و فراوانی ناهنجاری را برقرار کنند (رابطه ۱۴):

رابطه ۱۴:

$$\text{Combined Score}_{\text{DBSCAN}}(\text{for each model}) = \alpha \times (\text{Fraction of Outliers}) + (1-\alpha) \times (\text{Average Adjusted Distance Score})$$

در طول بهینه‌سازی، به آپچونا اجازه دادیم تا به طور خودکار فضای ابرپارامترها را برای مقادیر بهینه موارد زیر جست‌وجو کند:

eps (شعاع): به صورت لگاریتمی در محدوده ۰/۱ تا ۱۰۰۰ به صورت لگاریتمی جست‌وجو شد.

min_samples: محدوده ۱ تا ۲۰۰.

k: مؤلفه اندازه همسایگی که در محاسبه فاصله استفاده می‌شود، بین ۳ تا ۲۰ بهینه‌سازی شد.

α: ضریب تعادل بین کسر نقاط پرت و امتیاز مبتنی بر فاصله، در محدوده بین ۰ (کاملاً مبتنی بر فاصله) و ۱ (کاملاً مبتنی بر نقاط پرت) قرار می‌گیرد و مقدار آن برابر ۰/۵ انتخاب شد که انتخاب این مقدار به این معنی است که به تعداد نقاط آنومال تشخیص‌داده شده و فاصله نقطه آنومال از مرکز دسته‌ها به یک اندازه اهمیت می‌دهیم.

بهینه‌سازی بیزی به طور کارآمد این فضای پیچیده ابرپارامتر را با ارزیابی تکراری ترکیب‌های مختلف مؤلفه‌ها در حدود ۱۵۰۰۰ آزمایش، کاوش کرد که مقادیر بهینه‌شده در جدول ۳ و نمودارهای مرتبط با این بهینه‌سازی در شکل ۶-A، B و C نشان‌داده شده است که در این نمودارها مقادیری که به ازای آنها کمترین مقدار تابع هدف به دست آید، به عنوان مقادیر بهینه انتخاب می‌شود. در شکل ۶، مقادیر تابع هدف (رابطه ۱۴) یادشده به ازای مقادیر مختلف هر هایپرپارامتر نشان‌داده شده است که با توجه به توضیح داده شده، مدلی به عنوان مدل بهینه انتخاب می‌شود که مقدار تابع هدف به ازای مقادیر هایپرپارامترهای آن کمینه باشد. در شکل ۶-B مقادیر این تابع به ازای مقادیر مختلف هایپرپارمتر min-sample نمایش‌داده شده است که از طریق این نمودار می‌توان

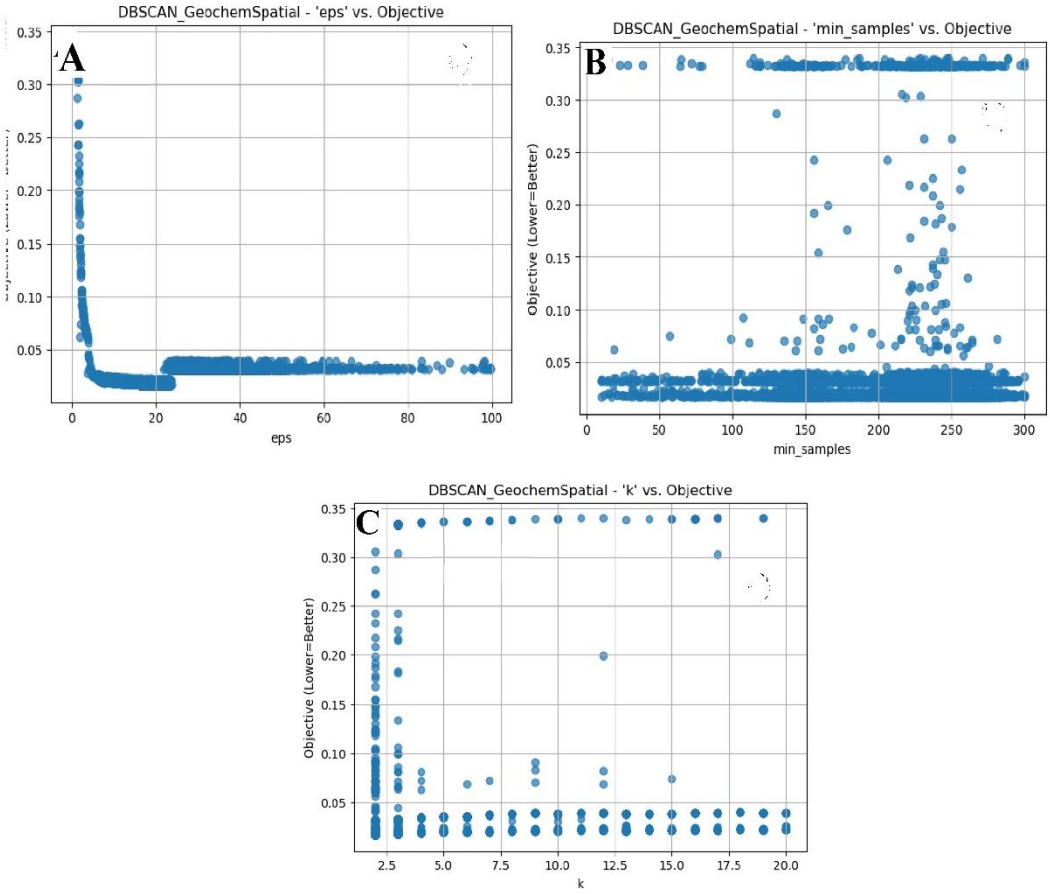
DOI: 10.22067/econg.2025.1157

متوجه شـــد که به ازای مقادیر مختلف این هایپرپارامتر، تابع هدف می‌تواند کمینه باشـد؛ ولی این موضـوع برای دو هایپرپارامتر دیگر به این شـــکل نیســـت و روند تأثیر انتخاب این هایپرپارامترها در نمودارها مشخص است.

جدول ۳. مقادیر انتخاب‌شده برای دی‌بی اسکن بر اساس جست‌وجوی بایزین در منطقه پاریز و چهارگنبد

Table 3. The selected values for the DBscan model parameters are based on Bayesian search Pariz and Chargonbad area

Parameter	Value
eps	21.25
Min_samples	195
k	2



شـــکل ٦. نتیجه مدل‌ســازی‌های دی‌بی اســکن انجام‌شــده در منطقه پاریز و چهارگنبد با مقادیر مختلف ابرپارامترها نســبت به تابع هدف (رابطه ۱۴). A: نمودار مربوط به تأثیر مقدار فاصـــله بر روی مقدار تابع هدف، B: نمودار مربوط به تأثیر حداقل نقاط مورد نیاز برای تشـــکیل خوشـــه بر روی مقدار تابع هدف و C: نمودار مربوط به تأثیر مقدار k بر روی مقدار تابع هدف

Fig. 6. Results of DBSCAN modeling in Pariz and Chargonbad performed with different hyperparameter values in relation to the objective function (Equation 17). A: Graph showing the effect of distance value on the objective function, B: Graph showing the effect of min-samples on the objective function, and C: Graph showing the effect of k value on the objective function

DOI: 10.22067/econg.2025.1157

خودرمزنگار

در این پژوهش از یک خودرمزگذار عمیق برای تشخیص آنومالی چند دسته‌ای در داده‌های زمین‌شیمیایی استفاده شده است. برای دستیابی به بهترین عملکرد مدل، فرایند بهینه‌سازی هایپرپارامترها با استفاده از الگوریتم بیزی مبتنی بر کتابخانه آپچونا پیاده‌سازی شد.داده‌ها به دو مجموعه آموزشی و اعتبارسنجی تقسیم شدند. خودرمزگذار طراحی‌شده به گونه‌ای پیاده‌سازی شد که تعداد نورون‌ها در لایه‌های پنهان به صورت نزولی کاهش‌یابد. به این صورت که تعداد نورون‌ها در نخستین لایه پنهان برابر یک مقدار ثابت بوده و در هر لایه بعدی، نصف تعداد نورون‌های لایه قبلی می‌شود تا در نهایت به تعداد نورون‌های لایه پنهان مرکزی برسد. این روند کاهشی تضمین می‌کند که ساختار مدل، بهینه و به صورت هرمی طراحی‌شده است. برای بهینه‌سازی هایپرپارامترها، تابع هدف به گونه‌ای طراحی‌شد که ترکیبی از خطای بازسازی میانگین و انحراف استاندارد خطاها باشد که این مقادیر قبل از ترکیب برای یکی‌شدن اثر آنها در ترکیب استانداردسازی شده‌اند. به طور مشخص تابع هدف به صورت رابطه ۱۵ تعریف‌شد:

رابطه ۱۵:

$$\text{Combined Score (for each model)} = \text{Mean reconstruction error} + 0.5 * \text{STD(Eror)}$$

که در این رابطه Mean Error (خطای بازسازی میانگین) از محاسبه میانگین مربعات تفاوت بین داده‌های اعتبارسنجی اصلی و مقادیر بازسازی‌شده توسط خودرمزگذار به دست می‌آید. هرچه این مقدار کمتر باشد، نشان‌دهنده دقت بالاتر مدل در بازسازی داده‌هاست.

STD(Error) (انحراف استاندارد خطاها) بیانگر پراکندگی خطاهای بازسازی در مجموعه داده اعتبارسنجی است. هرچه این مقدار کمتر باشد، عملکرد مدل باثبات‌تر است. هدف از این ترکیب، ایجاد تعادل میان دقت بازسازی میانگین و ثبات عملکرد مدل (کاهش پراکندگی خطاها) است. در رابطه ۱۵، هرچه مقدار امتیاز ترکیبی کمتر باشد، عملکرد مدل مطلوب‌تر است و هدف از بهینه‌سازی رسیدن به کمترین مقدار آن است. هایپرپارامترهای بهینه‌شده در این پژوهش شامل تعداد لایه‌های پنهان، ابعاد لایه‌های پنهان، ابعاد لایه پنهان مرکزی یا توابع فعال‌سازی و اندازه دسته است. برای آموزش خودرمزگذار از الگوریتم بهینه‌ساز آدام با کاهش نرخ یادگیری به همراه توقف زودهنگام استفاده شد.

در نهایت، الگوریتم به صورت هوشمند فضای هایپرپارامترها را جست‌وجو کرده و ترکیبی از هایپرپارامترها را یافت که کمترین مقدار رابطه ۱۵ را فراهم‌کرد که این نتایج در جدول ۴ آمده است. این رویکرد تضمین‌کرد که مدل بهینه دارای بهترین دقت بازسازی همراه با کمترین نوسان و پراکندگی خطا در مجموعه داده‌های زمین‌شیمیایی باشد.

جدول ٤. مقادیر انتخاب‌شده برای خودرمزنگار بر اساس جست‌وجوی بایزین بر روی داده‌های منطقه پاریز و چهارگنبد

Table 4. The selected values for the Autoencoder model parameters are based on Bayesian search on Pariz and Chargonbad data

Parameter	Value
n_layers	2
hidden_dim	64
Latent (latent_dim)	16
hidden_activation	elu
Output activation	Tanh

DOI: 10.22067/econg.2025.1157

خودرمزنگار واریانسی

در این پژوهش، معماری یک خودرمزنگار واریانسی با رویکرد یادگیری عمیق به منظور شناسایی آنومالی‌های زمین‌شیمیایی در داده‌های زمین‌شیمیایی اکتشافی، بهینه‌سازی شد. هدف اصلی، تشخیص بهینه و خودکار زون‌های آنومال زمین‌شیمیایی بود که می‌توانند به عنوان شاخص‌هایی برای حضور احتمالی مناطق کانی‌سازی‌شده تلقی شوند. استفاده از مدل خودرمزنگار واریانسی به دلیل قابلیت آن در استخراج الگوهای پنهان و مدل‌سازی توزیع داده‌های پیچیده، به ویژه در شرایطی با عدم قطعیت بالا و داده‌های پراکنده زمین‌شیمیایی از اهمیت ویژه‌ای برخوردار است. چنان‌که در قبل نیز بیان‌شد از بهینه‌سازی ابرپارامترها به روش بیزی با استفاده از کتابخانه آپچونا بهره گرفتیم و به صورت نظام‌مند، مؤلفه‌های کلیدی معماری و آموزش مدل خودرمزنگار واریانسی را تنظیم کردیم. هدف از این فرایند، کاهش خطای بازسازی مدل بر روی داده‌های اعتبارسنجی بود تا دقت تشخیص ناهنجاری‌ها افزایش یابد. ابرپارامترهایی که بهینه‌سازی شدند، شامل موارد زیر بودند:

تعداد لایه‌های مدل: تعداد لایه‌ها بین ۱ تا ۸ تغییرداده شد تا عمق بهینه‌ای به دست آید که بین پیچیدگی مدل و کارایی محاسباتی تعادل برقرار کند.

ابعاد لایه‌های پنهان: تعداد نورون‌های مختلف شامل ۱۶، ۳۲، ۶۴، ۱۲۸، ۲۵۶، ۵۱۲ و ۱۰۲۴ بررسی شد تا ظرفیت مناسبی برای مدل‌سازی روابط پیچیده زمین‌شیمیایی شناسایی‌شود.

ابعاد فضای نهفته: اندازه‌های مختلفی از نورون‌ها در لایه تنگنا شامل ۸، ۱۶، ۳۲ و ۶۴ مورد ارزیابی قرارگرفت تا ظرفیت بهینه برای نمایش ویژگی‌های اساسی داده‌ها تعیین‌شود.

توابع فعال‌ساز: عملکرد توابع فعال‌ساز مختلف بررسی‌شد. برای لایه خروجی، بین توابع خطی، تانژانت هایپربولیک و سیگموید انتخاب صورت‌گرفت؛ در حالی‌که برای لایه‌های پنهان، تابع tanh به صورت ثابت استفاده شد؛ زیرا که در مدل‌سازی روابط غیرخطی عملکرد مناسبی دارد.

اندازه بچ: مقادیر مختلفی شامل ۱۶، ۳۲ و ۶۴ آزمایش شدند تا اندازه‌ای بهینه برای آموزش مؤثر و تعمیم‌پذیری مناسب مدل تعیین‌شود.

فرایند بهینه‌سازی با استفاده از بهینه‌ساز آدام و نرخ یادگیری اولیه ۰/۰۱ انجام‌شد. برای پایدارسازی فرایند آموزش، از برش گرادیان با مقدار clipnorm برابر ۱/۰ بهره‌گرفته شد. همچنین، برای افزایش کارایی آموزش، از تنظیم پویای نرخ یادگیری استفاده شد. برای جلوگیری از بیش‌برازش، از روش توقف زودهنگام استفاده شد که در آن، مقدار خطای آموزش پایش می‌شد و در صورت عدم بهبود طی ۵ دوره متوالی، فرایند آموزش متوقف می‌شد. تابع هدف در فرایند بهینه‌سازی، مجموع کل خطای مدل خودرمزنگار واریانسی بود که شامل دو مؤلفه اصلی است:

خطای بازسازی: به صورت میانگین مربع خطا بین داده‌های اولیه و داده‌های بازسازی‌شده محاسبه می‌شود.

مقدار واگرایی کولبک-لیبلر: که نشان‌دهنده فاصله بین توزیع نهفته مدل و توزیع نرمال استاندارد است. این تابع هدف روی یک مجموعه داده اعتبارسنجی مستقل ارزیابی‌شد. در نهایت، ترکیب بهینه ابرپارامترها بر اساس کمینه‌شدن مقدار کل خطای اعتبارسنجی یا تابع هدف انتخاب‌شد. نتایج مربوط به انتخاب بهترین هایپرپارامترها در جدول ۵ آمده است. این نتیجه حاصل ۹۵۰ بار اجرای فرایند بهینه‌سازی و آزمون ترکیب‌های مختلف مؤلفه‌ها بود که نشان‌دهنده دقت و جامعیت رویکرد آزمایشی اتخاذشده در این پژوهش است. بهینه‌سازی بیزی با کاوش هدفمند فضای مؤلفه‌ها به افزایش دقت جداسازی در هر ۴ مدل منجرشد. این روش در مقایسه با تنظیم دستی یا جست‌وجو تصادفی، روش کارآمدتری بوده است؛ اگرچه زمان محاسباتی بیشتری صرف‌شد.

نتایج و بحث

پس از انتخاب بهینه هایپرپارامترهای مرتبط به هر مدل، با توجه به راهبرد بیان‌شده، نوبت به اجرا، بررسی و مقایسه عملکرد هر مدل رسید. برای بررسی و مقایسه عملکرد مدل‌های به کاربرده شده در

DOI: 10.22067/econg.2025.1157

این پژوهش از موقعیت نقاط کانی‌زایی شناخته‌شده در منطقه استفاده شد که تعداد ۳۹ کانی‌زایی شناخته‌شده در منطقه حضور داشت و موقعیت آنها در شکل ۱-B نشان‌داده شده است. با استفاده از این نقاط عملکرد مدل‌های مورد استفاده مورد مقایسه قرارگرفت که در ادامه به آن می‌پردازیم.

جدول ۵. مقادیر انتخاب‌شده برای خودرمزنگار واریانسی بر اساس جست‌وجوی بایزین بر روی داده‌های پاریز و چهارگنبد

Table 5. The selected values for the VAE model parameters are based on Bayesian search on Pariz and Chargonbad

Parameter	Value
n_layers	1
hidden_dim	1024
Latent or dimension of the bottleneck	128
hidden_activation	Tanh
Output activation	Linear

جنگل ایزوله

پس از به دست آوردن هایپرپارامترهای بهینه برای مدل جنگل ایزوله، این مدل با استفاده از کتابخانه سایکیت‌لرن اجرا شد و سپس از مقدار امتیاز به دست آمده برای هر نمونه، برای انجام جداسازی آنومالی از زمینه چند دسته‌ای به سه طبقه تقسیم‌شد که این مدل ۲۱ درصد از منطقه مورد بررسی را به عنوان منطقه با امکان کانی‌زایی بالا و یا منطقه آنومال انتخاب‌کرد و تنها ۱۰ نقطه کانی‌زایی شناخته‌شده در این منطقه از ۳۹ نقطه در طبقه‌بندی آنومال و یا طبقه سوم قرارگرفت. چنان که در قسمت‌های قبلی بیان‌شد، هرچه مقدار تابع تصمیم‌گیری برای یک نمونه بالاتر باشد؛ یعنی آن نمونه نرمال‌تر است و هرچه کمتر باشد، به معنی آن است که نمونه آنومال‌تر است. در شکل ۷ نقشه خروجی حاصل از روش جنگل ایزوله به همراه نقاط کانی‌زایی نشان‌داده شده است.

دی‌بی اسکن

پس از به دست آوردن پارامترهای بهینه دی‌بی اسکن (eps، min_samples) از طریق بهینه‌سازی بیزی، به انجام تشخیص آنومالی چند دسته‌ای بر روی داده‌های زمین‌شیمیایی پرداختیم. از آنجایی که دی‌بی اسکن به طور اساسی فقط خروجی باینری (زمینه در مقابل آنومال) ارائه می‌دهد، یک رویکرد نوین امتیازدهی پیوسته بر اساس روابط مکانی معرفی کردیم که امکان طبقه‌بندی دقیق‌تر آنومالی‌ها را فراهم می‌کند. برای هر نمونه داده، یک امتیاز ناهنجاری پیوسته را با ترکیب دو عامل تعریف کردیم:

شاخص آنومال باینری: مقدار ۱۰۰ برای نمونه‌های آنومال تشخیص‌داده شده توسط دی‌بی اسکن (نقاط نویز، برچسب‌خورده به عنوان -۱)، یا در غیر این صورت صفر که می‌توان برای بیشترکردن تأثیر نمونه‌های آنومال مقدار ۱۰۰ را افزایش‌داد.

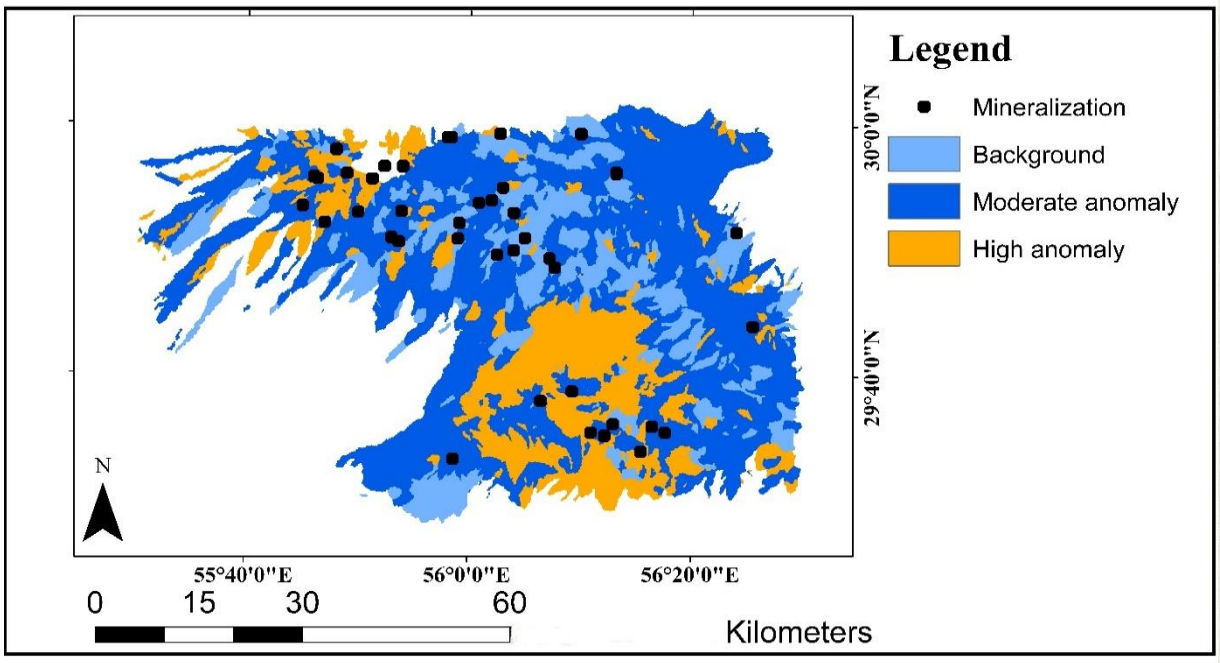
شاخص مبتنی بر فاصله: محاسبه شده به عنوان فاصله اقلیدسی از هر نقطه به k-امین همسایه نزدیک آن. برای نقاط برچسب‌خورده به عنوان غیر-هسته (پرت)، همچنین فاصله اقلیدسی تا نزدیک‌ترین نقطه هسته را اضافه کردیم، بنابراین به مقدار امتیاز نقاط آنومال به شدت اضافه‌شده و آنها را به وضوح از نقاط خوشه‌بندی شده متمایز می‌کنیم. این دو معیار با استفاده از مؤلفه وزن‌دهی بتا β بر اساس رابطه زیر ترکیب‌شدند. رابطه ۱۶ نشان‌دهنده رابطه این امتیاز ترکیبی است. بعد از به دست آوردن مقادیر امتیاز هر نمونه که مقادیر بزرگ‌تر این امتیاز ترکیبی نشان‌دهنده نمونه آنومال است، مقادیر امتیازها برای انجام جداسازی آنومالی از زمینه چند طبقه‌ای به سه دسته تقسیم‌بندی شدند. نتایج مربوطه به این روش در شکل

DOI: 10.22067/econg.2025.1157

۸ نشان‌داده شده است. ۴۷ درصد از منطقه مورد بررسی در این مدل به دسته آنومالی بالا اختصاص‌داده شد و تعداد ۲۴ کانی‌زایی شناخته‌شده از ۳۹ کانی‌زایی در این دسته قرار گرفته‌اند.

رابطه ۱۶:

Combined Score (point i)=β·(DBSCAN Outlier Indicator i)+(1−β)×(Adjusted Distance Score i)



شکل ۷. نقشه خروجی حاصل از مدل جنگل ایزوله در منطقه پاریز و چهارگنبد با هایپرپارامترهای بهینه‌شده به همرا نقاط کانی‌زایی که در این نقشه ۲۱ درصد از منطقه مورد بررسی به طبقه آنومالی بالا تعلق گرفته است و تنها ۱۰ نقطه کانی‌زایی در طبقه آنومالی بالا قرارگرفته و بقیه ۲۹ نقطه کانی‌زایی در دسته‌های دیگر قرار گرفته‌اند.

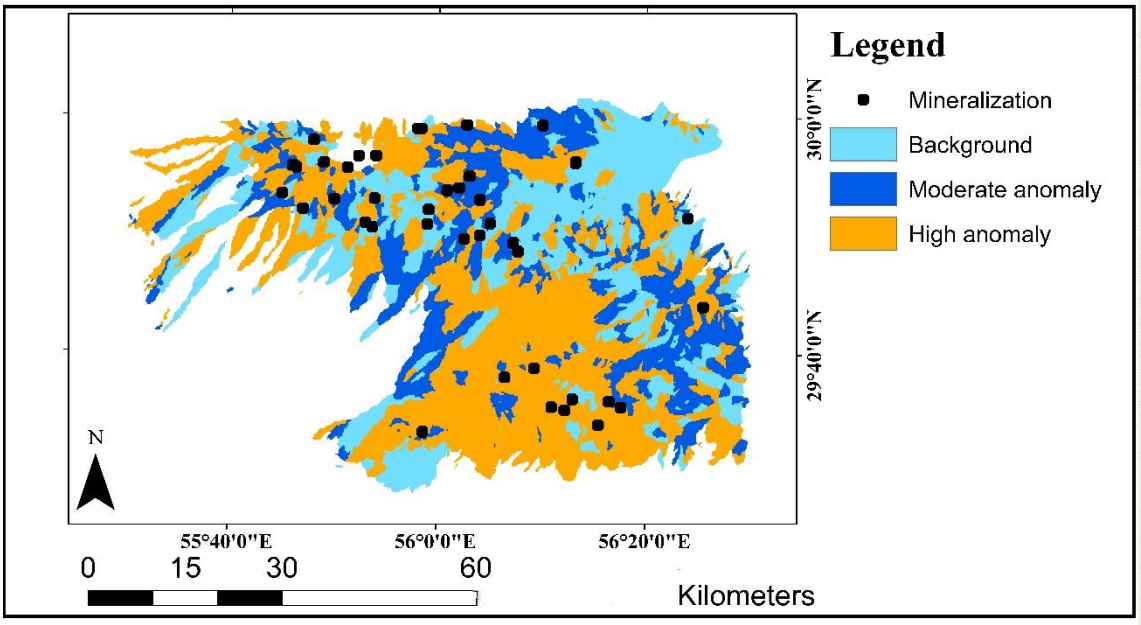
Fig. 7. The output map resulting from the Isolation Forest model in Pariz and Chargonbad area with optimized hyperparameters, along with mineralization points, shows that 21% of the study area falls into the high anomaly class. However, only 10 mineralization points are located in the high anomaly class, while the remaining 29 points fall into other classes.

خودرمزنگار

بعد از به دست آورن ساختار بهینه خودرمزنگار در قسمت قبلی که تعداد لایه‌های آن ۲، تعداد نورون‌ها به ترتیب ۶۴ و ۳۲ و تعداد نورون‌های فضای پنهان ۱۶ انتخاب‌شد و همچنین توابع فعال‌ساز داخلی برابر elu و تابع فعال‌ساز لایه آخر نیز tanh انتخاب‌شد، برای جلوگیری از بیش‌برازش مدل بعد از هر تابع فعال‌ساز داخلی از لایه حذف تصادفی با نرخ ۲۰ درصد استفاده شد. پس از پیاده‌سازی ساختار انتخاب‌شده و تقسیم‌بندی داده‌ها به دو نوع آموزشی و ارزیابی، مدل با تعداد ایپاک مناسب برای جلوگیری از بیش‌برازش آموزش داده شد و خطای ساخت برای هر نمونه به دست آمد. هرچه مقدار این خطا بالاتر باشد، نشان‌دهنده این است که نمونه آنومال‌تر است. سپس خطای به دست آمده برای هر نمونه به سه دسته تقسیم‌شد و نقشه جداسازی آنومالی از زمینه چند دسته‌ای به دست آمد. شکل ساختار خورمزنگار استفاده شده در

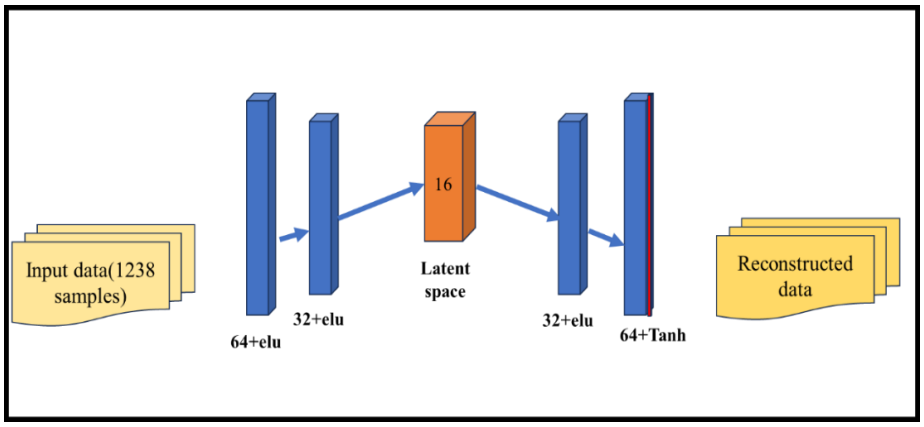
DOI: 10.22067/econg.2025.1157

شکل ۹ نشان‌داده شده و همچنین خروجی این مدل نیز در شکل ۱۰ نشان‌داده شده است. در این مدل ۳۶ درصد از منطقه مورد بررسی به عنوان محدوده آنومالی بالا انتخاب‌شد و تعداد ۲۴ نقطه کانی‌زایی از ۳۹ نقطه کانی‌زایی نیز در این منطقه قرار گرفته‌اند.



شکل ۸. نقشه خروجی حاصل از مدل دی‌بی اسکن در منطقه پاریز و چهارگنبد با هایپرپارامترهای بهینه‌شده به همرا نقاط کانی‌زایی که در این نقشه ۴۷ درصد از منطقه مورد بررسی به دسته آنومالی بالا تعلق‌گرفته و ۲۴ نقطه کانی‌زایی در دسته آنومالی بالا قرارگرفته و بقیه ۱۵ نقطه کانی‌زایی در دسته‌های دیگر قرار گرفته‌اند.

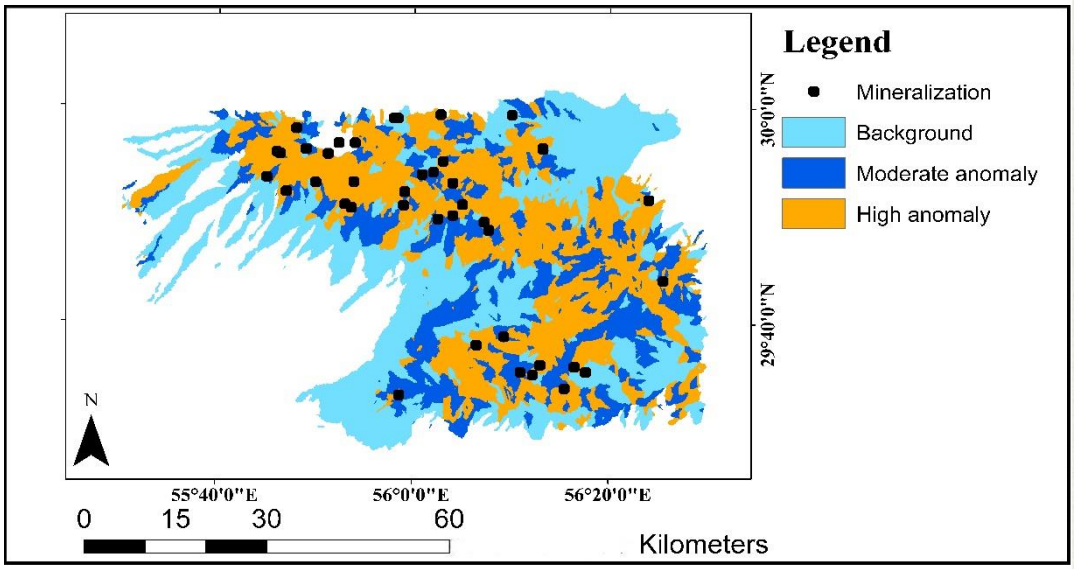
Fig. 8. The output map resulting from the DBSCAN model in Pariz and Chargonbad with optimized hyperparameters, along with mineralization points, shows that 47% of the study area falls into the high anomaly class. In this class, 24 mineralization points are located, while the remaining 15 points fall into other classes.



شکل ۹. تصویر ساختار شبکه خودرمزنگار استفاده شده در جداسازی آنومالی از زمینه چند دسته‌ای در منطقه پاریز و چهارگنبد

Fig. 9. The image of the autoencoder network structure used in this study for anomaly-background separation in a multi-class setting in Pariz and Chargonbad

DOI: 10.22067/econg.2025.1157



شکل ۱۰. نقشه خروجی حاصل از مدل خودرمزنگار در منطقه پاریز و چهارگنبد با هایپرپارامترهای بهینه‌شده به همرا نقاط کانی‌زایی که در این نقشه ۳۶ درصد از منطقه مورد بررسی به دسته آنومالی بالا تعلق گرفته و ۲۴ نقطه کانی‌زایی در دسته آنومالی بالا قرارگرفته و بقیه ۱۵ نقطه کانی‌زایی در دسته‌های دیگر قرار گرفته‌اند.

Fig. 10. The output map resulting from the Autoencoder model in Pariz and Chargonbad area with optimized hyperparameters, along with mineralization points, shows that 36% of the study area falls into the high anomaly class. In this class, 24 mineralization points are located, while the remaining 15 points fall into other classes.

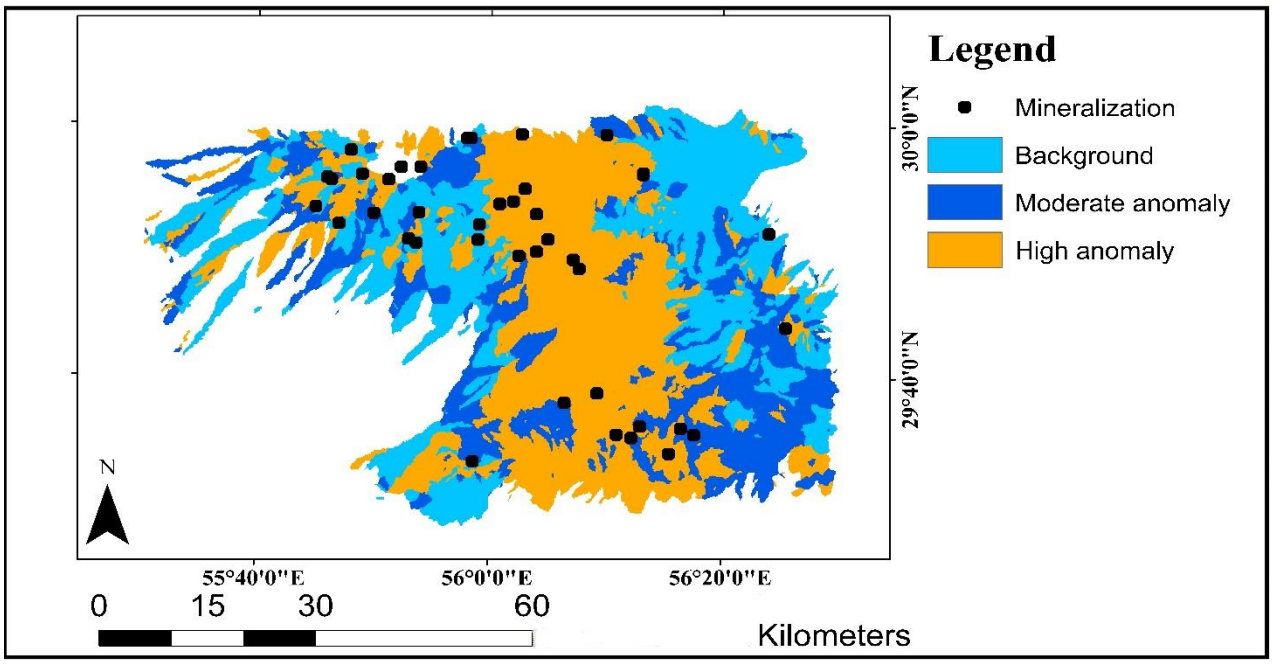
خودرمزنگار واریانسی

خورمزنگار واریانسی با هایپرپارامترهای به دست آمده از نتایج جست‌وجوی بیزی پس از تقسیم داده‌ها به دو دسته آموزشی و ارزیابی، مورد آموزش قرارگرفت و پس از کنترل‌کردن مدل برای بیش‌برازش مدل ساخته‌شده برروی تمامی نمونه‌ها آزمایش‌شد و امتیاز بازسازی یا خطای بازسازی برای تمام نمونه‌ها به دست آمد. همانند خودرمزنگار هرچه خطای بازسازی بیشتر باشد به معنی این است که نمونه آنومال‌تر است؛ پس این خطای به دست آمده برای انجام جداسازی آنومالی از زمینه چند دسته‌ای به سه دسته تقسیم‌بندی شد که هر دسته به ترتیب نمایش‌دهنده زمینه، آنومالی با احتمال متوسط و آنومالی با احتمال بالاست. نقشه حاصل از این دسته‌بندی در شکل ۱۱ نشان‌داده شده است که ۴۲ درصد از منطقه مورد بررسی در این مدل به دسته آنومالی با احتمال بالا اختصاص‌داده شده است و تعداد ۲۴ نقطه از نقاط کانی‌زایی که نقاط آنومال هستند، در این دسته قرار گرفته‌اند. در شکل ۱۲، آرایش شبکه یا مدل خودرمزنگار واریانسی به کاربرد شده در این پژوهش نشان‌داده شده است.

همبستگی نتایج با نقاط کانی‌زایی شناخته‌شده

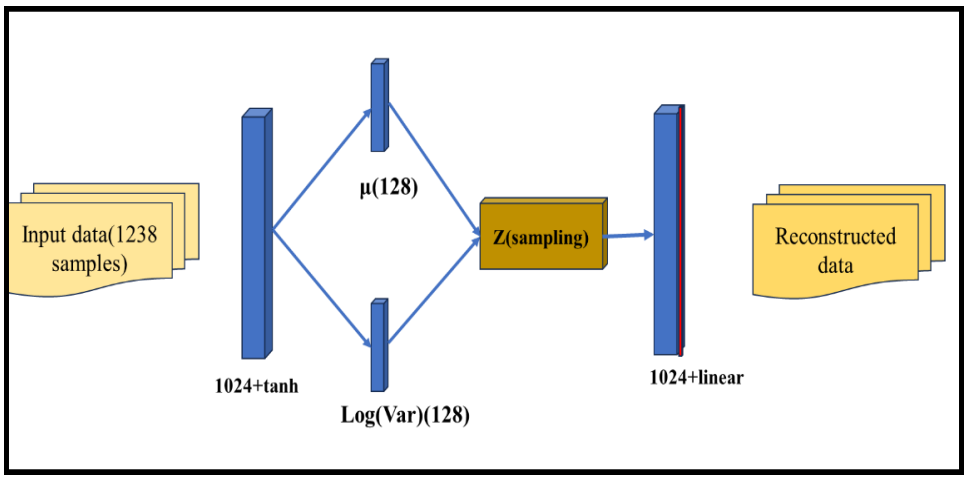
برای سنجش همبستگی فضایی خروجی مدل‌ها با شواهد کانی‌زایی شناخته‌شده، ابتدا امتیاز آنومالی هر نمونه بر اساس حوضه آبریز متصل به نمونه (SCB) منتشر شد و سپس نقشه‌ها به نقشه‌های چند دسته‌ای تبدیل شدند. در ادامه، این نقشه‌های چند دسته‌ای با ۳۹ رخداد کانی‌زایی موجود در محدوده بررسی هم‌پوشانی داده شد. ارزیابی با دو شاخص انجام‌شد: ۱- نرخ اصابت به معنای تعداد رخدادهایی که در طبقه آنومالی بالا قرار می‌گیرند و ۲- سهم سطحی همان طبقه یا دسته‌ای است که نشان‌دهنده گستره فضایی آنومالی و مبنایی برای داوری درباره توازن دقت و پوشش است.

DOI: 10.22067/econg.2025.1157



شکل ۱۱. نقشه خروجی حاصل از مدل خودرمزنگار واریانسی در منطقه پاریز و چهارگنبد با هایپرپارامترهای بهینه‌شده به همرا نقاط کانی‌زایی که در این نقشه ۴۲ درصد از منطقه مورد بررسی به دسته آنومالی بالا تعلق گرفته و ۲۴ نقطه کانی‌زایی در دسته آنومالی بالا قرار گرفته و بقیه ۱۵ نقطه کانی‌زایی در دسته‌های دیگر قرار گرفته‌اند.

Fig. 11. The output map resulting from the VAE model in Pariz and Chargonbad with optimized hyperparameters, along with mineralization points, shows that 42% of the study area falls into the high anomaly class. In this class, 24 mineralization points are located, while the remaining 15 points fall into other classes.



شکل ۱۲. تصویر ساختار شبکه خودرمزنگار واریانسی استفاده شده در منطقه پاریز و چهارگنبد در جداسازی آنومالی از زمینه چند دسته‌ای در این پژوهش

Fig. 12. The image of the VAE network structure used in Pariz and Chargonbad area in this study for anomaly-background separation in a multi-class setting

نتایج جدول ۵ نشـــان می‌دهد همه روش‌ها، هم‌مکانی معناداری با رخدادها دارند؛ اما توازن دقت و پوشش در آنها متفاوت است. در مدل خودرمزنگار (AE) تعداد ۲۴ رخداد از ۳۹ رخداد در طبقه آنومالی بالا قرارگرفته و سهم سطحی این دسته ۳۶ درصد است؛ بنابراین AE متعادل‌ترین تراز میان دقت و پوشش را ارائه می‌کند و پهنه‌های پیوسته‌ای را هم‌مکان با واحدهای آتشفشانی–نفوذی و زون‌های دگرسانی آشکار می‌سازد. روش دی‌بی اسکن نیز ۲۴ رخداد از ۳۹ رخداد را پوشـش می‌دهد؛ اما با سـهم سـطحی ۴۷ درصد، پوشش گسترده‌تر باعث افزایش بازخوانی می‌شود. امکان نمایش گســترده‌تر از حد بهینه در برخی پهنه‌ها وجود دارد؛ پس در دی‌بی اسکن، پوشش وسیع‌تر معمولاً به افزایش نرخ پوشش رخدادها منجر می‌شـود. در این داده، هر دو روش AE و دی‌بی اسکن، ۲۴ رخداد را پوشش می‌دهند؛ اما دی‌بی اسکن این پوشش برابر را با مسـاحت بیشـتری (۴۷ درصـد در برابر ۳۶ درصـد) حاصل کرده است. بنابراین دقت مکانی آن نسبت به AE پایین‌تر اسـت. در VAE نیز ۲۴ رخداد از ۳۹ رخداد با سـهم سـطحی ۴۲ درصـد ثبت‌شـده اسـت که از نظر الگوی فضـایی و توازن دقت–پوشـش بین AE و دی‌بی اسـکن قرار می‌گیرد. در جنگل ایزوله (IF)، ۱۰ رخداد از ۳۹ رخداد با سـهم سـطحی ۲۱ درصـد دیده می‌شـود. این روش نقشـه‌هایی فشـرده‌تر و محافظه‌کارانه‌تر تولید می‌کند و هاله‌های ضعیف‌تر کمتر در نظر گرفته می‌شوند. نقشه‌های طبقه‌بندی شـده آنومالی که در سـطح حوضـه‌های آبریز (SCB) منتشـر شـدند، با رخدادهای پورفیری شـناخته‌شـده در کمربند ارومیه–دختر در منطقه مورد بررسـی هم‌مکان هسـتند. در این مجموعه، خودرمزنگار (AE) کمربندهای پیوسـته‌ای را آشـکار می‌کنـد کـه بـا مراکز نفوذی میوسـن و زون‌هـای دگرسـانی فلدسپات‌کائولینیت–کلریتی همراستاست (شکل‌های ۱۰ و ۱۱). همراسـتایی نقشـه‌های آنومالی حاصـل با مراکز نفوذی میوسـن و زون‌های دگرسـانی پورفیری یک یافته فقط توصـیفی نیسـت؛ بلکه نشان‌دهنده اعتبار زمین‌شناختی نتایج مدل است. این هم‌پوشانی به این معنی اسـت که الگوریتم‌های ما به طور دقیق همان نواحی را برجسته کرده‌اند که زمین‌شناسان نیز به واسطه شواهد زمین‌شیمیایی و کانی‌سـازی انتظار دارند، غنی باشـند. از دید کاربردی، چنین تطابقی بسـیار حائز اهمیت اسـت. هرجا آنومالی‌های مدل با واحدهای زمین‌شـناسـی مسـاعد (مثلاً توده‌های نفوذی یا مناطق دگرسان‌شده) انطباق دارند، احتمالاً یک سامانه کانی‌سازی واقعی را منعکس می‌کننـد. بنـابراین می‌توان نتیجـه‌گرفـت کـه نواحی آنومالی مشابه در امتداد همان کمربند پورفیری – حتی اگر تاکنون ذخیره‌ای در آنها کشـف نشـده باشـد – اهداف مناسـبی برای پی‌جویی‌های اکتشافی آینده خواهند بود. به بیان ساده‌تر، مدل ما نه تنها گذشته را تأیید می‌کند؛ بلکه راهنمایی برای آینده نیز به دست می‌دهد و این افزایش کارایی و دقت در هدف‌گذاری اکتشـافی را ممکن می‌سازد. بر اساس جدول ۵، نقشه AE با بازیابی ۲۴ رخداد از ۳۹ رخداد؛ در حالی‌که ۳۶ درصد از مساحت را به دسته آنومالی بالا تخصیص می‌دهد، توازن دقت–پوشش مطلوب‌تری نسبت به سـایر مدل‌ها نشـان می‌دهد. از این رو، تلفیق این نقشـه با شـواهد زمین‌شـناسـی (نفوذی‌ها، زون‌های دگرسـانی) می‌تواند سـودمندی اکتشـافی را به صـورت مسـتقیم افزایش‌دهد. به طور کلی، نقشـه حاصـل از AE همخوانی بیشـتری با مراکز پورفیری محتمل و زون‌های دگرسـانی پتاسـیک، فیلیک و پروپیلیتیک نشـان می‌دهد. چهار روش به‌کارگرفته شـده در این پژوهش، هر یک بر مبنای اصـول آماری متفاوتی عمل می‌کنند و در نتیجه به جنبه‌های مختلفی از الگوهای زمین‌شـیمیایی حسـاس هسـتند. جنگل ایزوله (IF) با رویکرد درختی خود تمایل دارد فقط قوی‌ترین ناهنجاری‌ها را (که به سـادگی از بقیه نمونه‌ها جدا می‌شـوند) آشـکار کند. بنابراین نقاطی با غلظت بسیار بالای یک یا چند عنصـر (مثل Cu) را به عنوان آنومالی برمی‌گزیند و بخش‌هایی از هاله‌های وسیع‌تر و ضـعیف‌تر را ممکن اسـت نادیده بگیرد. بر عکس، خودرمزنگار عمیق (AE) به دلیل توانایی در یادگیری هم‌زمان روابط غیرخطی چندین عنصـر، قادر اسـت هاله‌های پراکنده و چندعنصـری را نیز مدل‌سازی و شناسایی کند. برای مثال، اگر مجموعه‌ای از عناصر ردیـاب بـه میزان کمی بـالاتر از زمینـه در محـدوده‌ای گسـترده

DOI: 10.22067/econg.2025.1157

افزایش‌یابند (الگویی که معمولاً پیرامون کانسارهای پنهان مشاهده می‌شود)، AE این الگوی پنهان را از طریق خطای بازسازی بالا تشخیص خواهد داد؛ در حالی‌که IF ممکن است به دلیل توزیع نسبتا پیوسته این داده‌ها، آنها را از زمینه تفکیک نکند. الگوریتم دی‌بی اسکن نیز به طور ذاتی به ساختار فضایی داده‌ها حساس است.در یک پس‌زمینه ناهمگن (متشکل از چند توزیع مختلف)، دی‌بی اسکن احتمال دارد برخی نقاط متعلق به یک جمعیت پس‌زمینه را به عنوان نویز جدا کند و در نتیجه محدوده دسته آنومالی را بیش از حد گسترش دهد. این همان دلیلی است که در نقشه ما، دسته آنومالی دی‌بی اسکن وسیع‌تر از AE شد (۴۷ درصد در برابر ۳۶ درصد مساحت). در مجموع، این تفاوت‌ها نشان می‌دهد AE به خوبی توانسته است هاله‌های ضعیف ناشی از پراکندگی زمین‌شیمیایی را تشخیص‌دهد؛ در حالی که IF بیشتر مناسب یافتن مناطق کانونی با آنومالی قوی (هسته‌های پرعیار) است و دی‌بی اسکن برای داده‌های با توزیع‌های چندگانه کمتر دقیق عمل می‌کند. این تحلیل علّی کاملاً با ماهیت داده‌های ما – شامل هاله‌های پهن و مناطق کانونی – سازگار است و بر درستی انتخاب AE به عنوان مدل برتر صحه می‌گذارد.

نتیجه‌گیری

روش‌های یادگیری عمیق، به ویژه در سال‌های اخیر، به شکل گسترده‌ای در حوزه‌های مختلف صنعتی و علمی مورد استفاده قرار گرفته‌اند و کاربردهای آنها در زمینه اکتشاف مواد معدنی نیز رو به افزایش است. در اکتشافات معدنی، روش‌های زمین‌شیمیایی و ژئوفیزیکی با استفاده از فناوری‌های نوین مانند یادگیری عمیق و شبکه‌های عصبی، امکان تحلیل داده‌های پیچیده را به شکل موثرتری فراهم می‌آورند. با پیشرفت‌هایی که در فناوری‌های مربوط به جمع‌آوری و پردازش داده‌ها ایجادشده است، از جمله استفاده از پهپادها و ماهواره‌ها برای جمع‌آوری گسترده و پردازش سریع اطلاعات پیچیده توسط الگوریتم‌های یادگیری عمیق، دقت و کارایی در شناسایی آنومالی‌ها افزایش چشمگیری داشته است.

در این پژوهش، ما برای هرکدام از مدل‌های جنگل ایزوله، دی‌بی اسکن، خودرمزنگار و خودرمزنگار واریانسی یک فرایند مؤثر برای پیداکردن مقادیر هایپرپارامترهای هر مدل معرفی کردیم و سپس نتایج به دست آمده برای مدل‌ها را با استفاده از نقاط کانی‌زایی شناخته‌شده در منطقه مورد بررسی مورد مقایسه قرار دادیم. چنان‌که در شکل ۳ مشخص است، مجموعه داده انتخابی یک مجموعه داده پیچیده متشکل از دو محدوده زمین‌شناسی یکصدهزار است که همان‌طور که در شکل ۳ نیز مشخص است، در این مجموعه داده رابطه‌های اکتشافی و زمین‌شیمیایی به خوبی مشخص نبوده و عناصر خاص ضریب همبستگی کمی را از خود نشان می‌دهند. در این فرایند از روش جست‌وجوی بایزی بهره بردیم که امکان بهینه‌سازی مؤلفه‌های مختلف و اجرای تعداد زیادی مدل با تنظیمات متفاوت را فراهم کرد. این فرایندها به ما کمک‌کرد تا بهترین مدل‌های ممکن را انتخاب‌کرده و به مقایسه عملکرد آنها بپردازیم.

به طور کلی ۳۹ نقطه کانی‌زایی شناخته‌شده در منطقه مورد بررسی قرار داشت که طبق جدول ۵، مدلی که بیشترین تعداد نقاط کانی‌زایی را در دسته آنومالی بالا تشخیص‌داده و کمترین درصد پوشش این دسته را در بین مدل‌های دیگر داشته، به عنوان مدل بهینه انتخاب‌شده است. مدل خودرمزنگار با تعداد ۲۴ نقطه کانی‌زایی و درصد پوشش دسته آنومالی بالای ۳۶ درصد به عنوان مدلی که بهترین عملکرد را داشت، انتخاب‌شد و بعد از آن مدل خودرمزنگار واریانسی با ۴۲ درصد پوشش و تعداد ۲۴ نقطه کانی‌زایی به عنوان مدل دوم انتخاب‌شد. جنگل ایزوله به سبب ساختار درختی و تأکید بر طول مسیر، هسته‌های با سیگنال قوی را بازیابی می‌کند؛ اما بخش‌هایی از هاله‌های ضعیف می‌تواند از دست برود. دی‌بی اسکن در زمینه‌های ناهمگن، با وجود توانایی تفکیک خوشه–نویز، مستعد بارز کردن دسته‌ها با مساحت بالاست.AE/VAE به‌دلیل یادگیری روابط غیرخطی چندعنصری، هاله‌ها را بهتر مدل می‌کنند.

DOI: 10.22067/econg.2025.1157

به طور کلی، این نتایج نشـــان‌دهنده پتانســـیل بالای روش‌های یادگیری عمیق در رفع محدودیت‌های روش‌های ســـنتی و افزایش دقت در شناسایی آنومالی‌های زمین‌شیمیایی است. این پیشرفت‌ها، همراه بـا بهبود در فن‌آوری‌هـای جمع‌آوری داده‌هـا و پردازش اطلاعات، نشان می‌دهد که آینده اکتشافات معدنی در گرو استفاده از چنین روش‌های نوین و قدرتمندی است.

برای طراحی پیمایش میدانی، گام‌های زیر توصیه می‌شود:

اولویت اولیه: حوضـه‌هایی که در نقشـه AE (شـکل ۱۰) در دسـته آنومالی بالا قرار دارند، بررسی شوند. این نقشه با ۳۹/۲۴ رخداد در ۳۶ درصد مساحت (جدول ۶) بهترین نقطه شروع را فراهم می‌کند.

اولویت دوم: برای گســترش دامنه اهداف به پیرامون هســته‌ها، از دی‌بی اسـکن (شـکل ۸) و VAE (شـکل ۱۱) به عنوان لایه‌های پشـتیبان اسـتفاده شـود (به‌ترتیب ۴۷ و ۴۲ درصـد پوشـش با ۳۹/۲۴ رخداد).

هسته‌های پرعیار: نقشه IF (شکل ۷) به دلیل تمرکز بر سیگنال‌های قوی برای نشـان‌دار کردن هسـته‌های پرعیار مناسـب اسـت (۳۹/۱۰ رخداد در ۲۱ درصد مساحت).

جدول ٦. درصد مناطق تحت پوشش از دسته آنومالی بالا و تعداد نقاط کانی‌زایی شناخته شده در این دسته در منطقه پاریز و چهارگنبد

Table 6. Percentage of areas covered by the high anomaly class and the number of known mineralized points in this class in Pariz and Chargonbad area

Model	Percentage of areas covered by high anomaly class	Number of known mineralized points in the high anomaly class
Isolation forest	21%	10
DBscan	47%	24
Autoencoder	36%	24
Variational Autoencoder	42%	24

به طور کلی، یافته‌های کلیدی زمین‌شناسی و زمین‌شیمی را از رفتار مدل‌ها می‌توان به صورت زیر تقسیم‌بندی کرد:

- هاله‌های زمین‌شـیمیایی چندعنصـری به خوبی توسـط مدل‌های یادگیری عمیق شـناسـایی شـدند. در این بررسـی، مشـخص‌شـد که آنومالی‌های گسـترده مرتبط با کانی‌سـازی پورفیری (که با افزایش هم‌زمان عناصـر متعددی نظیر Cu، Mo، Pb در پیرامون توده‌های نفوذی همراه هسـتند) توسـط خودرمزنگار عمیق (AE) و مدل واریانسـی (VAE) به صـورت یک دسـته آنومالی پیوسـته نمایان می‌شـوند. این نشـان می‌دهد روش‌های یادگیری عمیق قادر به تشـخیص سـیگنال‌های مرکب حاصـل از فرایندهای زمین‌شـیمیایی پیچیده (مانند دگرسـانی‌های وسـیع گرمابی) هسـتند؛ سـیگنال‌هایی که در رویکردهای سنتی آستانه‌ای ممکن بود پنهان بمانند.

- الگوی فضایی آنومالی‌ها منعکس‌کننده کنترل‌های زمین‌شناختی منطقه اسـت. آنومالی‌های به دسـت‌آمده اغلب همراسـتا با سـاختارهای زمین‌شـناسـی شـناخته‌شـده و واحدهای سـنگی میزبان کانی‌سازی‌اند. برای مثال، نقشه‌های ما نشان‌داد که بالاترین تراکم آنومالی‌ها در مجاورت توده‌های نفوذی ائوسـن–الیگوسـن و در امتداد امتداد گسـل‌های NE–SW منطقه دیده می‌شـود. این تطابق به این معناسـت که مدل، اثر پراکندگی ثانویه کانی‌سـازی را به درستی دنبال‌کرده و رد پای ساختاری–زمین‌شناختی کانه‌زایی را

DOI: 10.22067/econg.2025.1157

گرفته است. در نتیجه آنومالی‌های شناسایی‌شده معنای زمین‌شناختی مشخصی دارند و فقط تصادفی نیستند.

- تفاوت نقشه‌های آنومالی مدل‌های مختلف، بیانگر مقیاس‌های مختلف پراکندگی زمین‌شیمیایی است. جنگل ایزوله یک سری آنومالی‌های نقطه‌ای کوچک و پراکنده ارائه‌داد که بیانگر زون‌های کانونی و بسیار پربار (احتمالاً نزدیک به مرکز کانسار یا رخنمون‌های سطحی) است. در مقابل، خودرمزنگار عمیق و دی‌بی اسکن نواحی بسیار وسیع‌تری را به عنوان آنومالی مشخص کردند که نشان‌دهنده هاله‌های دورتری است که توسط فرایندهای رسوبی و آبراهه‌ای جابه‌جا شده‌اند. این به این معناست که در داده‌های منطقه ما هر دو نوع آنومالی حضور دارد: آنومالی‌های موضعی قوی و آنومالی‌های منطقه‌ای ضعیف‌تر. مدل AE توانست هر دوی این مقیاس‌ها را پوشش دهد (هاله‌های وسیع با حفظ تمرکز بر نواحی پربار درون آنها) و این یکی از دلایل عملکرد برتر آن از نظر توازن دقت و پوشش بود.

یافته‌های این پژوهش فقط به تکرار کشفیات پیشین محدود نمی‌شود؛ بلکه افق‌های تازه‌ای برای اکتشاف ترسیم می‌کند. نقشه‌های آنومالی ما تعدادی محدوده را نمایان ساختند که با وجود نبود هرگونه رخداد معدنی شناخته‌شده در آنها، از دید مدل دارای پتانسیل بالایی برای کانی‌سازی هستند. به طور مشخص، چند حوضه آبریز در بخش‌هایی از منطقه که تاکنون اکتشاف نظام‌مندی به خود ندیده‌اند، در نقشه AE به عنوان آنومالی‌های رده بالا ظاهر شده‌اند. این محدوده‌ها را می‌توان به عنوان اهداف اکتشافی جدید معرفی کرد تا در برنامه‌های آتی، بررسی‌های میدانی تفصیلی (نمونه‌برداری تکمیلی، نقشه‌برداری زمین‌شناسی و حتی حفاری اکتشافی) بر روی آنها صورت گیرد. مدل‌های ما بر روی داده‌های یک منطقه خاص (پاریز– چهارگنبد) آموزش و تنظیم شده‌اند. هرچند اقدامات اعتبارسنجی مکانی را انجام دادیم؛ اما هنوز ضمانتی نیست که به طور دقیق با همین کارایی در یک منطقه کاملاً جدید با زمین‌شناسی و توزیع‌های آماری متفاوت عمل کنند. برای مثال، اگر منطقه‌ای دارای آنومالی‌های بسیار موضعی‌تر یا برعکس آنومالی‌های بسیار پخش‌تر باشد، تنظیمات بهینه فعلی ممکن است نیاز به بازنگری داشته باشد. همچنین برخی فرضیه‌های مدل (مانند درصد کوچک آنومالی در داده) ممکن است در یک بررسی منطقه‌ای وسیع صدق نکند. بنابراین، نتیجه‌های ما باید در حوزه مورد بررسی فعلی تعبیر شوند و هنگام تعمیم به سایر کمربندها، احتمال نیاز به بازآموزی یا تنظیم مجدد مدل‌ها وجود دارد. این محدودیت را می‌توان با گردآوری داده‌های بیشتر از مناطق مختلف و ایجاد یک مدل کلی‌تر رفع کرد.

تعارض منافع

هیچ‌گونه تعارض منافعی توسط نویسندگان بیان‌نشده است.

DOI: 10.22067/econg.2025.1157

1. Optuna
2. Principal component analysis (PCA)
3. Support vector machine (SVM)
4. Contamination Factor
5. Grid search
6. Random search
7. contamination
8. Density-Based Spatial Clustering of Applications with Noise (DBSCAN)
9. Encoder
10. Decoder
11. Latent space
12. Mean Squared Error
13. Fe polymetallic
14. Bayesian optimization
15. Evolutionary Algorithms
16. number of estimators
17. contamination fraction
18. maximum samples
19. average decision function score
20. min_samples
21. Average Adjusted Distance Score
22. Mean Reconstruction Error
23. Bottleneck Layer
24. Batch Size
25. Gradient Clipping
26. Overfitting
27. Kullback-Leibler Divergence - KL
28. scikit learn
29. DBSCAN Outlier Indicator
30. Adjusted Distance Score
31. Hit-rate
32. recall

References

Adankon, M.M., Dansereau, J. and Cheriet, F., 2012. Analysis of scoliosis trunk deformities using ica. 11th International Conference on Information Science, Signal Processing and Their Applications (ISSPA). https://doi.org/10.1109/isspa.2012.6310543

Andriyani, F. and Puspitarani, Y., 2022. Performance comparison of k-means and dbscan algorithms for text clustering product reviews. SinkrOn, 6(3): 944–949. https://doi.org/10.33395/sinkron.v7i3.11569

Berberian, M. and King, G.C.P., 1981. Towards a paleogeography and tectonic evolution of Iran. Canadian Journal of Earth Sciences, 18(2): 210–265. https://doi.org/10.1139/e81-019

Bulut, O., Gorgun, G. and He, S., 2024. Unsupervised anomaly detection in sequential process data. Zeitschrift Für Psychologie, 232(2): 74–94. https://doi.org/10.1027/2151-2604/a000558

Chen, L., Guan, Q., Feng, B., Yue, H., Wang, J. and Zhang, F., 2019a. A multi-convolutional autoencoder approach to multivariate geochemical anomaly recognition. Minerals, 9(5): 270. https://doi.org/10.3390/min9050270

Chen, L., Guan, Q., Xiong, Y., Liang, J., Wang, Y. and Xu, Y., 2019b. A spatially constrained multi-autoencoder approach for multivariate geochemical anomaly recognition. Computers & Geosciences, 125: 43–54. https://doi.org/10.1016/j.cageo.2019.01.016

Chen, J. and Okle, R.A.N., 2024. A study on the application of response surface methodology and bayesian optimization in parameter tuning of genetic algorithms. Fourth International Conference on Applied Mathematics, Modelling, and Intelligent Computing (CAMMIC 2024), Kaifeng, China. https://doi.org/10.1117/12.3036921

Ester, M., Kriegel, H., Sander, J. and Xu, X., 1996. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. 2nd International Conference on Knowledge Discovery and Data Mining. Retrieved July 19, 2025 from https://www.dbs.ifi.lmu.de/Publikationen/Papers/KDD-96.final.frame.pdf

Farhadi, S., Tatullo, S., Boveiri Konari, M. and Afzal, P., 2024. Evaluating StackingC and ensemble models for enhanced lithological classification in geological mapping. Journal of Geochemical Exploration, 260: 107441 https://doi.org/10.1016/j.gexplo.2024.107441

Feng, B., Chen, L. and Xu, Y., 2022. Comparative study on three autoencoder-based deep learning algorithms for geochemical anomaly identification. Earth and Space Science, 9(11). https://doi.org/10.1029/2022ea002626

Ghawi, R. and Pfeffer, J., 2019. Efficient hyperparameter tuning with grid search for text categorization using knn approach with bm25 similarity. Open Computer Science, 9(1): 160-180. https://doi.org/10.1515/comp-2019-0011

Hariri, S., Kind, M.C. and Brunner, R., 2021. Extended isolation forest. IEEE Transactions on Knowledge and Data Engineering, 33(4): 1479–1489. https://doi.org/10.1109/tkde.2019.2947676

Hezarkhani, A., 2006. Petrology of the intrusive rocks within the Sungun Porphyry Copper Deposit, Azerbaijan, Iran. Journal of Asian Earth Sciences, 27(3): 326–340. https://doi.org/10.1016/j.jseaes.2005.04.005

Horrocks, T., 2019. Integrated analysis of geological, geophysical, and geochemical data of the kevitsa ni-cu-pge deposit: machine learning approaches. Ph.D. Thesis, The University of western Australia, Crawley, Western Australia, Australia,180 pp. https://doi.org/10.26182/5c5ba16237b57

Iwamori, H., Yoshida, K., Nakamura, H., Kuwatani, T., Hamada, M., Haraguchi, S., and Ueki, K., 2017. Classification of geochemical data based on multivariate statistical analyses: complementary roles of cluster, principal component, and independent component analyses. Geochemistry, Geophysics, Geosystems, 18(3): 994–1012. https://doi.org/10.1002/2016gc006663

Jin, Z., Risk, B.B. and Matteson, D.S., 2019. Optimization and testing in linear non-gaussian component analysis. Statistical Analysis and Data Mining: The ASA Data Science Journal, 12(3): 141–156. https://doi.org/10.1002/sam.11403

Kingma, D.P. and Welling, M., 2013. Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114. https://doi.org/10.48550/arXiv.1312.6114

Li, H., Correa, N.M., Rodríguez, P. and Calhoun, V. D., 2011. Application of independent component analysis with adaptive density model to complex-valued fmri data. IEEE Transactions on Biomedical Engineering, 58(10): 2794–2803. https://doi.org/10.1109/tbme.2011.2159841

Li, L., Jamieson, K.G., DeSalvo, G., Rostamizadeh, A. and Talwalkar, A., 2016. Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization. Journal of Machine Learning Research 18(2018): 1–52. https://doi.org/10.48550/arxiv.1603.06560

Li, H., Li, X., Yuan, F., Jowitt, S.M., Zhang, M., Zhou, J., Zhou, T., Li, X., Ge, C. and Wu, B., 2020. Convolutional neural network and transfer learning based mineral prospectivity modeling for geochemical exploration of au mineralization within the guandian–zhangbaling area, anhui province, china. Applied Geochemistry, 122: 104747. https://doi.org/10.1016/j.apgeochem.2020.104747

Luo, Z., Xiong, Y. and Zuo, R., 2020. Recognition of geochemical anomalies using a deep variational autoencoder network. Applied Geochemistry, 122: 104710. https://doi.org/10.1016/j.apgeochem.2020.104710

Luo, Z., Zuo, R., Xiong, Y. and Wang, X., 2021. Detection of geochemical anomalies related to mineralization using the ganomaly network. Applied Geochemistry, 131: 105043. https://doi.org/10.1016/j.apgeochem.2021.105043

Pourgholam, M.M., Adib, A., Afzal, P., Rahbar, K. and Gholinejad, M., 2025. Deep learning and fractal-wavelet techniques for magnetite-apatite exploration in Tarom Iran. Scientific Reports 15 (1): 31907. https://doi.org/10.1038/s41598-025-16040-2

Praveena, S.M., Kwan, O.W. and Aris, A.Z., 2011. Effect of data pre-treatment procedures on principal component analysis: a case study for mangrove surface sediment datasets. Environmental Monitoring and Assessment, 184(11): 6855–6868. https://doi.org/10.1007/s10661-011-2463-2

Rezende, D., Mohamed, S. and Wierstra, D., 2014. Stochastic Back-propagation and Variational Inference in Deep Latent Gaussian Models. ArXiv,abs/1401.4082. https://doi.org/10.48550/arxiv.1401.4082

Richards, J.P., 2003. Tectono-magmatic precursors for porphyry Cu-(Mo-Au) deposit formation. Economic Geology, 98(8): 1515–1533. https://doi.org/10.2113/gsecongeo.98.8.1515

Rijn, J.N.v. and Hutter, F., 2018. Hyperparameter importance across datasets. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 2367-2376. https://doi.org/10.1145/3219819.3220058

Sarikhani, P., Ferleger, B.I., Mitchell, K.T., Ostrem, J.L., Herron, J.A., Mahmoudi, B. and Miocinovic, S., 2022. Automated deep brain stimulation programming with safety constraints for tremor suppression in patients with parkinson's disease and essential tremor. Journal of Neural Engineering, 19(4): 046042. https://doi.org/10.1088/1741-2552/ac86a2

Snoek, J., Larochelle, H. and Adams, R.P., 2012. Practical Bayesian Optimization of Machine Learning Algorithms. Neural Information Processing Systems. https://doi.org/10.48550/arxiv.1206.2944

Taghipour, N., Aftabi, A. and Mathur, R., 2008. Geology and Re-Os geochronology of mineralization of the miduk porphyry copper deposit, Iran. Resource Geology, 58(2): 143–160. https://doi.org/10.1111/j.1751-3928.2008.00054.x

Thavasimani, K. and Srinath, N.K., 2022. Hyperparameter optimization using custom genetic algorithm for classification of benign and malicious traffic on internet of things–23 dataset. International Journal of Electrical and Computer Engineering (IJECE), 12(4): 4031. https://doi.org/10.11591/ijece.v12i4.pp4031-4041

Thu, T.N.T. and Armitage, G., 2008. A survey of techniques for internet traffic classification using machine learning. IEEE Communications Surveys & Tutorials, 10(4): 56–76. https://doi.org/10.1109/surv.2008.080406

Ueki, K., Hino, H. and Kuwatani, T., 2018. Geochemical discrimination and characteristics of magmatic tectonic settings: a machine-learning-based approach. Geochemistry, Geophysics, Geosystems, 19(4): 1327–1347. https://doi.org/10.1029/2017gc007401

Verma, S.P., 1997. Sixteen statistical tests for outlier detection and rejection in evaluation of international geochemical reference materials: example of microgabbro pm-s. Geostandards Newsletter, 21(1): 59–75. https://doi.org/10.1111/j.1751-908x.1997.tb00532.x

Wein, S., Tomé, A.M., Goldhacker, M., Greenlee, M.W. and Lang, E., 2020. A constrained ica-emd model for group level fmri analysis. Frontiers in Neuroscience, 14. https://doi.org/10.3389/fnins.2020.00221

Xiong, Y. and Zuo, R., 2016. Recognition of geochemical anomalies using a deep autoencoder network. Computers & Geosciences, 86: 75–82. https://doi.org/10.1016/j.cageo.2015.10.006

Xiong, Y. and Zuo, R., 2020. Recognizing multivariate geochemical anomalies for mineral exploration by combining deep learning and one-class support vector machine. Computers & Geosciences, 140: 104484. https://doi.org/10.1016/j.cageo.2020.104484

Yang, N., Zhang, Z., Yang, J., Hong, Z. and Shi, J., 2021. A convolutional neural network of googlenet applied in mineral prospectivity prediction based on multi-source geoinformation. Natural Resources Research, 30(6): 3905–3923. https://doi.org/10.1007/s11053-021-09934-1

Yasukawa, K., Nakamura, K., Fujinaga, K., Iwamori, H. and Kato, Y., 2016. Tracking the spatiotemporal variations of statistically independent components involving enrichment of rare-earth elements in deep-sea sediments. Scientific Reports, 6: 29603. https://doi.org/10.1038/srep29603

Zhang, S., Carranza, E.J.M., Fu, C., Wen-zhi, Z. and Xiang, Q., 2024. Interpretable machine learning for geochemical anomaly delineation in the yuanbo nang district, gansu province, china. Minerals, 14(5): 500. https://doi.org/10.3390/min14050500

Zhang, S., Xiao, K., Carranza, E.J.M., Yang, F. and Zhao, Z., 2019. Integration of auto-encoder network with density-based spatial clustering for geochemical anomaly detection for mineral exploration. Computers & Geosciences, 130: 43–56. https://doi.org/10.1016/j.cageo.2019.05.011

Zhang, C. and Zuo, R., 2021. Recognition of multivariate geochemical anomalies associated with mineralization using an improved generative adversarial network. Ore Geology Reviews, 136: 104264. https://doi.org/10.1016/j.oregeorev.2021.104264

Zhang, C. and Zuo, R., 2024. Incorporating geological knowledge into deep learning to enhance geochemical anomaly identification related to mineralization and interpretability. Mathematical Geosciences, 56(6): 1233–1254. https://doi.org/10.1007/s11004-023-10133-2

Zuo, R., 2011. Identifying geochemical anomalies associated with cu and pb–zn skarn mineralization using principal component analysis and spectrum–area fractal modeling in the gangdese belt, tibet (china). Journal of Geochemical Exploration, 111(1–2): 13–22. https://doi.org/10.1016/j.gexplo.2011.06.012