



Comparison of efficiency of random forest and support vector machine methods for Mineral Potential Mapping of copper deposits - Case study: Dahaj-Bazman

Mohammad Karimi^{*1}, Parastoo Pilehforooshha², Ali Safari³

Date of receive:

2023.09.26

Date of last review:

2023.12.30

Date of accept:

2024.04.20

Date of online publication:

2024.04.20

Extended Abstract

Introduction

The exploration and preparation of the potential map of mineral reserves requires the use of various methods and techniques, based on the geological and mining knowledge of the investigated area, and the use of predictive models of mineral potential (Bonham-Carter, 1994; Carranza et al., 2008a). According to the investigations, the common models of map integration that are used in the discovery of mineral reserves in the initial exploration stage include index overlap model, fuzzy operators, weighted indicators and smart methods such as random forests and artificial networks. Determining the values of weights and scores that show the relative importance of the effective factors is the primary requirement in combining the maps and preparing the mineral potential map (Agterberg, 1992; Brown et al., 2000).

The purpose of this research is to prepare a potential map of copper deposits in Dehj-Bazman region using two methods of random forest and support vector machine. In addition, in order to compare the potential map of porphyry copper reserves resulting from the random forest method, the support vector machine method and the knowledge-based methods of index overlap and fuzzy logic were used.

Materials & Methods

The area studied in this research is a part of the magmatic belt of Kerman region, known as the Dehj-Sardouye belt. The information layers controlling mineralization in Dehj-Bazman area include rock units, structures, alterations, geochemistry, geophysics and copper deposits. In practical applications of machine learning algorithms, mineral potential mapping is essentially a bimodal classification problem, such that each undiscovered area is classified as prospective or non-prospective according to some combination of mapping criteria (Zuo, 2011). The final results are a set of predictive maps that show target areas with high ore formation potential. In order to model, training was done. Before training the random forest model, the input data set and the target variable should be prepared and then the model should be trained. The target variables for entering the random forest model and support vector machine were determined as deposit points (values of 1) and non-deposit points (values of 0). Then the genetic algorithm was used to adjust the parameters ... ▶ Page 48

Keywords:

Geospatial Information System,
Porphyry Copper Potential Mapping,
Random forests,
Support vector machine

1- (*Corresponding author) Associate Professor, Faculty of Geodesy and Geomatic, K.N.Toosi University of Technology, mkarimi@kntu.ac.ir

2- Assistant Professor, Faculty of Geodesy and Geomatic, K.N.Toosi University of Technology, pilehforoosh.p@gmail.com

3- Master of Science in GIS, Faculty of Geodesy and Geomatic, K.N.Toosi University of Technology, asafarigis97@email.kntu.ac.ir

Evaluation of the predictive performance of random forest model and support vector machine can be described by the ambiguity matrix. In this matrix, there are four components, which are defined as: (1) a deposit sample that is correctly classified as a deposit (TP); (2) a deposit sample incorrectly classified as a non-deposit sample (FN), (3) a non-deposit sample correctly classified as a non-deposit sample (TN), and (4) a non-deposit sample that is wrongly classified as a deposit sample (FP) (Liu et al., 2005; Tien Bui et al., 2016):

(8) "Sensitivity" = $TP/(TP+FN)$

(9) "Specificity" = $TN/(TN+FP)$

(10) "Positive predictive value" = $TP/(TP+FP)$

(11) "Negative predictive value" = $TN/(TN+FN)$

(12) "Accuracy" = $(TP+TN)/(TP+TN+FP+FN)$

After training and evaluating different models, the best model was obtained by adjusting different parameters and it was used to integrate factor maps in order to predict areas with high potential of porphyry copper deposits. Also, knowledge-based methods of fuzzy logic and index overlap were used to combine factor maps to compare with the results of intelligent methods.

Results & Discussion

At this stage, the desired information layers were collected and prepared in the GIS environment, and then factor maps were prepared. Accuracy, sensitivity, specificity, predicted positive value, predicted negative value, kappa index and OOB error were used to evaluate the performance of random forest model and support vector machine. Also, the importance of the predictor variables in the random forest model was evaluated through the mean decrease in accuracy and the mean decrease in node impurity or the Gini impurity index (Breiman, 2001). According to the results, the most important predictor in the random forest model is the geochemical map, while the structures factor has the least impact in predicting the preparation of the mineral potential map with the final random forest model. In the potential maps of porphyry copper deposits obtained from two methods of random forest and support vector machine, the target areas cover 14% of the studied area, in which there are 92% and 87% of known deposits, respectively. Finally, the efficiency of machine learning methods and knowledge-based methods were compared. In order to produce porphyry

copper potential map with knowledge-based methods, the judgment of expert experts was used to assign weights to each criterion map. For this purpose, weights of 0.3, 0.25, 0.25, 0.1, 0.1 were assigned to produce maps of alteration factor, geochemistry, geology, geophysics and structures respectively. In the potential map obtained from the method of index overlap and fuzzy logic (fuzzy sum), the areas predicted as copper mines cover 16 and 17 percent of the studied area, respectively, in which 83 and 79 percent of the existing mines are located.

Conclusion

This research was conducted with the aim of evaluating and comparing the effectiveness of random forest method and support vector machine method and knowledge-based methods to prepare porphyry copper potential map of Dehaj-Bozman region of Kerman province. Based on the results, the random forest model works well in the field of porphyry copper potential map preparation with geochemical, geophysical, geological, alteration and structures datasets. In addition, the random forest algorithm can estimate the importance of factor maps.

The results of this research show that the geochemical factor map is the most important and the structure factor map is the least important in predicting the data-driven model of random forests. This estimate of importance is consistent with geological knowledge about porphyry copper mineralization in Dehaj-Buzman region. In order to produce porphyry copper potential map with knowledge-based methods, the judgment of expert experts was used to assign weights to each criterion map. According to the obtained results, the performance of the random forest model is better than the vector machine model, and also, the performance of the support vector machine model is better than the knowledge-based methods.



مقایسه کارایی روش‌های جنگل تصادفی و ماشین بردار پشتیبان

به منظور پتانسیل‌یابی ذخایر معدنی مس - مطالعه موردی: منطقه دهج - بزمان

محمد کریمی^۱، پرستو پیله‌فروش‌ها^۲، علی صفری^۳

چکیده

با توجه به وسعت زیاد کشور ایران و گستردگی مناطق پتانسیل‌دار ذخایر معدنی (وجود کمربند ولکانیکی ارومیه - دختر) و لزوم شناسایی و مدیریت صحیح این ذخایر، استفاده از سیستم اطلاعات مکانی به همراه مدل‌های پیش‌بینی‌کننده داده و دانش‌محور، نقش بسیار مهمی در تهیه نقشه پتانسیل از احتمال یافتن ذخایر معدنی در یک مکان خاص دارد. هدف این تحقیق پیش‌بینی ذخایر مس پورفیری در منطقه دهج - بزمان استان کرمان با استفاده از دو روش جنگل‌های تصادفی^۴ و ماشین بردار پشتیبان^۵ است. به این منظور، از یک پایگاه داده مکانی متشکل از نقشه‌های جنس واحدهای سنگی، ساختارها، آلتراسیون، ژئوشیمی، ژئوفیزیک و موقعیت ۲۴ کانسار مس پورفیری شناخته‌شده در منطقه استفاده شد. با توجه به نتایج حاصل شده، مدل جنگل‌های تصادفی توانست با صحت ۹۳/۳۳ درصد مناطق امیدبخش ذخایر مس پورفیری را پیش‌بینی کند. همچنین، در نقشه پتانسیل به‌دست آمده از این مدل، مناطق هدف ۱۴ درصد از منطقه مورد مطالعه را در بر گرفته است، که در آن ۹۲ درصد ذخایر شناخته شده مشخص شده‌اند. علاوه بر این، به‌منظور مقایسه نقشه پتانسیل ذخایر مس پورفیری منتج از روش جنگل‌های تصادفی، از روش ماشین بردار پشتیبان و روش‌های دانش‌محور همپوشانی شاخص و منطق فازی استفاده شد. در نقشه‌های پتانسیل ذخایر مس پورفیری به‌دست آمده از سه روش ماشین بردار پشتیبان، همپوشانی شاخص و منطق فازی به‌ترتیب مناطق هدف ۱۷،۱۶،۱۴ درصد از منطقه مورد مطالعه را در بر گرفته است که در آن‌ها ۷۹،۸۳،۸۷ درصد ذخایر شناخته شده وجود دارند. براساس نتایج این تحقیق، مدل جنگل‌های تصادفی از نظر صحت پیش‌بینی از کارایی بالاتری نسبت به مدل‌های دیگر برخوردار بوده و مدل‌های ماشین بردار پشتیبان، همپوشانی شاخص و منطق فازی به‌ترتیب در رتبه‌های بعدی قرار دارند.

تاریخ دریافت:

۱۴۰۲/۰۷/۰۴

تاریخ آخرین بازنگری:

۱۴۰۲/۱۰/۰۹

تاریخ پذیرش:

۱۴۰۳/۰۲/۰۱

تاریخ انتشار:

۱۴۰۳/۰۲/۰۱

واژه‌های کلیدی:

سیستم اطلاعات مکانی؛

نقشه پتانسیل مس پورفیری؛

جنگل‌های تصادفی؛

ماشین بردار پشتیبان

۱- دانشیار دانشکده مهندسی نقشه‌برداری، دانشگاه صنعتی خواجه نصیرالدین طوسی (نویسنده مسئول) mkarimi@kntu.ac.ir

۲- استادیار دانشکده مهندسی نقشه‌برداری، دانشگاه صنعتی خواجه نصیرالدین طوسی pilehforoosh.p@gmail.com

۳- دانش‌آموخته کارشناس ارشد سیستم اطلاعات مکانی، دانشکده مهندسی نقشه‌برداری، دانشگاه صنعتی خواجه نصیرالدین طوسی asafarigis97@email.kntu.ac.ir

4- Random Forest (RF)

5- Support Vector Machine (SVM)

۱- مقدمه

امتیازهایی که اهمیت نسبی عوامل مؤثر را نشان می‌دهند، نیاز اولیه در تلفیق نقشه‌ها و تهیه نقشه پتانسیل معدنی است (Agterberg, 1992; Brown et al., 2000).

به‌طور کلی، دو دسته اصلی روش‌های داده‌محور و دانش‌محور برای وزن‌دهی اطلاعات وجود دارد. در روش‌های داده‌محور، اهمیت داده‌ها را خود آن‌ها تعیین می‌کنند، در حالی که در روش‌های دانش‌محور، فردی متخصص یا گروهی از متخصصان این امر مهم را برعهده دارند. علاوه بر روش‌های فوق، روش‌های یادگیری ماشین به‌عنوان ابزاری بسیار کارآمد برای وزن‌دهی اطلاعات کاربرد دارند (Lary et al., 2016; Li et al., 2017). در مقایسه با روش‌های داده‌محور و دانش‌محور، روش‌های یادگیری ماشین مزیت را دارند که هیچ‌گونه فرضی در مورد توزیع داده را نیاز نداشته و می‌توانند روابط غیرخطی بین ذخایر معدنی و خصوصیات نقشه‌های فاکتور را کنترل کنند (Zuo, 2017).

تحقیقات متعددی در زمینه‌ی پتانسیل‌یابی مواد معدنی با استفاده از روش‌های داده‌محور، روش‌های دانش‌محور و روش‌های یادگیری ماشین انجام شده است. در زمینه روش‌های داده‌محور، روش‌های متداول مورد استفاده شامل شبکه‌های عصبی مصنوعی (Brown et al., 2000; Porwal et al., 2005) و اوزان شواهد (Bonham-Carter, 2003a; Chen et al., 2005) (WofE) (Bonham-Carter et al., 1988; Bonham-Carter et al., 1989; Carranza and Hale, 2000; Rigol-Sanchez et al., 2003) و تابع گمان شواهد (Carranza and Hale, 2003; Carranza et al., 2005) می‌شود.

در زمینه استفاده از روش‌های دانش‌محور می‌توان به استفاده از منطق بولین (کریمی و همکاران، ۱۳۸۷)، منطق فازی (Bonham-Carter, 1994; Alaei Moghadam et al., 2015) و همکاران، ۱۳۹۳) و روش همپوشانی شاخص چند کلاسه (Alaei Moghadam et al., 2015) و همکاران، ۱۳۸۷) اشاره کرد. علاوه بر روش‌های فوق، قابلیت مدل‌سازی بسیار بالای روش‌های یادگیری ماشین منجر به کاربرد گسترده آن‌ها در تهیه نقشه پتانسیل معدنی مبتنی بر سیستم اطلاعات مکانی شده است. در این خصوص می‌توان به بررسی روش‌های

اکتشاف منابع معدنی و ذخایر زیرزمینی از فعالیت‌های اقتصادی-راهبردی در کشور ایران شمرده می‌شود. در این راستا، کشف کانسارهای جدید به‌منظور جایگزینی معادن در حال بهره‌برداری یکی از اهداف بلندمدت و استراتژیک کشورهای دارای منابع معدنی محسوب می‌شود. از این‌رو، انجام اکتشاف سیستماتیک ناحیه‌ای امری ضروری است. با این حال، از آنجایی که عملیات حفاری با صرف هزینه‌های سنگین و محدودیت سرمایه‌گذاری همراه است، امکان حفاری کل منطقه وجود نداشته و در این زمینه روش‌های اکتشافی می‌توانند مناطقی را به‌عنوان مناطق مستعد کانی‌سازی معرفی نمایند تا اقدام به حفاری در آن مناطق شود (مومنی، ۱۳۹۲). در این زمینه، سیستم اطلاعات مکانی^۱ به ابزاری اساسی با هدف تصمیم‌گیری در مورد مناطق مناسب برای اکتشاف مواد معدنی تبدیل شده است (Bonham-Carter, 1994; Carranza et al., 2008a).

اکتشاف معدن فعالیت چندمرحله‌ای است که با انتخاب محل‌هایی به‌عنوان هدف برای حفاری به‌منظور دستیابی به ذخایر معدنی پایان می‌پذیرد. مرحله اول اکتشافات ذخایر معدنی، پی‌جویی اولیه، معمولاً در مقیاس ۱:۲۵۰,۰۰۰ و ۱:۱۰۰,۰۰۰ انجام می‌شود. به‌طور کلی این فرایند شامل چهار مرحله: ۱. تعیین فاکتورهای تشخیص کانی‌سازی، ۲. تهیه نقشه‌های فاکتور، ۳. تلفیق نقشه‌ها و ۴. تعیین مناطق مناسب برای مطالعات آتی است (کریمی و همکاران، ۱۳۸۷).

مدل‌های متداول تلفیق نقشه که در اکتشاف ذخایر معدنی در مرحله پی‌جویی اولیه به‌کار گرفته شده‌اند، شامل مدل همپوشانی شاخص^۲، عملگرهای فازی^۳، نشانگرهای وزن‌دار^۴ و روش‌های هوشمند مانند جنگل‌های تصادفی^۵ و شبکه‌های مصنوعی^۶ هستند. تعیین مقادیر وزن‌ها و

1- Geospatial Information Systems (GIS)

2- Index Overlay

3- Fuzzy Operators

4- Weight of Evidence

5- Random Forests

6- Neural Network

فصلنامه علمی - پژوهشی اطلاعات جغرافیایی (...)

مقایسه کارایی روش‌های جنگل تصادفی و ماشین بردار پشتیبان به‌منظور پتانسیل‌یابی ذخایر ... / ۵۱

علاوه بر این، یکی از مشکلاتی که در اکثر روش‌های داده‌محور وجود دارد این است که با تعداد کم داده‌های آموزشی نمی‌تواند عملیات مدل‌سازی را انجام دهند و از آنجایی که براساس تحقیقات پیشین، عملکرد مطلوب مدل جنگل تصادفی در دیگر مناطق با تعداد داده آموزشی کم به اثبات رسیده است، بررسی عملکرد این مدل در منطقه مورد مطالعه که تعداد ۲۴ کانسار شناخته شده مس وجود دارد، مورد نیاز است.

با توجه به آنچه گفته شد، هدف تحقیق حاضر، تهیه نقشه پتانسیل ذخایر مس منطقه دهج - بزمان با استفاده از دو روش یادگیری ماشین جنگل‌های تصادفی و ماشین بردار پشتیبان است. علاوه بر این، به‌منظور مقایسه نتایج، از روش‌های دانش‌محور همپوشانی شاخص و منطق فازی نیز استفاده شده است.

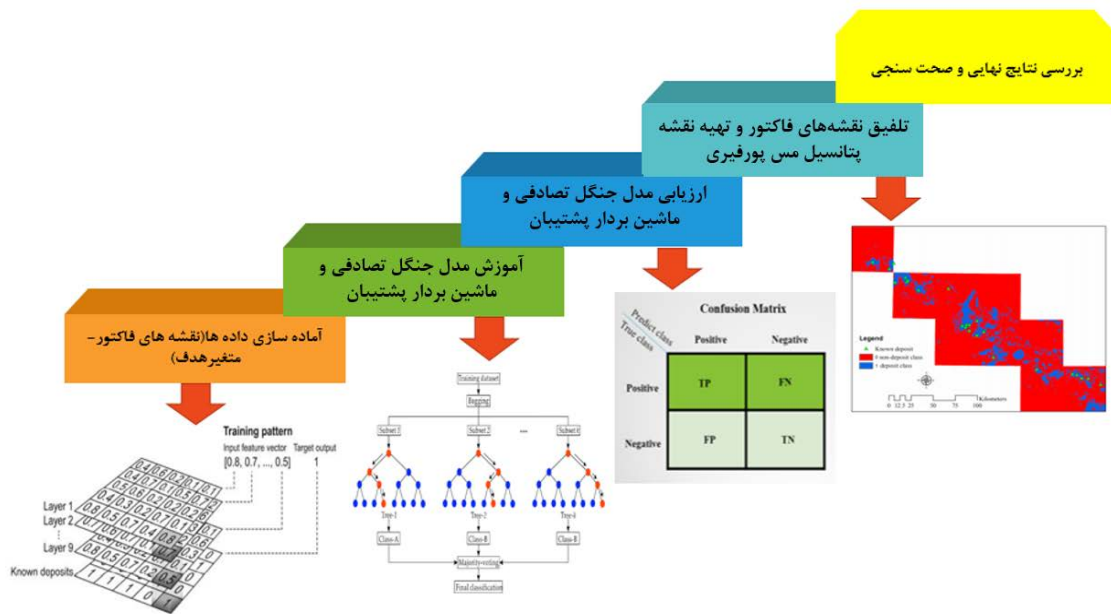
سؤالات کلیدی مطرح در این زمینه شامل تعیین نقشه‌های فاکتور مؤثر به‌عنوان ورودی روش‌های یادگیری ماشین و ماشین بردار پشتیبان، تعیین روش مناسب برای هریک از گروه‌های روش‌های جنگل تصادفی، ماشین بردار پشتیبان و روش‌های کارشناسی و همچنین تعیین اینکه آیا روش‌های یادگیری ماشین جنگل‌های تصادفی و ماشین بردار پشتیبان می‌توانند در فرآیند پتانسیل‌یابی ذخایر معدنی مس در مناطق با تعداد داده آموزشی کم مورد استفاده قرار گیرند، می‌شود.

۲- مواد و روش‌ها

مراحل اصلی تهیه نقشه پتانسیل معدنی شامل تعیین فاکتورهای تشخیص کانی‌سازی، تهیه نقشه‌های فاکتور، تلفیق نقشه‌ها و ارزیابی نتایج است. نگاره ۱ فلوچارت مراحل تهیه نقشه پتانسیل مس پورفیری منطقه دهج - بزمان با مدل جنگل‌های تصادفی و ماشین بردار پشتیبان را نشان می‌دهد. در این تحقیق به‌منظور مدل‌سازی مراحل مختلف فرایند اکتشافات ذخایر معدنی مس در منطقه دهج - بزمان از نرم‌افزار ArcMap و زبان برنامه‌نویسی R استفاده شده است.

مورد استفاده در پتانسیل‌یابی ذخایر معدنی (Li et al., 2022)، استفاده از مدل ماشین بردار پشتیبان برای تهیه نقشه پتانسیل مس پورفیری (Abedi et al., 2012; Mohammadi and Hezarkhani, 2018; Ghezelbash et al., 2022)، استفاده از مدل ماشین بردار پشتیبان برای تهیه نقشه پتانسیل طلا (Geranian et al., 2016; Chen and Wu, 2017)، استفاده از روش جنگل‌های تصادفی برای تهیه نقشه پتانسیل طلا (Rodriguez-Galiano et al., 2014; Carranza and Laborte, 2015; McKay and Harris, 2016) استفاده از روش‌های شبکه‌های عصبی، جنگل‌های تصادفی، رگرسیون و ماشین‌های بردار پشتیبان به‌منظور تهیه نقشه پتانسیل مواد معدنی (Rodriguez-Galiano et al., 2015) تهیه نقشه پتانسیل معدنی طلا-مس با استفاده از روش‌های جنگل‌های تصادفی و گمان شواهد (Carranza and Laborte, 2016) مقایسه کارایی روش‌های جنگل تصادفی، ماشین بردار پشتیبان و شبکه عصبی برای تهیه نقشه پتانسیل معدنی مس (Sun et al., 2019) و مقایسه مدل جنگل تصادفی و ماشین بردار پشتیبان برای پتانسیل‌یابی ذخایر معدنی (Zheng et al., 2023; Lachaud et al., 2023) اشاره کرد. براساس بررسی مطالعات صورت گرفته، روش‌های جنگل‌های تصادفی و ماشین بردار پشتیبان از سایر مدل‌ها عملکرد بهتری دارند. علاوه بر این، نتایج نشان می‌دهند که روش جنگل تصادفی می‌تواند با تعداد کم داده آموزشی در منطقه مورد مطالعه نیز مفید باشد. همچنین، براساس مطالعه انجام شده توسط Carranza و Laborte (2015)، از آنجا که نتایج روش‌های یادگیری ماشین به‌منظور تهیه نقشه پتانسیل ذخایر مس در مناطق مختلف متفاوت است، نیاز است تا به‌منظور تأیید عملکرد مطلوب آن‌ها در دیگر مناطق نیز بررسی شوند.

بنابراین، با توجه به آنکه در ایران و همچنین در منطقه مورد مطالعه از روش‌های هوشمند یادگیری ماشین به‌منظور تهیه نقشه پتانسیل ذخایر مس پورفیری استفاده نشده است، تحلیل و بررسی نتایج روش‌های هوشمند یادگیری ماشین به‌این منظور و مقایسه نتایج با روش‌های کارشناسی ضرورت می‌یابد.



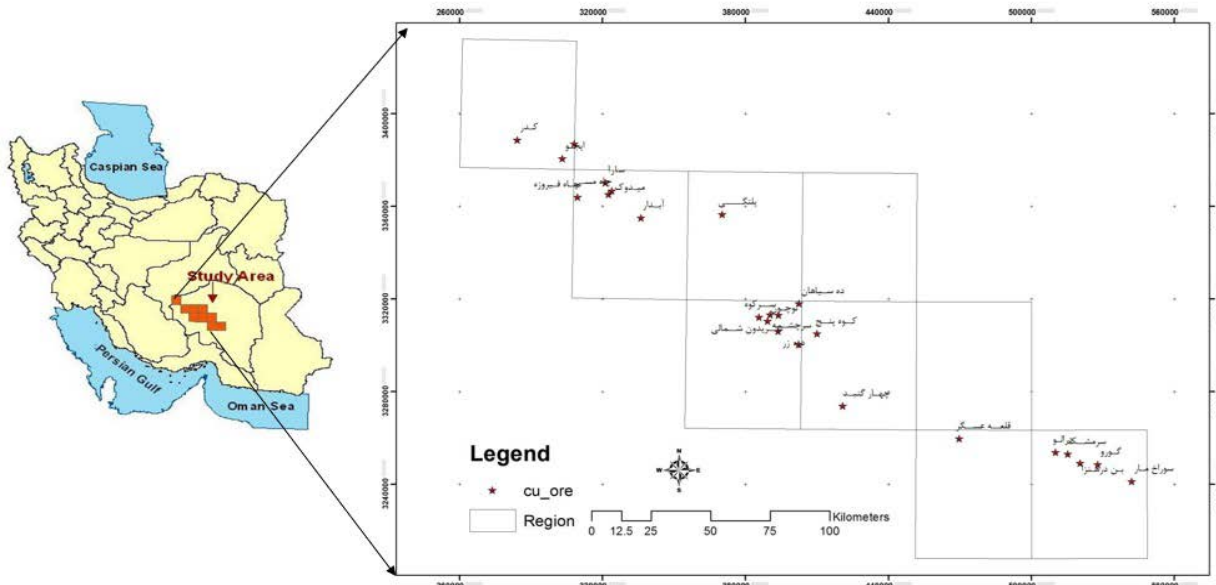
نگاره ۱: فلوجارت مراحل تهیه نقشه پتانسیل مس پورفیری منطقه دهج - بزمان با مدل جنگل های تصادفی و ماشین بردار پشتیبان

۱-۲- منطقه مورد مطالعه

۲-۲- داده های مورد نیاز

منطقه مورد مطالعه در این تحقیق بخشی از کمربند ماگمائی منطقه کرمان، موسوم به کمربند دهج - ساردوئیه، است که در استان کرمان واقع شده است. این ناحیه شامل ۹ ورقه ۱:۱۰۰,۰۰۰ می شود (نگاره ۲).

لایه های اطلاعاتی ورودی این تحقیق شامل واحدهای سنگی، ساختارها، آلتراسیون ها، ژئوشیمی، ژئوفیزیک و کانسارهای مس است که در ادامه هر یک به اختصار تشریح شده اند. لازم به ذکر است که فرض مقاله بر دارا بودن کیفیت



نگاره ۲: موقعیت محدوده مورد مطالعه در کشور

هدف با پتانسیل تشکیل سنگ معدن بالا را نشان می‌دهد. در ادامه، به اختصار به بررسی مدل جنگل‌های تصادفی و ماشین بردار پشتیبان پرداخته شده است.

۲-۳-۱- مدل جنگل‌های تصادفی

این الگوریتم، چندین درخت تصمیم را ترکیب می‌کند و پیش‌بینی‌های مکرر از پدیده را ارائه داده که توسط مجموعه داده‌های آموزشی ایجاد می‌شوند؛ (Rodriguez-Galiano, 2014)

(Rodriguez-Galiano, 2015; Carranza and Laborte, 2015)

جنگل تصادفی از تقریب‌های زیادی برای اندازه‌گیری خلوص اطلاعات از جمله شاخص جینی، نسبت بهره و Chi-square، استفاده می‌کند (Rodriguez-Galiano et al., 2014). متداول‌ترین روش، شاخص جینی است که در این مطالعه برای اندازه‌گیری خلوص گره‌های برگ در مقایسه با گره‌های ریشه آن‌ها استفاده می‌شود، که به صورت رابطه (۱) تعریف می‌شود (Breiman, 2017):

$$I_G(f) = \sum_{i=1}^m f_i(1 - f_i) \quad \text{رابطه (۱)}$$

که f_i احتمال کلاس i در گره m است و به صورت رابطه (۲) محاسبه می‌شود:

$$f_i = \frac{n_j}{n} \quad \text{رابطه (۲)}$$

که n_j تعداد نمونه‌های مربوط به کلاس j است و n تعداد کل نمونه‌ها در یک گره خاص را مشخص می‌کند. طبقه‌بندی نهایی با رأی اکثریت کلیه درختان تصمیم‌گیری در جنگل تعیین می‌شود.

۲-۳-۲- ماشین بردار پشتیبان

در این مطالعه، ماشین بردار پشتیبان غیرخطی به منظور طبقه‌بندی باینری برای شناسایی الگوی «کانسار» و «غیرکانسار» استفاده شد. با توجه به یک مجموعه داده آموزشی $\{x_i\}_{i=1}^n$ با n بردار ویژگی، یک مجموعه داده هدف دارای برچسب $\{y_i\}_{i=1}^n$ با هر بردار x_i همراه است.

قابل قبول داده‌های ورودی است.

• **واحدهای سنگی:** واحدهای سنگی بر اساس نقشه‌های زمین‌شناسی ۱:۱۰۰۰۰۰ تهیه شده توسط سازمان زمین‌شناسی و اکتشافات معدنی کشور، تعیین شده‌اند.

• **ساختارها:** گسل‌های اصلی و فرعی منطقه دهج - بزمان به‌عنوان ساختارهای کانی‌سازی تعیین شدند.

• **آلتراسیون‌ها:** پنج آلتراسیون رسی پیشرفته، رسی، فلیک، ژاروسیت و سیلیس در منطقه دهج - بزمان شناسایی شده‌اند.

• **ژئوشیمی:** داده‌های ژئوشیمی استفاده شده در این تحقیق مربوط به نمونه‌های برداشت شده توسط سازمان زمین‌شناسی و اکتشافات معدنی کشور است. با پردازش آماری تک متغیره و چند متغیره و همچنین محاسبه ضریب غنی‌شدگی، نقشه‌های ژئوشیمیایی عناصر و فاکتورهای موردنظر تهیه شدند.

• **ژئوفیزیک:** داده‌های ژئوفیزیکی استفاده شده در این تحقیق شامل داده‌های ژئونکس آبرودات که در سال ۱۹۹۲ و داده‌های پراکلا هستند که در سال‌های ۱۹۷۷ تا ۱۹۷۹ برداشت شده‌اند. با پردازش داده‌های مذکور، نقشه توده‌های نفوذی پنهان و آنومالی‌های ژئوفیزیکی تهیه و تلفیق شدند.

• **کانسارهای مس:** ذخایر مس پورفیری در دو ناحیه به شرح ذیل تمرکز بیشتری یافته‌اند: (۱) ناحیه اردستان - سرچشمه - بزمان در بخش مرکزی و جنوب شرقی کمربند، (۲) ناحیه معادن مس پورفیری سرچشمه، میدوک، دره زار، چاه فیروزه، دره زرشک، علی‌آباد، نوچون.

۲-۳-۳- الگوریتم‌های یادگیری ماشین

در کاربردهای عملی الگوریتم‌های یادگیری ماشین، تهیه نقشه پتانسیل معدنی در اصل یک موضوع طبقه‌بندی دوحالتی است، به گونه‌ای که هر منطقه کشف نشده از نظر برخی ترکیبات معیارهای تهیه نقشه به‌عنوان آینده‌نگر یا غیرآینده‌نگر طبقه‌بندی می‌شود (Zuo, 2011). نتایج نهایی، مجموعه‌ای از نقشه‌های پیش‌بینی‌کننده است که مناطق

به این صورت می توان داده های آموزشی کافی را به دست آورد و تعداد نمونه های مثبت و منفی را متعادل کرد. در این مطالعه ۲۴ موقعیت کانسار و ۲۴ موقعیت غیر کانسار به عنوان متغیر هدف مشخص شد.

۲- مکان های غیر کانسار می بایست از مکان های کانسار فاصله داشته باشند. در این مطالعه از آنالیز بافر به منظور ایجاد ۹ مجموعه داده با فواصل ۱۵۰، ۲۵۰، ۵۰۰، ۱۰۰۰، ۳۰۰۰، ۷۰۰۰، ۹۰۰۰ و ۱۲۰۰۰ متر از مکان های کانسار استفاده شد تا مشخص شود که تا چه اندازه می توان از مکان های ذخیره به اندازه کافی فاصله داشت. در این خصوص، زمانی که نقاط غیر کانسار در فاصله ۱۲ کیلومتری از مکان های کانسار قرار دارند با احتمال ۸۸ درصد می تواند کانسار جدید کشف شود. بنابراین، در این مطالعه حداقل فاصله بین نقاط کانسار و غیر کانسار ۱۲ کیلومتر در نظر گرفته شد.

۳- برخلاف توزیع فضایی خوشه بندی شده ذخایر معدنی که محصول حوادث نادر و فرآیندهای تشکیل سنگ غیر تصادفی هستند، مکان های غیر کانسار حاصل از فرآیندهای زمین شناسی مشترک، باید به طور تصادفی از لحاظ مکانی توزیع شود.

قبل از اجرای مدل، نقشه های فاکتور باید به نقشه های رستری تبدیل شوند. در این خصوص، اندازه سلول باید طوری باشد که یک ذخیره در هر سلول رخ دهد. در این مطالعه، حد بالایی اندازه سلول ۱۵۰ متر در نظر گرفته شده است. حد پایین سلول بر اساس مقیاس نقشه های فاکتور مکانی در نظر گرفته می شود. بهترین وضوح مکانی خوانا را می توان با رابطه (۶) برآورد کرد (Hengl, 2006):

$$R_f = MS \times 0,00025 \quad \text{رابطه (۶)}$$

که MS مقیاس نقشه را نشان می دهد. در این مطالعه، بزرگ ترین مقیاس نقشه ۱:۱۰۰۰۰۰ است و بنابراین حد پایین سلول اندازه ۲۵ متر در نظر گرفته می شود. در محدوده اندازه مناسب، از اندازه سلول ۱۰۰ متر برای تهیه نقشه های

که $y = 1$ نشان دهنده کانسار و $y = -1$ نشان دهنده وقوع غیر کانسار است. از آنجا که داده های ورودی نمی توانند به طور خطی در فضای ویژگی اصلی از هم جدا شوند، ابتدا توسط یک تابع جستجو Φ به فضایی با ابعاد بالاتر H نگاشت می شوند (Borges, 1998; Asadi and Hale, 2001; Mohammadi and Hezarkhani, 2018):

$$\Phi: R^n \rightarrow H \quad \text{رابطه (۳)}$$

بردار x_i در فضای اصلی را می توان با $\Phi(x_i)$ در فضای H نشان داد. یک هایپرپلان برای جدا کردن دو کلاس، رابطه (۴) را برآورده می کند (Huang et al., 2002):

$$\begin{cases} w\Phi(x_i) + b \geq 1 - \xi_i & \text{for } y_i = 1 \\ w\Phi(x_i) + b \leq -1 + \xi_i & \text{for } y_i = -1 \\ \xi_i \geq 0, \quad i = 1, 2, \dots, n \end{cases} \quad \text{رابطه (۴)}$$

که w بردار وزن نرمال هایپرپلان، b بایاس هایپرپلان و ξ_i متغیر slack مثبت است. لازم به ذکر است که چهار تابع کرنل که معمولاً در SVM استفاده می شود شامل تابع خطی، شعاعی، چند جمله ای و سیگموئید است که تابع پایه شعاعی^۱ به دلیل خطاهای کم و سادگی پارامترها در کاربرد داده های علوم زمین در این تحقیق انتخاب شده است (Zuo, 2011). RBF را می توان به صورت رابطه ۵ تشریح کرد.

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad \gamma > 0 \quad \text{رابطه (۵)}$$

که γ عرض RBF است.

۲-۴- مدل سازی

۲-۴-۱- آموزش مدل

در این تحقیق، متغیرهای هدف برای ورود به مدل جنگل تصادفی و ماشین بردار پشتیبان به صورت نقاط کانسار (با مقادیر ۱) و نقاط غیر کانسار (با مقادیر ۰) تعیین شدند. موقعیت نقاط غیر کانسار با استفاده از سه شرط زیر تعیین شدند (Carranza and Laborte, 2016):

۱- تعداد نقاط غیر کانسار برابر با تعداد نقاط کانسار باشد،

مقدار گاما برابر با ۰/۶۸۱۴ و هزینه ۱/۸۴۳۸ برآورد شد.

۲-۴-۳- ارزیابی مدل

ارزیابی کارایی پیش‌بینی مدل جنگل تصادفی و ماشین بردار پشتیبان می‌تواند به خوبی توسط ماتریس ابهام توصیف شود. در ماتریس ابهام چهار مؤلفه خروجی طبقه‌بندی وجود دارد که به صورت زیر تعریف می‌شوند:

(۱) یک نمونه کانسار که به صورت درست به عنوان کانسار طبقه بندی شده است،

(۲) یک نمونه کانسار که به صورت غلط به عنوان نمونه غیرکانسار طبقه‌بندی شده است،

(۳) یک نمونه غیرکانسار که به صورت درست به عنوان نمونه غیرکانسار طبقه‌بندی شده است، و

(۴) یک نمونه غیرکانسار که به صورت غلط به عنوان نمونه کانسار طبقه‌بندی شده است.

بر این اساس، یک سری از شاخص‌های آماری به شرح روابط (۸ تا ۱۲) برای ارزیابی عملکرد پیش‌بینی مدل‌های آموزش دیده به کار گرفته می‌شود (Liu et al., 2005; Tien Bui et al., 2016)

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad \text{رابطه (۸)}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad \text{رابطه (۹)}$$

$$\text{Positive predictive value} = \frac{TP}{TP + FP} \quad \text{رابطه (۱۰)}$$

$$\text{Negative predictive value} = \frac{TN}{TN + FN} \quad \text{رابطه (۱۱)}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad \text{رابطه (۱۲)}$$

1- TP: True Positive

2- FN: False Negative

3- TN: True Negative

4- FP: False Positive

فاکتور رستری استفاده شد. در مجموع ۴۸ سلول شامل نقاط کانسار و غیرکانسار به عنوان یک مجموعه داده دارای برچسب مورد استفاده قرار گرفتند. سپس این داده‌ها به دو بخش تقسیم شدند. دو سوم داده‌ها به عنوان مجموعه داده آموزشی برای آموزش مدل استفاده شدند، درحالی‌که دیگر داده‌ها به عنوان مجموعه داده تست مورد استفاده قرار گرفتند.

در این مطالعه ۲۰۰۰۰ درخت برای آنالیز جنگل تصادفی انتخاب شدند. علاوه بر این تعداد مطلوب متغیرهای پیش‌بینی‌کننده برای وارد شدن در هر گره توسط تابع «tuneRF» بسته «randomForest» انتخاب شد که تعداد متغیرهای پیش‌بینی‌کننده بهینه با حداقل خطای OOB را ارائه می‌دهد (Micheletti et al., 2014; Carranza and Laborte, 2016).

ارزش بهره اطلاعاتی برای یک نقشه‌ی فاکتور F_i متناسب با کلاس خروجی Y (ذخیره یا غیرذخیره) با رابطه (۷) قابل محاسبه است (Micheletti et al., 2014; Sun et al., 2019):

$$IG(Y, F_i) = H(Y) - H(Y | F_i) \quad \text{رابطه (۷)}$$

که در آن $H(Y)$ ارزش آنتروپی Y است و $H(Y|F_i)$ آنتروپی Y پس از مرتبط کردن مقادیر ویژگی‌های شاهد F_i است.

همچنین، مدل SVM بر مبنای مقدار پارامتر گاما و هزینه ۱۰۰ ساخته شد. (برای جزئیات بیشتر به Tien Bui et al., 2016 مراجعه کنید).

۲-۴-۲- تنظیم پارامترها با الگوریتم ژنتیک

در این مطالعه علاوه بر استفاده از اعداد فوق برای تنظیم پارامترها از الگوریتم ژنتیک و اعتبارسنجی متقابل K-fold به منظور تعیین بهینه تعداد درختان در مدل جنگل تصادفی و پارامترهای گاما و هزینه در ماشین بردار پشتیبان به کار برده شده است. برای جنگل‌های تصادفی زمانی که تعداد درختان از عدد ۱ تا ۱۰۰ در برنامه تنظیم شده بود مقدار بهینه ۵۶ و زمانی که تعداد درختان بین ۱ و ۱۰۰۰ تنظیم شد مقدار بهینه ۴۷۱ برآورد شد. برای ماشین بردار پشتیبان

۲-۵- تلفیق نقشه‌های فاکتور با روش‌های یادگیری ماشین و دانش‌محور

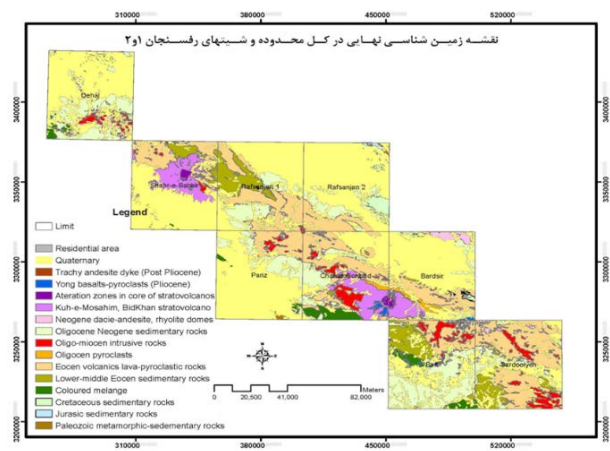
بعد از آموزش و ارزیابی مدل‌های مختلف، بهترین مدل با تنظیم پارامترهای مختلف به دست آمد و از آن برای تلفیق نقشه‌های فاکتور به منظور پیش‌بینی مناطق پتانسیل ذخایر مس پورفیری استفاده شد. همچنین از روش‌های دانش‌محور منطق فازی و همپوشانی شاخص برای تلفیق نقشه‌های فاکتور استفاده شد تا با نتایج روش‌های هوشمند مقایسه شوند.

به منظور نیل به این مهم، در مرحله اول نیاز است تا وزن نقشه‌های فاکتور تعیین شود. به این منظور با توجه به نظرات و قضاوت کارشناسان در این حوزه اوزان نقشه‌های فاکتور آلتراسیون، ژئوشیمی، زمین‌شناسی، ژئوفیزیک و ساختارها به ترتیب ۰/۳، ۰/۲۵، ۰/۲۵، ۰/۱، ۰/۱، ۰/۱ اختصاص یافت. سپس از آنالیزهای Weighted Sum و Fuzzy Overlay به منظور تلفیق لایه‌ها و تولید نقشه پتانسیل استفاده شد.

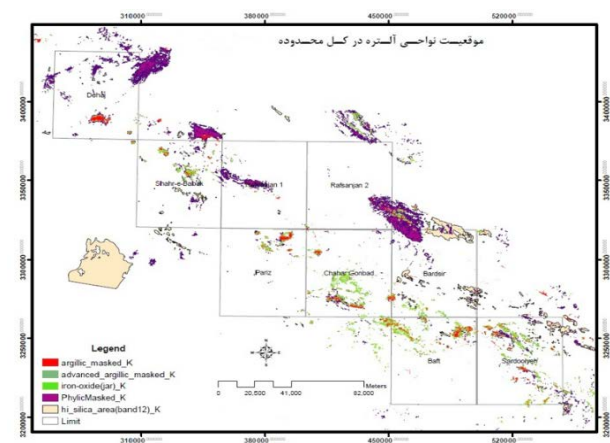
برای مقایسه کارایی الگوریتم‌های یادگیری ماشین و روش‌های دانش‌محور نیاز است تا نقشه‌های خروجی به دست آمده از هر کدام از روش‌ها به صورت باینری باشند. برای این منظور از آنالیز Reclaify به منظور طبقه‌بندی نقشه‌ها استفاده شد. به طور مثال برای تولید نقشه با روش همپوشانی شاخص نقشه پتانسیل به ۹ کلاس طبقه‌بندی شد. ۷ کلاس به عنوان مناطق مس پورفیری و ۲ کلاس به عنوان مناطق غیر مس طبقه‌بندی شدند.

۳- نتایج و بحث

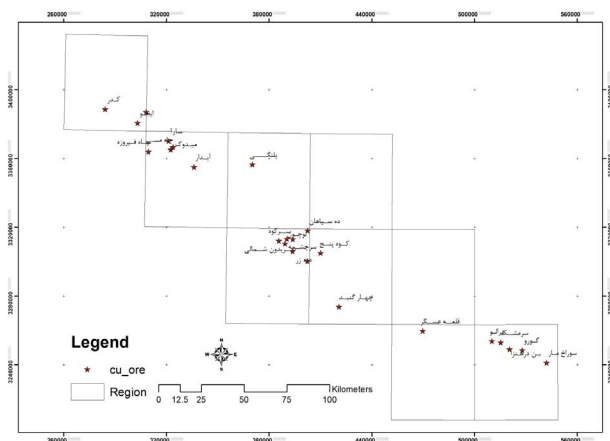
لایه‌های اطلاعاتی کنترل‌کننده کانی‌سازی در منطقه دهج - بزمان شامل واحدهای سنگی، ساختارها، آلتراسیون‌ها، ژئوشیمی، ژئوفیزیک و کانسارهای مس هستند که نحوه توزیع مکانی برخی از آن‌ها در نگاره ۳ ارائه شده است.



(الف)



(ب)



(ج)

نگاره ۳: الف) نقشه جنس واحدهای سنگی موجود در منطقه دهج - بزمان، ب) نقشه انواع آلتراسیون‌های موجود در منطقه دهج - بزمان، ج) نقشه موقعیت کانسارهای موجود در منطقه دهج - بزمان

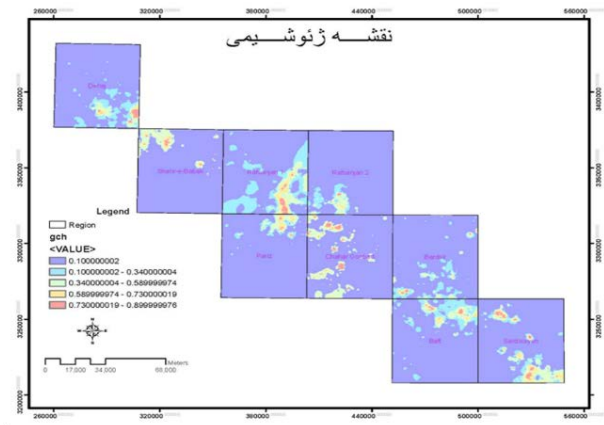
فصلنامه علمی - پژوهشی اطلاعات جغرافیایی (...)

مقایسه کارایی روش‌های جنگل تصادفی و ماشین بردار پشتیبان به منظور پتانسیل یابی ذخایر ... / ۵۷

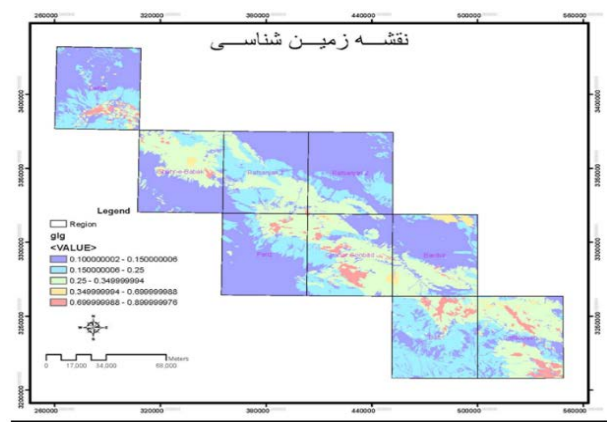
۳-۱- آماده‌سازی داده‌ها و نقشه‌های فاکتور

در این تحقیق، در تهیه نقشه فاکتور واحدهای سنگی جنس واحدهای سنگی کلیه ورقه‌ها با توجه به اهمیت هر جنس برای اکتشاف ذخایر مس یکسان‌سازی، کلاسه‌بندی و وزن‌دهی شد. در تهیه نقشه فاکتور گسل‌ها چهار بافر با فواصل ۴۰۰، ۸۰۰، ۱۴۰۰ و ۲۰۰۰ متر بر روی لایه گسل‌ها ایجاد و به ترتیب وزن‌های ۰/۹، ۰/۷، ۰/۵، ۰/۳ و ۰/۱ به آن‌ها منتسب شد. در تهیه نقشه فاکتور آلتراسیون، ابتدا به آلتراسیون‌های رسی پیشرفته، رسی، فیلیک، ژاروسیت و سیلیس به ترتیب وزن‌های ۶، ۶، ۲، ۲ و ۱ منتسب شد. سپس انواع لایه‌های آلتراسیون تلفیق و سپس نرمال‌سازی شدند. در تهیه نقشه فاکتور ژئوشیمی، ابتدا برای هر ورقه نقشه‌های

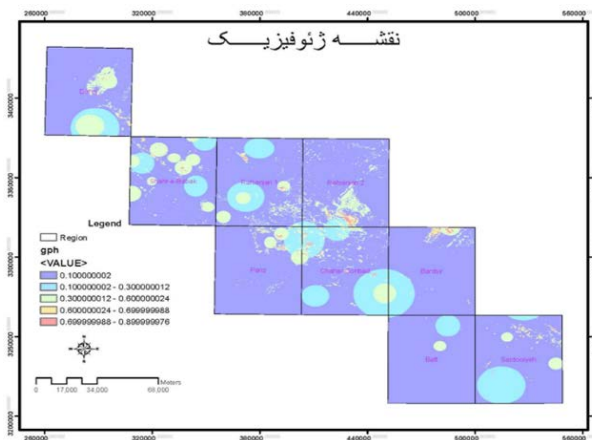
آنومالی عناصر مختلف از جمله مس، مولیبدن، طلا و نقره و همچنین نقشه‌های چند متغیره تهیه و کلاسه‌بندی شد. با تحلیل نقشه‌های تهیه شده، نقشه فاکتور ژئوشیمی هر ورقه از تلفیق نقشه‌های مناسب و با استفاده از روش میانگین وزن‌دار تهیه و سپس نقشه ورقه‌های مختلف تلفیق شد. در نهایت، در تهیه نقشه فاکتور ژئوفیزیک، ابتدا نقشه آنومالی ژئوفیزیک با سه کلاس به وزن‌های ۰/۹، ۰/۶ و ۰/۳ و نقشه توده‌های نفوذی با دو کلاس به وزن‌های ۰/۶ و ۰/۵ تهیه شدند. نقشه فاکتور ژئوفیزیک از تلفیق دو نقشه مذکور با توجه به جمع جبری وزن‌ها و نرمال‌سازی آن‌ها حاصل شد. نگاره ۴، برخی از نقشه‌های فاکتور تهیه شده را نشان می‌دهد.



(ب)



(الف)



(ج)

نگاره ۴: الف) نقشه فاکتور واحدهای سنگی مرتبط با کانی‌سازی، ب) نقشه فاکتور ژئوشیمی، ج) نقشه فاکتور ژئوفیزیک

جدول ۱: معیارهای آماری کیفیت مدل جنگل‌های تصادفی و ماشین بردار پشتیبان

شاخص آماری کیفیت	مدل جنگل‌های تصادفی	مدل ماشین بردار پشتیبان
Sensitivity	۱۰۰ درصد	۱۰۰ درصد
Specificity	۸۸ درصد	۷۵ درصد
POSITIVE PREDICTIVE VALUE	۸۸ درصد	۷۸ درصد
NEGATIVE PREDICTIVE VALUE	۱۰۰ درصد	۱۰۰ درصد
Accuracy	۹۳/۳۳ درصد	۸۶/۶۷ درصد
Kappa	۸۷ درصد	۷۴ درصد
OOB error	۶/۶۷ درصد	۱۳/۳۳ درصد

واقعی هستند. مقدار شاخص کاپا مدل جنگل‌های تصادفی و ماشین بردار پشتیبان به ترتیب ۸۷ و ۷۴ درصد هستند که نشان‌دهنده قابلیت اطمینان به مدل است.

۳-۲-۲- اهمیت نسبی نقشه‌های فاکتور

اهمیت متغیرهای پیش‌بینی‌کننده در مدل جنگل‌های تصادفی می‌تواند از طریق میانگین کاهش صحت^۱ و میانگین کاهش ناخالصی گره یا شاخص ناخالصی جینی ارزیابی شود (Breiman, 2001). معیار اول در هنگام محاسبه خطای OOB تعیین می‌شود، در حالی که معیار دوم نحوه عملکرد هر نقشه فاکتور برای همگنی گره‌ها و مدل نهایی جنگل‌های تصادفی است. هرچه میانگین صحت و شاخص ناخالصی جینی یک نقشه فاکتور بیشتر باشد، نقشه فاکتور از اهمیت بیشتری برخوردار خواهد بود (Breiman, 2001). با اندازه‌گیری دو معیار فوق، مهم‌ترین پیش‌بینی‌کننده در مدل جنگل‌های تصادفی، نقشه ژئوشیمی است، در حالی که فاکتور ساختارها کمترین تأثیر در پیش‌بینی تهیه نقشه پتانسیل معدنی با مدل جنگل‌های تصادفی را دارد.

۳-۲-۳- تجزیه و تحلیل حساسیت پارامترها

در این تحقیق تجزیه و تحلیل حساسیت پارامترها در سه بخش تغییر اندازه مساحت معدن مس، تغییر اندازه سلول نقشه‌های فاکتور و تغییر تعداد mtry در مدل جنگل‌های

۳-۲- نتایج روش های ارزیابی شده

۳-۲-۱- ارزیابی کارایی مدل جنگل‌های تصادفی و ماشین بردار پشتیبان

برای ارزیابی عملکرد مدل جنگل‌های تصادفی و ماشین بردار پشتیبان معیارهای صحت، حساسیت، ویژگی، ارزش مثبت پیش‌بینی شده، ارزش منفی پیش‌بینی شده، شاخص کاپا و خطای OOB مورد استفاده قرار گرفتند. معیارهای آماری کیفیت مدل جنگل‌های تصادفی و ماشین بردار پشتیبان در جدول (۱) ارائه شده‌اند. براساس این جدول، مقدار صحت و خطای OOB مدل جنگل‌های تصادفی به ترتیب ۹۳/۳۳ درصد و ۶/۶۷ درصد است که نرخ طبقه‌بندی صحیح هر دو مکان کانسار و غیرکانسار است که نشان می‌دهد مدل به نوعی کامل است. مقدار حساسیت مدل جنگل‌های تصادفی و مدل ماشین بردار پشتیبان به ۱۰۰ درصد می‌رسد، که نشان می‌دهد ۱۰۰ درصد سلول‌های کانسار به درستی در کلاس کانسار طبقه‌بندی شده‌اند. مقدار ویژگی مدل جنگل‌های تصادفی به ۸۸ درصد می‌رسد که نشان می‌دهد ۸۸ درصد سلول‌های غیر کانسار به درستی در کلاس غیر کانسار طبقه‌بندی شده‌اند. مقدار ارزش مثبت پیش‌بینی شده مدل جنگل تصادفی به ۸۸ درصد می‌رسد که نشان می‌دهد ۸۸ درصد سلول‌های غیرکانسار پیش‌بینی شده مکان غیرکانسار واقعی هستند. مقدار ارزش منفی پیش‌بینی شده مدل جنگل‌های تصادفی به ۱۰۰ درصد می‌رسد که نشان می‌دهد ۱۰۰ درصد سلول‌های کانسار پیش‌بینی شده مکان کانسار

1- mean decrease in accuracy

فصلنامه علمی - پژوهشی اطلاعات جغرافیایی (۳۳)

مقایسه کارایی روش‌های جنگل تصادفی و ماشین بردار پشتیبان به منظور پتانسیل‌یابی ذخایر ... / ۵۹

تصادفی انجام شده است که در ادامه ارائه شده‌اند. گرفته شده است وجود دارد یا خیر.

• تغییر اندازه مساحت معدن مس: در منطقه مطالعاتی دهج-بزمان ۲۴ مکان معدن وجود دارد که در این تحلیل مساحت و تعداد معادن با پیکسل‌هایی که در همسایگی ۲۴ معدن فوق وجود دارد، زیاد شد. به این ترتیب در پنج مرحله و با مساحت‌های معدن ۱، ۹، ۲۵، ۴۹ و ۸۱ هکتار صحت مدل جنگل‌های تصادفی و نقشه‌ی پتانسیل مس پورفیری منطقه مورد ارزیابی قرار گرفت. برای ارزیابی صحت مدل جنگل‌های تصادفی در ۵۰ تکرار مجموعه داده‌های آموزش و تست تغییر داده شد و در فرآیند آموزش و ارزیابی مدل قرار گرفت. سپس صحت میانگین برای ۱۰ مجموعه داده محاسبه و در نهایت از میانگین صحت مدل به‌عنوان صحت نهایی مدل جنگل‌های تصادفی استفاده شد (جدول ۳). با توجه به نتایج حاصل شده، فاکتورهای مؤثر با اندازه پیکسل‌های مربعی ۱۰۰ منجر به ایجاد بیشترین صحت مدل شد. برای ارزیابی صحت نقشه‌ی پتانسیل مس پورفیری بعد از اینکه برای هریک از مساحت‌های فوق نقشه پتانسیل به‌دست آمد، بررسی شد که آیا معادن مس داخل محدوده‌ای که به‌عنوان معدن در نظر گرفته شده است وجود دارد یا خیر.

• تغییر اندازه مساحت معدن مس: در منطقه مطالعاتی دهج-بزمان ۲۴ مکان معدن وجود دارد که در این تحلیل مساحت و تعداد معادن با پیکسل‌هایی که در همسایگی ۲۴ معدن فوق وجود دارد، زیاد شد. به این ترتیب در پنج مرحله و با مساحت‌های معدن ۱، ۹، ۲۵، ۴۹ و ۸۱ هکتار صحت مدل جنگل‌های تصادفی و نقشه‌ی پتانسیل مس پورفیری منطقه مورد ارزیابی قرار گرفت. برای ارزیابی صحت مدل جنگل‌های تصادفی در ۵۰ تکرار مجموعه داده‌های آموزش و تست تغییر داده شد و در فرآیند آموزش و ارزیابی مدل قرار گرفت. سپس صحت میانگین برای ۱۰ مجموعه داده محاسبه و در نهایت از میانگین صحت مدل به‌عنوان صحت نهایی مدل جنگل‌های تصادفی استفاده شد (جدول ۲). با توجه به نتایج حاصل شده، مساحت معدن ۹ هکتار منجر به ایجاد بیشترین صحت مدل شد. برای ارزیابی صحت نقشه‌ی پتانسیل مس پورفیری بعد از اینکه برای هریک از مساحت‌های فوق نقشه پتانسیل به‌دست آمد، بررسی شد که آیا معادن مس داخل محدوده‌ای که به‌عنوان معدن در نظر

جدول ۲: اطلاعات استخراج شده از نقشه پتانسیل مس پورفیری برای تحلیل تغییر اندازه مساحت معدن مس

مساحت معدن	تعداد معدن	صحت مدل	تعداد کل پیکسل‌ها	تعداد پیکسل‌های معدن	تعداد پیکسل‌های غیرمعدن	مساحت معدن به کل منطقه	ذخایر موجود در مناطق پیشنهادی
۱ هکتار	۲۴	۹۳/۳۳	۲۴۰۳۹۴۸	۳۴۰۰۰۶	۲۰۶۳۹۴۲	۱۴ درصد	۹۲ درصد
۹ هکتار	۲۱۶	۹۵/۶۷	۲۴۰۳۹۴۸	۵۷۲۲۰۵	۱۸۳۱۷۴۳	۲۴ درصد	۱۰۰ درصد
۲۴ هکتار	۶۰۰	۹۴/۱۷	۲۴۰۳۹۴۸	۵۷۲۲۱۵	۱۸۳۱۷۳۳	۲۴ درصد	۹۶ درصد
۴۹ هکتار	۱۱۷۶	۹۳/۱۸	۲۴۰۳۹۴۸	۵۷۲۱۷۰	۱۸۳۱۷۷۸	۲۴ درصد	۹۶ درصد
۸۱ هکتار	۱۹۴۴	۹۲/۴۶	۲۴۰۳۹۴۸	۵۷۲۱۸۶	۱۸۳۱۷۶۲	۲۴ درصد	۹۶ درصد

جدول ۳: اطلاعات استخراج شده از نقشه پتانسیل مس پورفیری برای تحلیل تغییر اندازه سلول نقشه‌های فاکتور

مساحت معدن	تعداد معدن	صحت مدل	تعداد کل پیکسل‌ها	تعداد پیکسل‌های معدن	تعداد پیکسل‌های غیرمعدن	مساحت معدن به کل منطقه	ذخایر موجود در مناطق پیشنهادی
۱۰۰*۱۰۰	۲۴	۹۳/۳۳	۲۴۰۳۹۴۸	۳۴۰۰۰۶	۲۰۶۳۹۴۲	۱۴ درصد	۹۲ درصد
۳۰۰*۳۰۰	۲۴	۸۷/۳۳	۲۶۷۵۷۸	۶۷۰۵۹	۲۰۰۵۱۹	۲۵ درصد	۱۰۰ درصد
۵۰۰*۵۰۰	۲۴	۸۶/۶۷	۹۶۱۵۸	۱۰۴۱۶	۸۵۷۴۲	۱۱ درصد	۸۸ درصد
۷۰۰*۷۰۰	۲۴	۸۹/۳۳	۴۹۱۲۳	۹۶۲۷	۳۹۴۹۶	۲۰ درصد	۹۶ درصد
۹۰۰*۹۰۰	۲۴	۸۰/۶۷	۲۹۷۴۰	۴۰۹۵	۲۵۶۴۵	۱۴ درصد	۸۰ درصد

۳-۲-۴- مقایسه روش‌های یادگیری ماشین و روش‌های دانش‌محور

برای مقایسه روش‌های یادگیری ماشین و روش‌های دانش‌محور، اطلاعات جدول ۴ مطابق با نقشه‌های موجود استخراج شدند. در نقشه پتانسیل به‌دست آمده از مدل جنگل‌های تصادفی (نگاره ۶-الف) مناطق هدف ۱۴ درصد از منطقه مورد مطالعه را در بر گرفته، که در آن ۹۲ درصد ذخایر شناخته شده مشخص شده‌اند.

در نقشه پتانسیل مس پورفیری به‌دست آمده از مدل ماشین بردار پشتیبان (نگاره ۶-ب) مناطق هدف ۱۴ درصد از منطقه مورد مطالعه را در بر گرفته، که در آن ۸۷ درصد ذخایر شناخته شده مشخص شده‌اند. به‌منظور تولید نقشه پتانسیل مس پورفیری با روش‌های دانش‌محور، قضاوت کارشناسان خبره برای اختصاص اوزان به هر نقشه معیار به کار برده شد. به‌این منظور برای تولید نقشه‌های فاکتور آلتراسیون، ژئوشیمی، زمین‌شناسی، ژئوفیزیک و ساختارها به‌ترتیب اوزان ۰/۳، ۰/۲۵، ۰/۲۵، ۰/۱، ۰/۱، ۰/۱ اختصاص یافت.

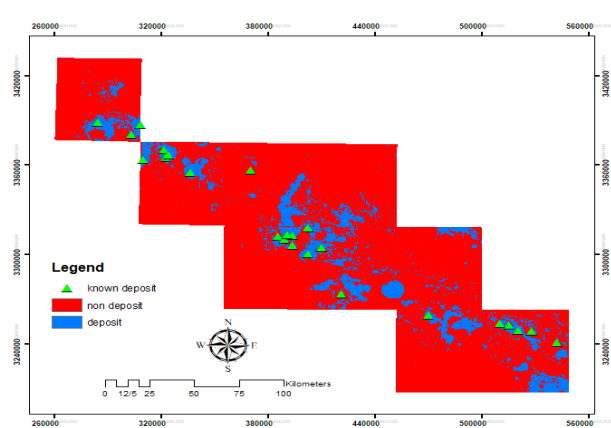
• تغییر تعداد mtry در مدل جنگل‌های تصادفی: برای آموزش مدل جنگل‌های تصادفی نیاز به تنظیم دو پارامتر است که یکی از این پارامترها تعداد ویژگی‌های شرکت‌کننده در هر درخت از جنگل‌های تصادفی (که در این تحقیق نقشه‌های فاکتور ژئوشیمی، ژئوفیزیک، آلتراسیون، ساختارها و جنس واحدهای سنگی هستند) یا همان mtry است. برای مجموعه داده‌های موجود در تحلیل تغییر اندازه مساحت معدن مس و با تنظیم مقدار mtry از اعداد ۱ تا ۵ صحت مدل جنگل‌های تصادفی با ۲۰ مجموعه داده آموزش و تست متفاوت مورد ارزیابی قرار گرفت تا مشخص شود با کدام تعداد mtry بهترین صحت حاصل می‌شود (نگاره ۵).



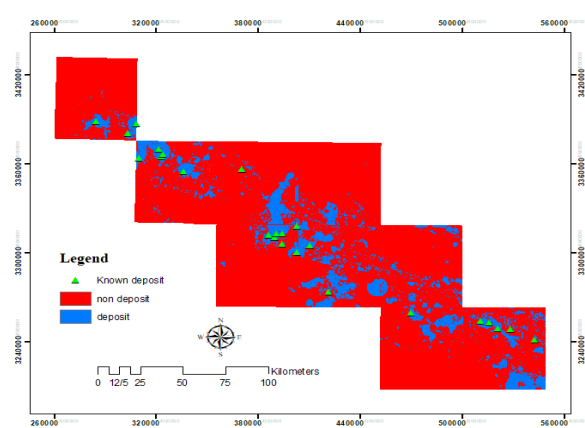
نگاره ۵: نتایج حاصل از تغییر mtry برای معادن ۱، ۹، ۲۵، ۴۹ و ۸۱ هکتاری

جدول ۴: مقایسه کارایی روش‌های یادگیری ماشین و روش‌های دانش‌محور

روش	تعداد کل پیکسل‌ها	تعداد پیکسل‌های معادن	تعداد پیکسل‌های غیرمعادن	مساحت معادن به کل منطقه	ذخایر موجود در مناطق پیشنهادی
جنگل‌های تصادفی	۲۴۰۳۹۴۸	۳۴۰۰۰۶	۲۰۶۳۹۴۲	۱۴ درصد	۹۲ درصد
ماشین بردار پشتیبان	۲۴۰۳۹۴۸	۳۳۵۵۲۱	۲۰۶۸۴۲۷	۱۴ درصد	۸۷ درصد
همپوشانی شاخص	۲۴۰۰۳۱۹	۳۶۶۱۷۹	۲۰۳۴۱۴۰	۱۶ درصد	۸۳ درصد
جمع فازی	۲۴۰۰۳۱۹	۴۰۷۵۷۶	۱۹۹۲۷۴۳	۱۷ درصد	۷۹ درصد
فازی or	۲۴۰۰۳۱۹	۵۳۱۳۶۱	۱۸۶۸۹۵۸	۲۲ درصد	۷۹ درصد
فازی گاما ۰/۸	۲۴۰۰۳۱۹	۳۱۰۷۸۱	۲۰۸۹۵۳۸	۱۳ درصد	۶۳ درصد
فازی and	۲۴۰۰۳۱۹	۵۱۶۷۶۶	۱۸۸۳۵۵۳	۲۱ درصد	۶۳ درصد
ماشین بردار پشتیبان-الگوریتم ژنتیک	۲۴۰۳۹۴۸	۳۸۵۶۳۰	۲۰۱۸۳۱۸	۱۶ درصد	۸۷ درصد
جنگل تصادفی-الگوریتم ژنتیک	۲۴۰۳۹۴۸	۳۴۴۲۲۴	۲۰۶۰۵۲۴	۱۴ درصد	۹۲ درصد



(ب)



(الف)

نگاره ۶: نقشه پتانسیل مس پورفیری منطقه دهج-بزمان با استفاده از (الف) مدل جنگل تصادفی، (ب) مدل ماشین بردار پشتیبان

۴- نتیجه‌گیری

مس پورفیری منطقه دارد. همچنین، با توجه به نتایج تحلیل حساسیت، در نظر گرفتن مساحت معدن به میزان ۹ هکتار و فاکتورهای مؤثر با اندازه پیکسل‌های مربعی ۱۰۰، بیشترین صحت مدل ایجاد می‌شود.

به‌منظور تولید نقشه پتانسیل مس پورفیری با روش‌های دانش‌محور نیز از قضاوت کارشناسان خبره برای اختصاص اوزان به هر نقشه معیار استفاده شد. با توجه به نتایج حاصل شده، عملکرد مدل جنگل‌های تصادفی بهتر از مدل ماشین بردار و همچنین، عملکرد مدل ماشین بردار پشتیبان بهتر از روش‌های دانش‌محور است. ذکر این نکته قابل توجه است که با توجه به این که در این تحقیق از تعداد نمونه‌های آموزشی کمی (۲۴ معدن مس پورفیری) استفاده شده است، پیشنهاد می‌شود که از این روش در مناطقی با داده‌های بیشتر استفاده شود. همچنین، با توجه به اینکه تهیه داده‌های اکتشافی از لحاظ اقتصادی هزینه‌بر و در بعضی از مناطق مطالعاتی در دسترس نیست، پیشنهاد می‌شود از داده‌های ماهواره‌ای سنجش از دور استفاده و کارایی مدل جنگل‌های تصادفی با استفاده از آن‌ها مورد بررسی قرار گیرد. افزون بر این، پیشنهاد می‌شود در تحقیقات آینده، معنی‌دار بودن اختلافات بین روش‌های مورد استفاده از نظر آماری مورد بررسی و تحلیل قرار گیرد.

در این تحقیق ارزیابی و مقایسه کارایی روش جنگل‌های تصادفی و روش ماشین بردار پشتیبان و روش‌های دانش‌محور برای تهیه نقشه پتانسیل مس پورفیری منطقه دهج-بزمان استان کرمان انجام شد. براساس نتایج، مدل جنگل‌های تصادفی در زمینه تهیه نقشه پتانسیل مس پورفیری به خوبی عمل می‌کند. علاوه بر این، الگوریتم جنگل‌های تصادفی می‌تواند اهمیت نقشه‌های فاکتور را تخمین بزند. نتایج این تحقیق نشان می‌دهد که نقشه فاکتور ژئوشیمی بالاترین اهمیت و نقشه فاکتور ساختارها کمترین اهمیت را در پیش‌بینی مدل داده‌محور جنگل‌های تصادفی دارا هستند. این برآورد اهمیت با دانش زمین‌شناسی در مورد کانی‌سازی مس پورفیری منطقه دهج-بزمان سازگار است. همچنین، معیارهای آماری مورد استفاده برای کیفیت نقشه نشان می‌دهند که روش جنگل‌های تصادفی عملکرد مناسبی در تهیه نقشه پتانسیل مس پورفیری منطقه دارد.

برای تهیه نقشه پتانسیل مس پورفیری با استفاده از روش مدل ماشین بردار پشتیبان از کرنل RBF به‌منظور انتقال داده‌های آموزش به یک فضای n بعدی استفاده شد. معیارهای آماری مورد استفاده برای کیفیت نقشه نشان می‌دهد که روش ماشین بردار پشتیبان بعد از روش جنگل‌های تصادفی، عملکرد مناسبی در تهیه نقشه پتانسیل

Engineering and Remote Sensing, 54, 1585–1592.

11- Bonham-Carter, G. F., Agterberg, F. P., & Wright, D. F., 1989. Weights of evidence modelling: A new approach to mapping mineral potential. In F. P. Agterberg & G.F. Bonham-Carter (Eds.), Statistical applications in the earth sciences. Geological Survey of Canada, Paper 89-9 (pp. 171–183).

12- Bonham-Carter, Graeme F., “ Geographical Information Systems for geoscientists: modeling with GIS”, Pergamon Press, Ontario, Canada, Chapter9, 1994.

13- Breiman, L., 2001. Random forests. Machine learning, 45, 5-32.

14- Breiman, L., 2017. Classification and regression trees. Routledge.

15- Brown, W. M., Gedeon, T. D., Groves, D. I., & Barnes, R. G., 2000. Artificial neural networks: A new method for mineral prospectivity mapping. Australian Journal of Earth Sciences, 47, 757–770.

16- Burges, C.J.C., 1998. A tutorial on support vector machines for pattern recognition, Data Min. Knowl. Discov., pp. 121–167.

17- Carranza, E. J. M., van Ruitenbeek, F. J. A., Hecker, C., van der Meijde, M., van der Meer, F. D., 2008a. Knowledge-guided data-driven evidential belief modeling of mineral prospectivity in Cabo de Gata, SE Spain. Int. J. Appl. Earth Obs. 10, 374–387.

18- Carranza, E. J. M., & Hale, M., 2000. Geologically constrained probabilistic mapping of gold potential, Baguio District, Philippines. Natural Resources Research, 9, 237–253.

19- Carranza, E. J. M., Woldai, T., & Chikambwe, E. M., 2005. Application of data-driven evidential belief functions to prospectivity mapping for aquamarine-bearing pegmatites, Lundazi District, Zambia. Natural Resources Research, 14, 47–63.

20- Carranza, E. J. M., & Hale, M. (2003). Evidential belief functions for data-driven geologically constrained mapping of gold potential, Baguio district, Philippines. Ore Geology Reviews, 22, 117–132.

21- Carranza, E. J. M., Laborte, A. G., 2015. “Data-driven predictive mapping of gold prospectivity, Baguio district, Philippines: Application of Random Forests algorithm,”

منابع و مأخذ

۱- درویش‌زاده، علی، ۱۳۷۰، کتاب زمین‌شناسی ایران، انتشارات امیرکبیر، ۱۳۷۰

۲- علائی‌مقدم، کریمی، مسگری، صاحب‌الزمانی؛ ساناز، محمد، محمدسعیدی، نادر؛ مدل‌سازی فرایند تهیه نقشه پتانسیل معدنی با استفاده از سیستم استنتاج‌گر فازی (منطقه مورد مطالعه: کانسار مس چاه فیروزه)، نشریه علوم زمین (سازمان زمین‌شناسی و اکتشاف معدنی کشور)، پاییز ۱۳۹۳، دوره ۲۴، شماره ۹۳- صفحه ۵۳-۶۶.

۳- کریمی، ولدان‌زوج، عبادی، صاحب‌الزمانی؛ محمد، محمد جواد، حمید، نادر؛ تهیه نقشه پتانسیل دخیار مس با استفاده از سامانه اطلاعات جغرافیایی (GIS)، فصلنامه علوم زمین، شماره ۶۸، ۱۳۸۷.

۴- مؤمنی؛ سمانه؛ شناسایی مناطق امیدبخش از نظر وجود منابع فلزی با استفاده از تلفیق و تحلیل داده‌های ماهواره‌ای و فتوگرامتری، پایان‌نامه کارشناسی ارشد، ۱۳۹۲

5- Abedi, M., Norouzi, G.H., Bahroudi, A., 2012. Support vector machine for multi-classification of mineral prospectivity areas, Comput. Geosci., vol. 46, pp. 272–283, doi: 10.1016/j.cageo.2011.12.014.

6- Agterberg, F. P., 1992. Combining indicator patterns in weights of evidence modeling for resource evaluation. Nonrenewable Resources, 1, 39–50.

7- Alaei Moghadam, S., Karimi, M., Sadi Mesgari, M. Application of a fuzzy inference system to mapping prospectivity for the Chahfiroozeh copper deposit, Kerman, Iran, Journal of Spatial Science, 2015. 60(2), pp.233-255 DOI:10.1080/14498596.2015.1004653.

8- Asadi, H.H., Hale, M., 2001. A predictive GIS model for mapping potential gold and base metal mineralization in Takab area, Iran. Comput. Geosci. 27, 901-912.

9- Bonham-Carter, G. F., 1994, Geographic Information Systems for Geoscientist: Modeling With GIS, Pergamon Press, Ontario, Canada.

10- Bonham-Carter, G. F., Agterberg, F. P., & Wright, D. F., 1988. Integration of geological datasets for gold exploration in Nova Scotia. Photogrammetric

- remote sensing, *Geosci. Front.*, vol. 7, no. 1, pp. 3–10, Jan. 2016, doi: 10.1016/j.gsf.2015.07.003.
- 32- Li, Y., Li, J.W., Li, X.H., Selby, D., Huang, G.H., Chen, L.J. and Zheng, K., 2017. An Early Cretaceous carbonate replacement origin for the Xinqiao stratabound massive sulfide deposit, Middle-Lower Yangtze metallogenic belt, China. *Ore Geology Reviews*, 80, 985–1003.
- 33- Li, S., Chen, J. and Liu, C., 2022. Overview on the Development of Intelligent Methods for Mineral Resource Prediction under the Background of Geological Big Data. *Minerals*, 12(5), p.616.
- 34- McKay, G., Harris, J. R., 2016. Comparison of the Data-Driven Random Forests Model and a Knowledge-Driven Method for Mineral Prospectivity Mapping: A Case Study for Gold Deposits Around the Huritz Group and Nueltin Suite, Nunavut, Canada, *Nat. Resour. Res.*, vol. 25, no. 2, pp. 125–143, 2016, doi: 10.1007/s11053-015-9274-z.
- 35- Micheletti, N., Foresti, L., Robert, S., Leuenberger, M., Pedrazzini, A., Jaboyedoff, M. and Kanevski, M., 2014. Machine learning feature selection methods for landslide susceptibility mapping. *Mathematical geosciences*, 46, 33–57.
- 36- Mohammadi, M., Hezarkhani, M., 2018. Application of support vector machine for the separation of mineralised zones in the Takht-e-Gonbad porphyry deposit, SE Iran. *J. Afr. Earth. Sci.* 143, 301–308.
- 37- Porwal, A., Carranza, E. J. M., & Hale, M. 2003a. Artificial neural networks for mineral-potential mapping: A case study from Aravalli Province, Western India. *Natural Resources Research*, 12, 155–171.
- 38- Rigol-Sanchez, J. P., Chica-Olmo, M., & Abarca-Hernandez, F., 2003. Artificial neural networks as a tool for mineral potential mapping with GIS. *International Journal of Remote Sensing*, 24, 1151–1156.
- 39- Rodriguez-Galiano, V.F., Chica-Olmo, M., and M. Chica-Rivas, “Predictive modelling of gold potential with the integration of multisource information based on random forest: a case study on the Rodalquilar area, Southern Spain,” *Int. J. Geogr. Inf. Sci.*, vol. 28, no. 7, pp. 1336–1354, 2014.
- Ore Geol. Rev.*, vol. 71, pp. 777–787.
- 22- Carranza, E. J. M., Laborte, A. G., 2015. “Random forest predictive modeling of mineral prospectivity with small number of prospects and data with missing values in Abra (Philippines),” *Comput. Geosci.*, vol. 74, pp. 60–70, doi: 10.1016/j.cageo.2014.10.004.
- 23- Carranza, E. J. M., Laborte, A. G. 2016. “Data-Driven Predictive Modeling of Mineral Prospectivity Using Random Forests: A Case Study in Catanduanes Island (Philippines),” *Nat. Resour. Res.*, vol. 25, no. 1, pp. 35–50, doi: 10.1007/s11053-015-9268-x.
- 24- Chen, J., Wang, G., & Hou, C, 2005. Quantitative prediction and evaluation of mineral resources based on GIS: A case study in Sanjiang region, southwestern China. *Natural Resources Research*, 15, 285–294.
- 25- Chen, Y., Wu, W., 2017. Mapping mineral prospectivity by using one-class support vector machine to identify multivariate geological anomalies from digital geological survey data, *Australian Journal of Earth Sciences*, 64:5, 639–65.
- 26- Geranian, H., Tabatabaei, S. H., Asadi, H. H., Carranza, E. J. M., 2016. Application of discriminant analysis and support vector machine in mapping gold potential areas for further drilling in the Sari-Gunay gold deposit, NW Iran. *Natural Resources Research*, 25, 145–159.
- 27- Ghezelbash, R., Maghsoudi, A., Shamekhi, M., Pradhan, B. and Daviran, M., 2023. Genetic algorithm to optimize the SVM and K-means algorithms for mapping of mineral prospectivity. *Neural Computing and Applications*, 35(1), pp.719–733.
- 28- Hengl, T., 2006. Finding the right pixel size,” *Comput. Geosci.*, vol. 32, no. 9, pp. 1283–1298.
- 29- Huang, C., Davis, L.S., Townshend, J.R.G., 2002, An assessment of support vector machines for land cover classification, *Int. J. Remote Sens.*, 725–749.
- 30- Lachaud, A., Adam, M. and Mišković, I., 2023. Comparative Study of Random Forest and Support Vector Machine Algorithms in Mineral Prospectivity Mapping with Limited Training Data. *Minerals*, 13(8), p.1073.
- 31- Lary, D. J., Alavi, A. H., Gandomi, A. H., and Walker, A. L. Machine learning in geosciences and

- Xinjiang, China, Earth Sci. Informatics, vol. 11, no. 4, pp. 553–566.
- 46- Zheng, C., Yuan, F., Luo, X., Li, X., Liu, P., Wen, M., Chen, Z. and Albanese, S., 2023. Mineral prospectivity mapping based on Support vector machine and Random Forest algorithm-A case study from Ashele copper-zinc deposit, Xinjiang, NW China. Ore Geology Reviews, p.105567.
- 47- Zuo, R. Machine Learning of Mineralization-Related Geochemical Anomalies: A Review of Potential Methods, Natural Resources Research, vol. 26, no. 4. Springer New York LLC, pp. 457–464, Oct. 01, 2017, doi: 10.1007/s11053-017-9345-4.
- 48- Zuo, R., Carranza, E.J.M., 2011. Support vector machine: A tool for mapping mineral prospectivity. Comput. Geosci. 37, 1967-1975.
- 40- Rodriguez-Galiano, V., , Sanchez-Castillo, M., Chica-Olmo, M., and Chica-Rivas, M. “Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines,” Ore Geol. Rev., vol. 71, pp. 804–818, 2015.
- 41- Saljoughi, B.S., Hezarkhani, A. A., 2018, comparative analysis of artificial neural network (ANN), wavelet neural network (WNN), and support vector machine (SVM) data-driven models to mineral potential mapping for copper mineralizations in the Shahr-e-Babak region, Kerman, Iran. Appl Geomat 10, 229–256. <https://doi.org/10.1007/s12518-018-0229-z>
- 42- Sun, T., Chen, F., Zhong, L., Liu, W., Wang, Y., 2019. “GIS-based mineral prospectivity mapping using machine learning methods: A case study from Tongling ore district, eastern China,” Ore Geol. Rev., vol. 109, pp. 26–49, doi: 10.1016/j.oregeorev.2019.04.003.
- 43- Tien Bui, D., Ho, T.C., Pradhan, B., Pham, B.T., Nhu, V.H. and Revhaug, I., 2016. GIS-based modeling of rainfall-induced landslides using data mining-based functional trees classifier with AdaBoost, Bagging, and MultiBoost ensemble frameworks. Environmental Earth Sciences, 75, pp.1-22.
- 44- Tien Bui, D., Tuan, T.A., Klempe, H., Pradhan, B. and Revhaug, I., 2016. Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. Landslides, 13, 361-378.
- 45- Zhang, N., Zhou, K., Li, D., 2018. Back-propagation neural network and support vector machines for gold mineral prospectivity mapping in the Hatu region,

COPYRIGHTS

©2024 by the authors. Published by National Geographical Organization. This article is an open access article distributed under the terms and conditions of the Creative Commons [Attribution-NoDerivs 3.0 Unported \(CC BY-ND 3.0\)](https://creativecommons.org/licenses/by-nd/3.0/)

