



Semnan University

Journal of Modeling in Engineering

Journal homepage: <https://modelling.semnan.ac.ir/>

ISSN: 2783-2538



Research Article

Assessment of Copper-Gold Mineral Potential in the Shadan Porphyry Area Using SVM and RF Machine Learning Algorithms

Hassan Hosseinzadeh ^{a,*}, Gholamreza Nowrouzi ^a, Arash Gourabjeri ^b, Masoumeh Dadpour ^c

^a Department of Mining Engineering, Birjand University, Birjand, Iran

^b Department of Geology, Faculty of Science, Miyaneh Branch Azad University, Miyaneh, Iran

^c Department of Geology, Faculty of Science, Shahrood University of Technology, Shahrood, Iran

PAPER INFO

Paper history:

Received: 2024-10-12

Revised: 2025-06-07

Accepted: 2025-06-15

Keywords:

Mineral potential map;
machine learning;
SVM algorithm;
RF algorithm;
Shadan.

ABSTRACT

This study applied machine learning algorithms, namely Support Vector Machine (SVM) and Random Forest (RF), to develop a mineral potential map for the Shadan region, situated within the Lut Block and Flysch-Ophiolite Belt of Eastern Iran. The research integrated multiple exploration datasets, including geological, geochemical, satellite imagery, and geomagnetic data, to identify promising areas for mineral exploration, specifically targeting porphyry copper and gold deposits. The performance of the models was evaluated using metrics like Accuracy, Sensitivity, ROC curves, and P-A plots. The SVM model demonstrated superior accuracy, successfully predicting 13% of the study area as high-potential zones for future drilling, which corresponded closely with existing drilling results. These identified target zones were predominantly located in regions with intense tectonic activity and were associated with rock units such as andesite, granite, and granodiorite. The study underscores the effectiveness of the SVM model in accurately delineating mineral-rich areas, providing a valuable basis for future exploration programs.

DOI: <https://doi.org/10.22075/jme.2025.35411.2743>

© 2026 Published by Semnan University Press.

This is an open access article under the CC-BY 4.0 license. (<https://creativecommons.org/licenses/by/4.0/>)

* Corresponding author.

E-mail address: Hosseinzadeh.hassan70@gmail.com

How to cite this article:

H. hosseinzadeh , G. nowrouzi , A. Gourabjeri and M. Dadpour, "Assessment of Copper-Gold Mineral Potential in the Shadan Porphyry Area Using SVM and RF Machine Learning Algorithms," Journal of Modeling in Engineering, 24 84 (2026): 227-251, doi: 10.22075/jme.2025.35411.2743

به کارگیری الگوریتم‌های یادگیری ماشین SVM و RF در تهیه نقشه پتانسیل معدنی محدوده مس - طلای پورفیری شادان، خراسان جنوبی، ایران

حسن حسین زاده^{۱*}، غلامرضا نوروزی^۱، آرش گورابجیری پور^۲، معصومه دادپور^۳

اطلاعات مقاله	چکیده
دریافت مقاله: ۱۴۰۳/۰۷/۲۱	
بازنگری مقاله: ۱۴۰۴/۰۳/۱۷	
پذیرش مقاله: ۱۴۰۴/۰۳/۲۵	
واژگان کلیدی:	
نقشه پتانسیل معدنی،	شناسایی مناطق با پتانسیل معدنی در اکتشافات تفصیلی به منظور طراحی دقیق تر برنامه‌های
یادگیری ماشین،	اکتشافی، از طریق تحلیل و ارزیابی داده‌های موجود انجام می‌گیرد. این فرایند در چارچوب
الگوریتم SVM،	مدل سازی پتانسیل معدنی توسعه داده شده و به بهینه‌سازی عملیات اکتشاف کمک می‌کند. در
الگوریتم RF،	این تحقیق، از الگوریتم‌های یادگیری ماشین نظارت‌شده شامل ماشین بردار پشتیبان (SVM)
شادان.	و جنگل تصادفی (Random Forest) برای تهیه نقشه پتانسیل معدنی محدوده شادان استفاده
	شده است. منطقه شادان که بخشی از بلوک لوت و زون فلیشی - کمر بند افیولیتی شرق ایران
	است. در این پژوهش، با به کارگیری الگوریتم‌های یادگیری ماشین بر روی داده‌های اکتشافی
	منطقه، شامل داده‌های زمین‌شناسی (سنگ‌شناسی و ساختاری)، ژئوشیمی نمونه‌های سنگی،
	تصاویر ماهواره‌ای و ژئومغناطیسی، مناطق هدف برای حفاری‌های بیشتر شناسایی شدند. سپس،
	عملکرد مدل‌های تولیدشده با استفاده از شاخص‌هایی مانند Sensitivity، Accuracy، نمودار
	ROC و نمودار P-A ارزیابی و مقایسه شد. نتایج نشان داد که مدل پتانسیل معدنی حاصل از
	روش SVM عملکرد بهتری در شناسایی اهداف اکتشافی مس و طلای پورفیری داشته است. این
	مدل توانست ۱۳ درصد از مساحت ناحیه مورد بررسی را به عنوان نواحی مطلوب برای حفاری‌های
	آتی که با حفاری‌های انجام‌شده قبلی نیز مطابقت داشت، به درستی پیش‌بینی کند. نواحی
	شناسایی‌شده عمدتاً در مناطقی با تکتونیزه بالا و مرتبط با واحدهای سنگی آندزیت، گرانیت و
	گرانودیوریت قرار دارند. این پژوهش نشان‌دهنده برتری روش SVM در شناسایی نواحی مطلوب
	اکتشافی است و می‌تواند به عنوان مبنایی برای برنامه‌ریزی مراحل بعدی فرایند اکتشاف استفاده
	شود.

DOI: <https://doi.org/10.22075/jme.2025.35411.2743>

© 2026 Published by Semnan University Press.

This is an open access article under the CC-BY 4.0 license. (<https://creativecommons.org/licenses/by/4.0/>)

۱- مقدمه

اندیس‌های معدنی و تبدیل اندیس‌ها به معدن از اهمیت ویژه‌ای برخوردار است. در حالت کلی، ابتدا اکتشاف در یک منطقه وسیع و با استفاده از روش‌های سریع و کم‌هزینه به‌ازای واحد سطح و در مقیاس کوچک شروع شده و پس

شناسایی و اکتشاف صحیح منابع معدنی، باعث پیشگیری از هدر رفتن منابع مالی و انسانی هر کشور محسوب می‌شود، لذا وجود نگرشی سامان‌مند جهت شناسایی

* پست الکترونیک نویسنده مسئول: Hosseinzadeh.hassan70@gmail.com

۱. دانشکده فنی و مهندسی، دانشگاه بیرجند، بیرجند، ایران

۲. دانشکده علوم، دانشگاه آزاد اسلامی واحد میانه، میانه، ایران

۳. دانشکده علوم، دانشگاه صنعتی شاهرود، شاهرود، ایران

استناد به این مقاله:

حسن حسین زاده، غلامرضا نوروزی، آرش گورابجیری پور و معصومه دادپور، "به کارگیری الگوریتم‌های یادگیری ماشین SVM و RF در تهیه نقشه پتانسیل معدنی محدوده مس - طلای پورفیری شادان، خراسان جنوبی، ایران"، مدل سازی در مهندسی، ۲۴ (۱۴۰۵): ۲۲۷-۲۵۱، doi: 10.22075/jme.2025.35411.2743

قطعیت بالایی باشد. از این رو به کارگیری روش‌هایی که توانایی تحلیل و کشف ارتباط پیچیده فضایی را در بین داده‌های اکتشافی اولیه دارا باشند توسعه یافته است. از جمله این روش‌ها می‌توان به خانواده روش‌های مبتنی بر یادگیری ماشین و تشخیص الگو اشاره کرد که به کارگیری آن‌ها بخصوص در سالیان اخیر محبوبیت بسیاری در بین دانشمندان علوم زمین پیدا کرده است [۱۲، ۱۳]. یکی از تعاریف یادگیری ماشین آن‌طور که از سوی تام میشل پروفیسور دانشگاه کارنگی ملون ارائه گردید بدین شرح است: نوعی برنامه کامپیوتری که با توجه به برخی وظایف گروه T و عملکرد P، تجربه E را شکل می‌دهد، اگر عملکرد آن در گروه وظایف T آن‌طور که توسط P اندازه‌گیری شده با تجربه E بهبود پیدا کند [۱۴]. روش‌های یادگیری ماشین به انواع نظارت شده، نظارت نشده، تقویت‌شونده و نیمه نظارتی تقسیم می‌شوند.

روش‌های با ناظر که به‌عنوان روش‌های طبقه‌بندی نیز شناخته می‌شوند، روش‌هایی هستند که در آن‌ها عموماً برچسب‌دهی به پیکسل‌ها بر اساس اندیس‌های معدنی شناخته شده در مقیاس ناحیه‌ای و یا کارهای اکتشافی نظیر گمانه‌ها و یا ترانسه‌های اکتشافی در مقیاس محلی صورت می‌گیرد. مدل بر اساس مجموعه داده‌های برچسب‌دار آموزش می‌بیند و در نهایت برای پیش‌بینی برچسب سایر داده‌های غیر آموزش استفاده می‌شود. در این روش آماده‌سازی داده‌های آموزش نیاز به وجود نواحی شناخته شده اکتشافی دارد درحالی‌که در مراحل اولیه اکتشاف این امر به دلیل عملیات اکتشافی اندک صورت گرفته با محدودیت همراه می‌باشد [۱، ۹، ۱۵].

کارهای اکتشافی صورت گرفته در محدوده شادان و حفاری-های انجام شده نشان می‌دهد تعیین محل‌های حفاری بر اساس اطلاعات اکتشافی موجود و یا تلفیق آن‌ها بر اساس روش‌های کلاسیک صورت گرفته نتوانسته مناطق مستعد کانی‌سازی را به‌خوبی مشخص کند؛ چراکه با بررسی گمانه-های موجود در شبکه حفاری طراحی شده مشاهده می‌شود برخی گمانه‌ها کانی‌سازی مطلوبی نداشته‌اند. اگرچه حفر گمانه‌های اکتشافی مهم‌ترین ابزار اکتشافی به شمار می‌رود، اما هزینه بالا، زمان‌بر بودن، پوشش منطقه کوچک و از همه مهم‌تر امکان عدم برخورد به ماده معدنی از معایب این کار اکتشافی به شمار می‌رود. به کمک روش‌های طبقه‌بندی بر روی داده‌های ژئوشیمی سطحی، ژئومغناطیس زمینی،

از کسب نتیجه در هر مرحله و تعیین مناطقی امیدبخش، عملیات مراحل بعدی که تفصیلی‌تر می‌باشند، بر روی مناطق امیدبخش به‌دست‌آمده از مراحل قبلی متمرکز می‌گردد. مسئله اصلی در این حالت، نحوه انتخاب بهترین مناطق امیدبخش می‌باشد و باید روشی در نظر گرفته شود که بتوان با استفاده از آن و با یک قاعده منطقی و معقول، منطقه مورد مطالعه را کوچک نمود و نواحی امیدبخش را معرفی کرد. در مراحل اولیه اکتشاف و معرفی مناطق امیدبخش، ممکن است از یکی از روش‌های مختلف اکتشافی نظیر مطالعات زمین‌شناسی، سنجش‌ازدور، ژئوشیمی، ژئوفیزیک و یا تلفیق نتایج چند روش استفاده شود. با توجه به اینکه عملیات اکتشافی با قطعیت صورت می‌پذیرد، از مهم‌ترین تصمیم‌ها، نحوه انتخاب مناطق هدف می‌باشد. از مراحل اولیه اکتشاف، مدل‌سازی پتانسیل معدنی به‌منظور معرفی مناسب‌ترین مناطق به‌عنوان اهداف، جهت اکتشافات بعدی می‌باشد [۱-۶].

به‌طور معمول برای مدل‌سازی پتانسیل معدنی، پس از اجرای روش‌های مختلف اکتشافی، از روش‌های داده‌محور باتکیه بر مناطق شناخته شده موجود در منطقه اکتشافی، دانش‌محور بر اساس تجربه کارشناسی، ترکیب دو روش داده‌محور و دانش‌محور و همچنین روش‌های پیوسته بدون استفاده از اندیس‌های شناخته شده و بدون یا با حداقل دخالت قضاوت کارشناسی برای تخصیص وزن به شواهد و الگوهای مختلف و تلفیق داده‌ها استفاده می‌شود. در این راستا استفاده از روش‌های مؤثر تلفیقی در محیط GIS که عدم قطعیت را نیز به‌ویژه در محیط‌های پیچیده زمین‌شناسی بیشتر در مدل‌ها دخالت می‌دهند، تأثیر بسزایی در افزایش دقت اهداف اکتشافی تعیین شده، دارد [۷-۱۰].

در یک مدل پتانسیل معدنی از یک تابع ادغام جهت تعیین ارتباط مجموعه‌ای از اطلاعات موضوعی مختلف شامل اطلاعات زمین‌شناسی، ژئوشیمی، ژئوفیزیک و غیره به‌عنوان ورودی و تیپ خاص کانی‌سازی مورد جستجو به‌عنوان نتیجه خروجی مطلوب استفاده می‌شود. عملکرد تابع ادغام، از عملگرهای منطقی ساده تا توابع ریاضی پیچیده با توجه به نوع مدل پیش‌یاب توسعه یافته است [۱۱]. حجم بالایی داده‌ها در مطالعات اکتشافی و ارتباط پیچیده موجود بین معیارهای اکتشافی و شواهد کانی‌سازی باعث شده است خروجی یک مدل پیش‌یاب پتانسیل معدنی دارای عدم

معدنی استفاده شده است. نتایج نشان داد که این دو الگوریتم در ترکیب با یکدیگر توانستند به دقت بالاتری در شناسایی نواحی معدنی دست یابند. مدل‌های RF و SVM توانستند به دلیل توانایی در مدیریت داده‌های پیچیده ژئوفیزیکی و ژئوشیمیایی، نقشه‌هایی با دقت بالا و توانایی پیش‌بینی دقیق ارائه کنند. این مطالعه نشان داد که روش‌های یادگیری ماشین به طور قابل توجهی کارآمدتر از روش‌های کلاسیک مانند حداکثر احتمال (MaxL) هستند [۱۰]. مقاله‌ای با عنوان "مطالعه مقایسه‌ای الگوریتم‌های جنگل تصادفی (RF) و ماشین بردار پشتیبان (SVM) در تهیه نقشه پتانسیل معدنی با داده‌های محدود آموزشی" توسط لاچود و همکاران در سال ۲۰۲۳ به بررسی اثربخشی دو الگوریتم یادگیری ماشین، یعنی RF و SVM، برای تهیه نقشه پتانسیل معدنی در شرایطی که داده‌های آموزشی محدودی در دسترس است، پرداخته است. مقاله نشان می‌دهد که هر دو روش RF و SVM برای شناسایی الگوها در داده‌های پیچیده ژئوشیمیایی و ژئوفیزیکی کارایی بالایی دارند. RF به دلیل پایداری بیشتر در برابر داده‌های نویزی و عدم نیاز به تنظیمات پیچیده، عملکرد بهتری در برخی شرایط نشان می‌دهد. از سوی دیگر، SVM برای داده‌های با ابعاد بالا مناسب‌تر است، اما ممکن است به حساسیت بیشتری در مقابل ناهماهنگی داده‌ها دچار شود. نتایج این مطالعه بیان می‌کند که حتی با داده‌های آموزشی محدود، هر دو مدل می‌توانند به خوبی عمل کنند، اگرچه کارایی آن‌ها به ویژگی‌های خاص داده‌ها بستگی دارد [۱۶]. ژنگ و همکاران در پژوهشی که در منطقه Rodalquilar اسپانیا اجرا شده است به بررسی استفاده از الگوریتم‌های یادگیری ماشین مانند جنگل تصادفی (Random Forest) و ماشین بردار پشتیبانی (SVM) در مدل‌سازی پتانسیل معدنی می‌پردازد. هدف اصلی مقاله مقایسه عملکرد این الگوریتم‌ها در مدل‌سازی پتانسیل معدنی است و نشان می‌دهد که جنگل تصادفی به دلیل پایداری و دقت بالا در طبقه‌بندی و تحلیل داده‌های ژئوشیمیایی و ژئوفیزیکی بر دیگر الگوریتم‌ها برتری دارد. این تحقیق نشان می‌دهد که این روش‌ها می‌توانند برای مکان‌یابی مناطق معدنی با دقت بالاتری به کار گرفته شوند. همچنین، مقاله بر قابلیت تفسیر پارامترهای مدل‌های جنگل تصادفی و درخت‌های تصمیم تأکید دارد، که می‌تواند به درک بهتر کنترل‌های زمین‌شناسی روی فرآیندهای معدنی کمک کند [۱۷]. در

آلتراسیون‌ها، واحدهای سنگ‌شناسی و گسل‌ها می‌توان با پیش‌بینی موقعیت کانی‌زایی زیرسطحی از معایب آن کاست؛ بنابراین با گمانه‌های حفر شده از قبل و ارتباط دادن آن‌ها با داده‌های سطحی به روش طبقه‌بندی می‌توان علاوه بر تهیه نقشه پتانسیل بهینه، در موقعیت‌یابی گمانه‌های تکمیلی، بهتر از نتایج قبل عمل کرد. تحقیق حاضر به بررسی پتانسیل معدنی طلا و مس در منطقه شادان با استفاده از الگوریتم‌های یادگیری ماشین نظارت شده و تجمیعی می‌پردازد. هدف اصلی تحقیق طراحی نقشه پتانسیل مس - طلا و تعیین نقاط حفاری تکمیلی در محدوده اکتشافی شادان است.

۲- مطالعات مربوطه و پیشینه تحقیق

روش‌های کلاسیک مانند وزن‌دهی شواهد (WofE)، منطق بولین و منطق فازی به دانش کارشناسانه وابسته هستند و از قوانین از پیش تعیین شده برای تحلیل داده‌ها استفاده می‌کنند. این روش‌ها به دلیل تفسیرپذیری بالا و توانایی در ادغام دانش زمین‌شناسی مناسب هستند؛ اما یک محدودیت بزرگ این روش‌ها در مواجهه با داده‌های بزرگ و پیچیده است. آن‌ها معمولاً نمی‌توانند روابط غیرخطی و پیچیده بین داده‌ها را به خوبی مدل‌سازی کنند و در مواجهه با تنوع فضایی و ناپوستگی داده‌ها عملکرد ضعیفی دارند. در مقابل، الگوریتم‌های یادگیری ماشین مانند جنگل تصادفی (RF)، ماشین بردار پشتیبان (SVM) به‌ویژه در دهه اخیر محبوبیت زیادی پیدا کرده‌اند. این الگوریتم‌ها قادر به پردازش داده‌های بزرگ، چندمنبعی (زمین‌شناسی، ژئوفیزیکی و ژئوشیمیایی) و کشف روابط پیچیده و غیرخطی هستند. بر اساس مطالعات، مدل‌های یادگیری ماشین مانند RF و SVM می‌توانند روابط فضایی و غیرخطی بین داده‌ها را بهتر شناسایی کرده و نقشه‌های پتانسیل معدنی دقیق‌تری ارائه دهند. این پیشرفت‌ها در مدل‌سازی پتانسیل معدنی نه تنها توانسته است دقت و سرعت فرایندهای اکتشافی را بهبود بخشد، بلکه روش‌های جدیدی برای تحلیل داده‌های پیچیده ارائه داده است که می‌تواند به طور قابل توجهی بر روند تصمیم‌گیری‌های اکتشافی تأثیر بگذارد. در ادامه تعدادی از مقالات بررسی شده‌اند.

در مطالعه‌ای که در حوضه کوبار استرالیا انجام شده است، از الگوریتم‌های RF و SVM برای تهیه نقشه پتانسیل

مگ کی و هریس (۲۰۱۶) در مقاله‌ای با عنوان "مقایسه نتایج روش داده‌محور جنگل تصادفی و روش دانش‌محور در پتانسیل‌یابی طلا" به بررسی عملکرد جنگل تصادفی در مقایسه با روش‌های دانش‌محور پرداختند. در این تحقیق، به منظور نقشه‌برداری از پتانسیل طلا، دو نقشه جداگانه با استفاده از چندین تکنیک مبتنی بر دانش و نام جنگل تصادفی (RF) تهیه شده است. هدف از این مطالعه مقایسه نتایج تکنیک RF با مدل KD بوده است. داده‌های مورد استفاده شامل مجموعه داده‌های ژئوشیمیایی رسوبات دریاچه، ساختارهای زمین‌شناسی، گسل‌ها، لیتولوژی، ویژگی‌های توپوگرافی و رخدادهای شناخته شده طلا می‌باشد. نتایج نشان داد که روش RF در پیش‌بینی رخدادهای شناخته شده طلا بهتر از روش KD عمل کرده است. همچنین، این روش از مزایایی مانند پذیرش داده‌های پیوسته و دسته‌ای، تخمین بدون طرف خطا (out-of-bag error) بدون نیاز به اعتبارسنجی متقابل و ارزیابی اهمیت هر متغیر ورودی برخوردار است. منحنی‌های کارایی پیش‌بینی نشان می‌دهند که روش RF به‌طور قابل توجهی عملکرد بهتری نسبت به روش KD دارد [۱۹]. در یک مطالعه کاربردی در منطقه رودالکیلاز اسپانیا برای شناسایی پتانسیل معدنی طلا، الگوریتم جنگل تصادفی (RF) نسبت به سایر الگوریتم‌های RT، ANN، MLA و SVM عملکرد بهتری از خود نشان داده است. RF با داشتن پایداری و استحکام بیشتر در مواجهه با پارامترهای آموزشی مختلف و نتایج بهتر در تحلیل‌های ROC و نرخ موفقیت، به عنوان الگوریتمی با دقت بالاتر معرفی شده است. همچنین، تمامی الگوریتم‌های MLA در شرایطی که شواهد معدنی محدود هستند قابل استفاده هستند و الگوریتم‌های RF و RT قابلیت تفسیر پارامترهای مدل را برای درک کنترل‌های زمین‌شناسی کانی‌سازی ارائه می‌دهند [۲۰]. هریس و همکاران (۲۰۱۵) در تحقیق خود با عنوان "مدل‌سازی داده‌محور و دانش‌محور پتانسیل معدنی در شمال کانادا" از روش جنگل تصادفی برای شناسایی نواحی معدنی استفاده کرده و نتایج آن را با روش‌های دانش‌محور مقایسه کردند. نتایج این تحقیق نشان‌دهنده برتری روش جنگل تصادفی نسبت به روش‌های داده‌محور مرسوم بوده است [۲۱]. کارانزا و لاپورت (۲۰۱۵) در مقاله‌ای با عنوان "مدل‌های داده‌محور پیش‌بینی نواحی پتانسیل‌دار طلا در منطقه باگویو فیلیپین" به بررسی

پژوهشی توسط کاردوسو و همکاران از الگوریتم‌های یادگیری ماشین برای اکتشاف مواد خام استراتژیک مانند لیتیوم (Li)، که در صنایع الکترونیک مصرفی و انرژی‌های سبز استفاده می‌شود، بهره گرفته شده است. منطقه مورد مطالعه، ناحیه Fregeneda-Almendra (بین اسپانیا و پرتغال) است که لیتیوم در پگماتیت‌ها یافت می‌شود. با این حال، اندازه کوچک رخنمون این پگماتیت‌ها چالشی برای روش‌های سنجش از دور محسوب می‌شود. برای غلبه بر این چالش، الگوریتم‌های ماشین بردار پشتیبان (SVM) و جنگل تصادفی (RF) به کار گرفته شدند. این مطالعه با هدف مقایسه دقت عملکرد SVM و RF در نقشه‌برداری لیتولوژیکی، ارزیابی حساسیت این دو طبقه‌بند به عدم توازن کلاس‌ها و مقایسه نتایج با تحقیقات قبلی انجام شده است. داده‌های Sentinel-2 برای این تحلیل استفاده شدند. نتایج نشان داد که SVM دقت کمی بالاتر دارد، اما RF تعداد بیشتری از پگماتیت‌های حاوی لیتیوم را به درستی طبقه‌بندی کرده است. عملکرد مدل‌ها برای تمامی کلاس‌ها یکسان نبوده و هر دو در برخی کلاس‌ها عملکرد ضعیفی داشتند. همچنین، RF تحت تأثیر عدم توازن کلاس قرار گرفت، در حالی که SVM حساسیت کمتری به این موضوع نشان داد [۱۸]. سان و همکاران در پژوهش خود به ارزیابی مدل‌سازی پیش‌بینی پتانسیل معدنی با استفاده از روش‌های یادگیری ماشین در منطقه Tongling در شرق چین می‌پردازند. در این مطالعه، از سه الگوریتم یادگیری ماشین شامل ماشین بردار پشتیبان (SVM)، شبکه‌های عصبی مصنوعی (ANN) و جنگل تصادفی (RF) برای نقشه‌برداری از پتانسیل معدنی استفاده شده است. این مدل‌ها با استفاده از ۱۲ نقشه پیش‌بیننده که فرآیندهای مختلف تشکیل معدنی را نمایان می‌کنند، آموزش دیده و با استفاده از اعتبارسنجی متقابل ۱۰ برابری ارزیابی شدند. نتایج نشان داد که مدل RF نسبت به SVM و ANN عملکرد بهتری داشته و دقت پیش‌بینی بیشتری را ارائه داده است. مدل RF موفق شد اکثر رسوبات شناخته شده را در کوچک‌ترین نواحی پتانسیل‌دار شناسایی کند و نقشه‌ای از مناطق با پتانسیل بسیار بالا، بالا، متوسط و کم ایجاد کرد که ۱۳.۹۷٪ از منطقه مطالعه را پوشش داده و ۸۰.۹۵٪ از رسوبات شناخته شده را در بر می‌گیرد. وجود رسوبات جدید در نواحی پیش‌بینی شده توسط مدل نشان‌دهنده قدرت و کارایی این مدل در شناسایی اهداف اکتشافی است [۱].

پیچیده هستند، عملکرد خوبی از خود نشان دهند؛ بنابراین، استفاده از این الگوریتم‌ها می‌تواند به بهبود فرآیندهای اکتشافی و تسهیل در تصمیم‌گیری‌های مربوط به پتانسیل‌یابی معدنی کمک کند.

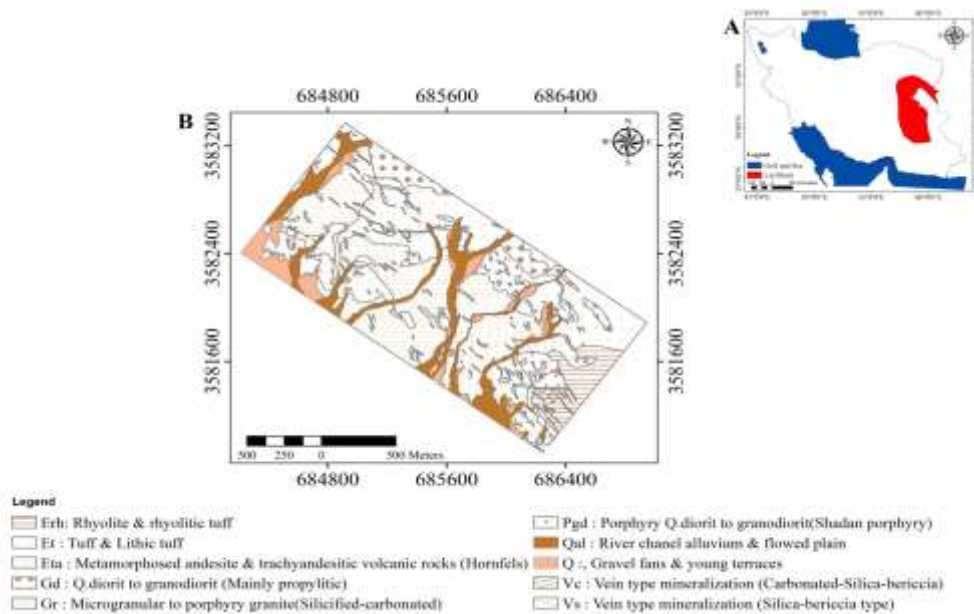
۳- معرفی محدوده مورد مطالعه

محدوده شادان در ۶۵ کیلومتری جنوب شهرستان بیرجند و نزدیک به روستای شیخ‌آباد در استان خراسان جنوبی واقع شده است. این محدوده در نقشه ۱:۱۰۰۰۰۰ سرچاه شور جای گرفته است. رخساره‌های آتشفشانی، دگرسانی آن‌ها در زون‌ها و شدت‌های متفاوت و نزدیکی به کویر لوت، از عوامل اصلی شکل‌گیری ویژگی‌های مورفولوژیکی در این منطقه محسوب می‌شوند. در بخش‌های شمالی و غربی منطقه، مورفولوژی خاص نواحی آتشفشانی مشاهده می‌شود که شامل ارتفاعات تند و پرتگاه‌های سنگی متعدد است. نواحی مرکزی و جنوبی‌تر به صورت کوهپایه‌ای بوده و از تپه‌های کم‌ارتفاع تشکیل شده‌اند. همچنین، از سمت جنوب و غرب منطقه به کویر لوت متصل می‌شود و مخروط افکنه‌های ناشی از مناطق مرتفع، به صورت بادبزن‌های آبرفتی، نقاط پیوند این منطقه با دشت‌های کویری را شکل می‌دهند. منطقه اکتشافی از لحاظ تقسیمات ساختاری در حاشیه شرقی بلوک لوت قرار گرفته است. بیش از ۱۷ توده نفوذی اسیدی تا حدواسط با ترکیب دیوریت تا سینیت در منطقه شناسایی شده است. نقشه زمین شناسی محدوده با مقیاس ۱:۱۰۰۰ در شکل (۱) ارائه شده است. زونهای دگرسانی هیدروترمالی مرتبط با توده‌های نفوذی شامل آرژیلیک، پروپیلیتیک، آرژیلیک پیشرفته، سیلیسی، کوارتز، سرسیت، پیریت، گوسان و برشهای هیدروترمالی است که اغلب با یک دیگر تداخل کرده و این امر به همراه هوازدگی شدید سبب پیچیدگی تفکیک زونهای دگرسانی شده است. کانی‌سازی به شکل سولفیدی پیریت و به مقدار کمتر کالکوپیریت و اکسیدهای آهن دانه پراکنده و رگچه‌های استوک ورک کوارتز اکسید آهن مشاهده می‌شود با توجه به نقشه زمین شناسی محدوده مطالعاتی، این ناحیه از سنگهای پورفیری، آندزیتی و لکانیک‌های ائوسن توف‌گدازه و پیروکسن آندزیت و نفوذی‌های میکرودیوریتی و گاه مونزونیتی کوارتز مونزونیتی تشکیل شده است. با وجود این که در کل محدوده میکرودیوریت نسبت به توده‌های پورفیری آندزیتی گسترش وسیعی یافته است کانی‌سازی

عملکرد الگوریتم جنگل تصادفی پرداختند و نتایج نشان داد که این الگوریتم قادر به ایجاد ارتباط مکانی میان نقشه‌های پیش‌بینی و داده‌های آموزشی بوده و نسبت به سایر روش‌های پتانسیل‌یابی مزایای نسبی دارد [۲۲]. رودریگز و همکاران (۲۰۱۴) در مقاله‌ای با عنوان "مدل‌سازی پیش‌بینی طلا با استفاده از داده‌های چندامتیازی با الگوریتم جنگل تصادفی" به بررسی نقشه پتانسیل طلا در منطقه رودالکویلا اسپانیا پرداختند و نتایج را با روش رگرسیون لجستیکی مقایسه کردند. این تحقیق از لایه‌های اطلاعاتی مختلف شامل آلتراسیون‌های هیدروترمال، داده‌های ژئوشیمیایی، ژئوفیزیکی و زمین‌شناسی استفاده کرده و مزایای استفاده از جنگل تصادفی را شامل تنظیمات ساده پارامترهای ورودی، تخمین نااریب خطای پیش‌بینی، توانایی پردازش داده‌های پیچیده و تشخیص اهمیت متغیرها ذکر کرده است [۲۳]. عابدی و همکاران (۲۰۱۲) در مقاله‌ای با عنوان "ماشین بردار پشتیبان به عنوان ابزار طبقه‌بندی چندکلاسه نقشه پتانسیل معدنی" از SVM برای تهیه نقشه پتانسیل معدنی مس در استان کرمان استفاده کردند. این مطالعه از دو تابع کرنل چندجمله‌ای و کرنل با تابع شعاعی (RBF) بهره برده و برای اعتبارسنجی مدل از روش کنار گذاشتن اطلاعات یک گمانه و تکرار استفاده کرده است. نتایج نشان‌دهنده نرخ موفقیت ۵۲ درصدی این الگوریتم در شناسایی نقاط آنومالی بوده است [۲۴]. ژو و کارانزا (۲۰۱۱) در مقاله خود تحت عنوان "ماشین برداری پشتیبان: ابزاری برای مدل‌سازی پتانسیل معدنی" به معرفی الگوریتم SVM به عنوان روشی داده‌محور و نظارت شده پرداخته و از آن برای شناسایی نواحی مستعد طلا در کانادا استفاده کرده‌اند. این تحقیق از توابع کرنل مختلف شامل کرنل‌های خطی، چندجمله‌ای با درجه آزادی d ، تابع پایه شعاعی (RBF)، و سیگموئید بهره برده و نتایج را با خروجی روش Weight of Evidence (WofE) مقایسه کرده است. نتایج حاصل نشان‌دهنده دقت ۱۰۰ درصدی در شناسایی نواحی معدنی و ۸۵ درصدی در شناسایی نواحی غیرمعدنی بوده است [۲۵]. به‌طورکلی، نتایج مطالعات مختلف نشان داده است که استفاده از SVM و RF می‌تواند به‌دقت بالاتری در شناسایی نواحی معدنی منجر شود. این روش‌ها به طور قابل‌توجهی نسبت به روش‌های کلاسیک کارآمدتر هستند و می‌توانند در شرایطی که داده‌های آموزشی محدود یا

دسته گسل‌هایی با روند شمال غرب جنوب شرق که دایک‌ها نیز از همین روند تبعیت می‌کنند و دسته دیگر با روند شمال شرق جنوب غرب می‌باشد تعداد کمتری نیز با روند شمالی جنوبی و با طولی کمتر در منطقه دیده می‌شود؛ اما اثر آن‌ها چندان معلوم نیست در بخش شمالی و شرقی عملکرد تکتونیک بیشتر است. وجود گسل‌ها و خطواره‌های فراوان با دگرسانی و کانی‌سازی بسیار مرتبط است. ناحیه مورد بررسی در تقسیم‌بندی‌های رایج زمین‌شناسی ایران بخشی از بلوک لوت خردقاره ایران مرکزی و منطقه فیلیش - کمربند افیولیتی شرق ایران بوده و از نظر ساختار زمین‌شناسی ویژگی و اختصاصات این دو بخش از ایران را توأمأ دارا است بلوک لوت به درازای ۹۰۰ کیلومتر میان دو گسل نایبند در غرب و گسل نه‌بندان در شرق محصور گشته است و از شمال به گسل درونه و از جنوب به فرونشست جازموریان محصور می‌گردد.

در یک فاز نفوذی موزونیتی دگرسان شده در بخش شمال شرقی وجود دارد. همچنین تورمالین تقریباً همیشه همراه کربنات در بخش شرقی بیش از بخش غربی منطقه دیده می‌شود که این بخش را می‌توان زون لیستونیتی نیز نامید در بخش مرکزی آثاری از آلتراسیون پتاسیک یک توده آندزیتی میکرودیوریتی با رگچه‌های کوارتز خاکستری حاوی مالاکیت و لیمونیت مشاهده می‌گردد. این بخش خود با آلتراسیون آرژیلیک متوسط پوشیده می‌شود که احتمالاً موجب افزایش رگچه‌های سولفید کوارتز شده است. در بخش شرقی نیز مالاکیت و رگچه‌های کوارتز ورقه‌ای برش‌های انفجاری به همراه اکتینولیت و کوارتز بسیار دانه‌درشت دیده می‌شود. در شمال و شرق محدوده آلتراسیون پروپیلیتی (کلریت، اپیدوت و پیریت آندزیت‌ها و ولکانی کلاستیک‌ها توف) و گدازه را تحت‌تأثیر قرار داده است. در این منطقه گسل‌های فراوانی وجود دارد که دودسته گسل و خطواره اصلی در آن دیده می‌شود یک



شکل ۱- A موقعیت بلوک لوت در شرق ایران و B: نقشه زمین‌شناسی محدوده شادان با مقیاس ۱:۱۰۰۰

مس، مولیبدن و در برخی موارد طلا و نقره، به طور گسترده‌ای مورد مطالعه قرار گرفته‌اند. مدل مفهومی کانسار مس پورفیری ترکیبی از ویژگی‌های زمین‌شناسی، ژئوشیمی، ژئوفیزیک و آلتراسیون هیدروترمال است که به شناسایی، اکتشاف و بهره‌برداری از این ذخایر کمک می‌کند. ۱- ویژگی‌های سنگ‌شناسی: کانسارهای مس پورفیری معمولاً در ارتباط با توده‌های نفوذی آذرین حد واسط تا

۴- مدل مفهومی کانسار مس پورفیری و داده‌های اکتشافی

کانسارهای مس پورفیری یکی از مهم‌ترین انواع کانسارهای فلزی در مقیاس جهانی هستند که اغلب با سیستم‌های ماگمایی و فرایندهای هیدروترمال پیچیده در اعماق کم پوسته زمین همراه هستند. این کانسارها به دلیل توزیع گسترده، حجم عظیم ذخایر و عیارهای اقتصادی عناصر

دهنده پیوسته بودن آنومالی می‌باشد. نقشه‌های میدان مغناطیس کل با ۵ متر ادامه فراسو و نقشه‌ی برگردان به قطب داده‌های ژئومغناطیس پس از انجام پردازش‌های اولیه و تصحیحات لازم تولید شده است. به دلیل اینکه این نقشه-ها مشخص‌کننده محل زون‌های دگرسانی این دو نقشه به عنوان نقشه‌های شاهد معیار ژئوفیزیکی انتخاب گردیده است (شکل ۹ و شکل ۱۰).

۴- آلتراسیون هیدروترمال آلتراسیون و توزیع آن در سیستم‌های پورفیری: آلتراسیون هیدروترمال در کنسارهای پورفیری به تغییرات مینرالوژیکی گسترده‌ای در سنگ‌های میزبان منجر می‌شود که به‌طور مستقیم با فرایندهای کانی‌سازی مرتبط است. این زون‌ها به‌طور کلی به چهار دسته تقسیم می‌شوند آلتراسیون پتاسیک در زون مرکزی و نزدیک به توده‌های نفوذی رخ می‌دهد. این زون به‌طور عمده شامل کانی‌های ارتوز، بیوتیت و گاهی مگنتیت است و به دلیل دما و فشار بالای سیالات هیدروترمال شکل می‌گیرد. آلتراسیون فیلیک در اطراف زون پتاسیک قرار دارد و شامل کانی‌هایی مانند سرسیت، کوارتز و پیریت است. این نوع آلتراسیون در دماهای پایین‌تر و محیط‌های اسیدی‌تر نسبت به زون پتاسیک رخ می‌دهد. آلتراسیون پروپلیتیک در مناطق حاشیه‌ای کنسار رخ داده و شامل کانی‌هایی مانند کلریت، اپیدوت و کربنات است. این زون نشان‌دهنده دماهای پایین‌تر و حضور سیالات ضعیف‌تر است. آلتراسیون آرژیلیک معمولاً در نواحی سطحی و بالایی کنسار یافت می‌شود و شامل کانی‌های رسی مانند کائولینیت و اسمکتیت است. حضور گسترده اکسیدهای آهن نشان‌دهنده فعالیت‌های هیدروترمالی است که می‌تواند با کانی‌سازی‌هایی مانند آهن، مس، طلا و حتی سولفیدهای فلزی مانند پیریت همراه باشد [۲۶، ۳۲].

برای بارزسازی زون‌های آلتراسیون در این محدوده، از تصاویر ASTER استفاده شده و تحلیل‌های سنجش از دور جهت بارزسازی آلتراسیون‌های مرتبط با کانه‌زایی انجام شده است. دگرسانی هیدروترمال منطقه گسترده است و مطالعات انجام شده نشان می‌دهد که این دگرسانی‌ها ارتباط بسیار خوبی با کانی‌سازی در منطقه دارد. در این منطقه از تصاویر استر برای شناسایی آلتراسیون‌ها استفاده شده است. آلتراسیون آرژیلیک، اکسید آهن و پروپلیتیک مهمترین آلتراسیون‌های مرتبط با ذخایر مس پورفیری هستند و می‌توانند در مراحل اکتشاف راهنمای بسیار

داده‌ها انجام شد. نتایج این تحلیل چهار مؤلفه اصلی را استخراج کرد که مجموعاً ۷۵ درصد از تغییرپذیری منطقه را تبیین می‌کنند. قرارگیری عناصر مس، طلا و مولیبدن در فاکتور اول به‌عنوان فاکتور کانی‌زایی، تأییدی بر وجود سیستم کانی‌زایی پورفیری در منطقه است. در این راستا نقشه‌های آنومالی عناصر مس، طلا و مولیبدن و همچنین نقشه امتیازات فاکتوری فاکتور اول به‌عنوان لایه‌های معیار ژئوشیمیایی انتخاب شدند (شکل ۵ تا شکل ۸).

۳- ویژگی‌های ژئوفیزیکی: ژئومغناطیس به دلیل ویژگی‌های حجمی و گستردگی کنسارهای پورفیری، یکی از ابزارهای کلیدی در اکتشاف این ذخایر است. این روش‌ها می‌توانند ناهنجاری‌های ساختاری، چگالی و مغناطیسی مرتبط با توده‌های نفوذی و زون‌های کانی‌سازی را شناسایی کنند. از جمله روش‌های مهم ژئوفیزیکی در اکتشاف کنسارهای پورفیری عبارت‌اند از مغناطیس‌سنجی: این روش برای شناسایی ناهنجاری‌های مغناطیسی ناشی از توده‌های نفوذی مافیک یا حضور کانی‌های مغناطیسی مانند مگنتیت و هماتیت استفاده می‌شود [۲۶، ۳۱].

در راستای انجام مطالعات ژئومغناطیسی در این محدوده، ۵۴۷ ایستگاه مغناطیس‌سنجی بر روی ۳۵ پروفیل به‌صورت منظم و مربعی طراحی شده‌اند. این برداشت‌ها در یک شبکه منظم با ابعاد ۷۰×۷۰ متر، با توجه به روند کانه‌زایی شرقی-غربی، به‌صورت پروفیل‌های شمالی-جنوبی انجام شده‌اند. ابزار مورد استفاده برای برداشت داده‌های مغناطیس‌سنجی، مگنتومتر پروتون مدل GEM بوده و موقعیت نقاط ایستگاه‌ها با استفاده از دستگاه GPS ثبت شده است.

ژئوفیزیک به دلیل ویژگی‌های حجمی و گستردگی کنسارهای پورفیری، یکی از ابزارهای کلیدی در اکتشاف این ذخایر است. این روش‌ها می‌توانند ناهنجاری‌های ساختاری، چگالی و مغناطیسی مرتبط با توده‌های نفوذی و زون‌های کانی‌سازی را شناسایی کنند. از جمله روش‌های مهم ژئوفیزیکی در اکتشاف کنسارهای پورفیری عبارت‌اند از مغناطیس‌سنجی: این روش برای شناسایی ناهنجاری‌های مغناطیسی ناشی از توده‌های نفوذی مافیک یا حضور کانی‌های مغناطیسی مانند مگنتیت و هماتیت استفاده می‌شود. در بررسی‌های مغناطیس‌سنجی در منطقه بر اساس نقشه میدان کل مغناطیسی یک آنومالی وسیع با امتداد شمال غربی - جنوب شرقی در محدوده را نشان می‌دهد. نتایج حاصل از مشتق قائم و سیگنال تحلیلی نشان

اکسید آهن به عنوان لایه‌های شاهد اطلاعاتی مرتبط با آلتراسیون‌ها استفاده شده است (شکل ۱۱ تا شکل ۱۳). وجود دگرسانی پروپلیتیک در حاشیه و دگرسانی سربسیتیک و آرژیلیک در مرکز بسیار نزدیک به الگوی دگرسانی در ذخایر مس پورفیری می‌باشد. جدول ۱ خلاصه‌ای از داده‌های استفاده شده در این مطالعه را ارائه داده است.

مناسبی جهت کشف این ذخایر باشند. آلتراسیون‌های پروپلیتیک و آرژیلیک با استفاده از روش SAM و آلتراسیون اکسید آهن با استفاده از روش نسبت باندی بارسازی شده است. در نتیجه یک تصویر حاصل از تلفیق rule فایل‌های مرتبط با آلتراسیون پروپلیتیک، یک تصویر حاصل از تلفیق rule فایل‌های مرتبط با آلتراسیون آرژیلیک و یک تصویر حاصل از نسبت باندی مرتبط با آلتراسیون

جدول ۱- خلاصه‌ای از داده‌های استفاده شده در این مطالعه

توضیح	مجموعه داده‌ای اکتشافی
نقشه زمین‌شناسی با مقیاس ۱:۱۰۰۰ محدوده	زمین‌شناسی
گسل‌ها و خطواره استخراج شده از نقشه زمین‌شناسی	ساختاری
۳۰۰ نمونه سنگی آنالیز شده برای مس، طلا و مولیبدن	ژئوشیمیایی
۱ شیت تصویر ASTER دانلود شده از سایت USGS	تصاویر ماهواره ASTER
۵۴۷ ایستگاه مغناطیس‌سنجی بر روی ۳۵ پروفیل به صورت منظم و مربعی در یک شبکه منظم با ابعاد ۷۰×۷۰ متر	ژئومغناطیس زمینی
۲۷ گمانه حفاری شده در محدوده	داده‌های حفاری

جدول ۲- خلاصه‌ای از لایه‌های اطلاعاتی مورد استفاده در این پژوهش (مجموع لایه‌ها ۱۱ عدد است)

معیار	لایه شاهد اطلاعاتی	دلیل استفاده
زمین‌شناسی	اکسید آهن	مناطق مناسب برای رخداد کانی‌سازی
	زون‌های آلتراسیون	
	پروپلیتیک	
ژئوشیمی	مناطق گسله	غنی‌شدگی کانی‌سازی
	واحدهای سنگی	
	نقشه آنومالی Au	
	نقشه آنومالی Cu	
ژئوفیزیک	نقشه آنومالی Mo	مشخص‌کننده زون‌های آلتراسیون
	نقشه فاکتور مرتبط با کانی‌سازی در آنالیز چندمتغیره	
	نقشه میدان مغناطیسی کل	
	نقشه برگردان به قطب	

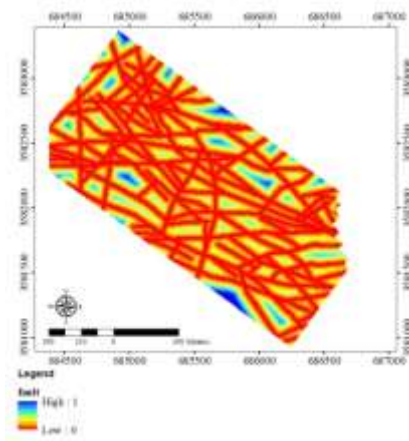
$$\bar{x} = \frac{x - \min(x)}{\max(x) - \min(x)}$$

که در آن x داده اصلی و \bar{x} مقدار نرمالایز شده می‌باشد [۳۴].

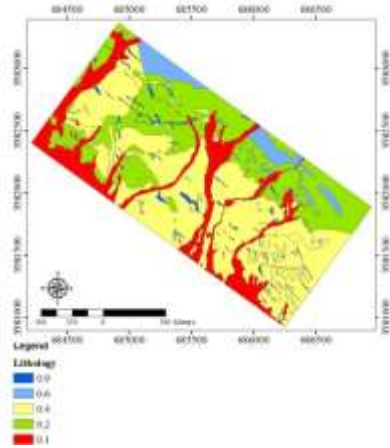
فقط برای لایه واحدهای سنگی با توجه به مقادیر گسسته آن‌ها مقادیر ۰/۳، ۰/۴، ۰/۵، ۰/۶، ۰/۷، ۰/۸، ۰/۹ و برای ۱ تا ۷ انتخاب شدند. همچنین مقادیر ۰/۱ و ۰/۲ برای واحدهای توف و رسوبات کواترنری که کد آن‌ها ۸ و ۹ بود، انتخاب شده است. تمام لایه‌های اطلاعاتی به صورت نرمالایز شده در شکل ۳ (۳) تا شکل ۱۳ (۱۳) شده است.

باتوجه به این که رنج داده‌ها و واحد اندازه‌گیری آن‌ها در لایه‌های اطلاعاتی متفاوت می‌باشند، در الگوریتم‌های یادگیری ماشین عملکرد تابع با محدودیت سخت‌افزاری و نرم‌افزاری همراه است و نمی‌تواند عملکرد مثبتی داشته باشد. برای مثال، در این پژوهش داده‌های ژئوشیمیایی مقادیری بسیار کوچک‌تر از داده‌های ژئومغناطیس دارند؛ بنابراین نیاز است تا مقادیر پردازش شده نرمالایز شوند [۳۳].

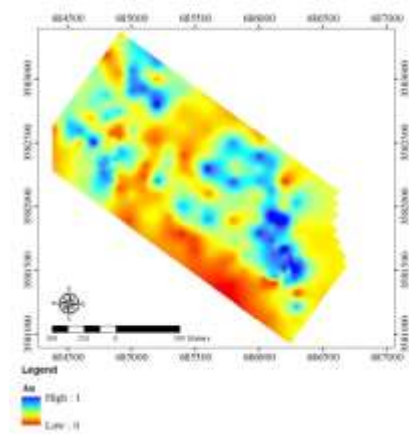
ساده‌ترین روش برای بی‌مقیاس کردن این است که مقادیر را به مقیاس [۰, ۱] تغییر دهیم. فرمول این روش به صورت زیر می‌باشد.



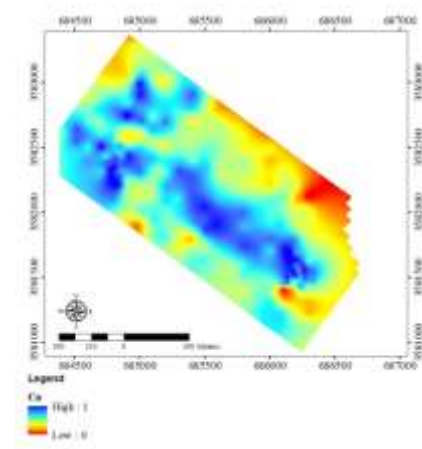
شکل ۴- نقشه فاصله از گسل‌ها به صورت بی‌مقیاس شده



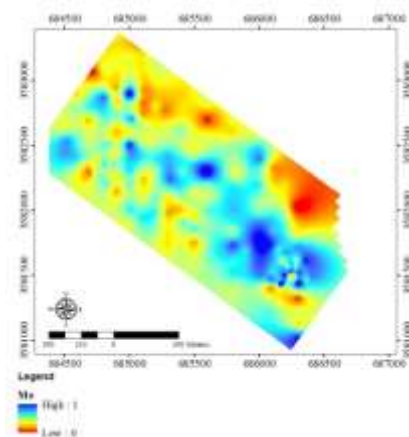
شکل ۳- نقشه زمین‌شناسی ۱:۱۰۰۰ محدوده شادان دسته‌بندی شده بر اساس دیاگرام سینگر



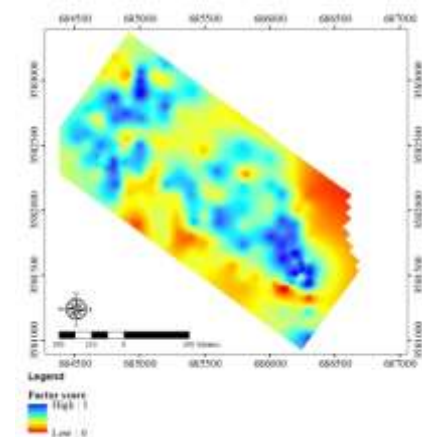
شکل ۶- نقشه توزیع جوامع ژئوشیمیایی طلا به صورت بی‌مقیاس شده



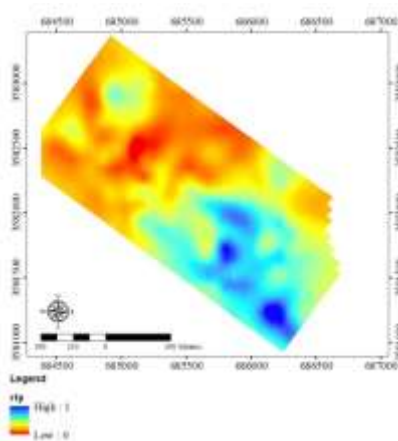
شکل ۵- نقشه توزیع جوامع ژئوشیمیایی مس به صورت بی‌مقیاس شده



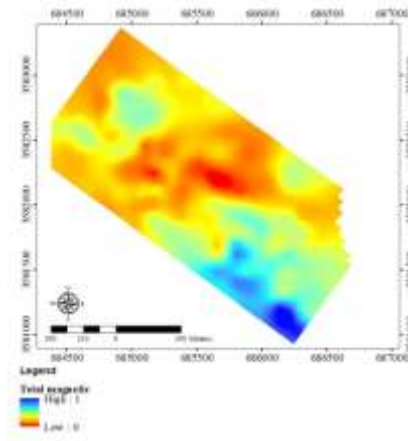
شکل ۸- نقشه توزیع جوامع ژئوشیمیایی مولیبدن به صورت بی‌مقیاس شده



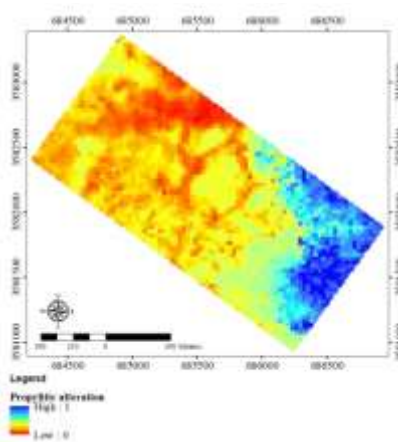
شکل ۷- نقشه توزیع جوامع ژئوشیمیایی امتیازات فاکتوری فاکتور اول به صورت بی‌مقیاس شده



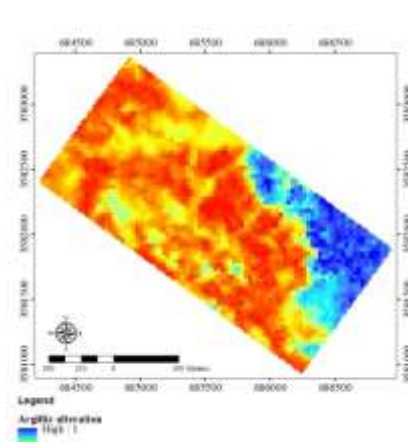
شکل ۱۰- نقشه برگردان به قطب به صورت بی‌مقیاس شده



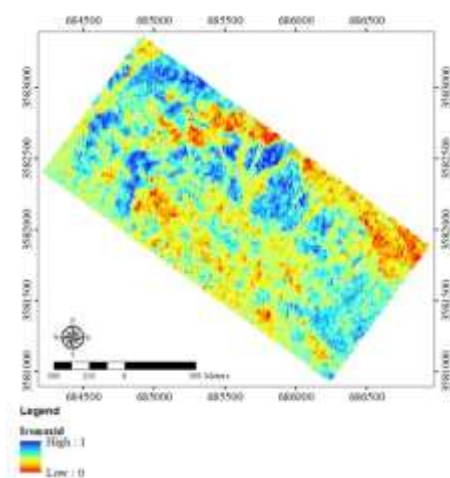
شکل ۹- نقشه میدان کل مغناطیسی به صورت بی‌مقیاس شده



شکل ۱۲- تصویر حاصل از تلفیق نتایج SAM مرتبط با آلتراسیون پروپلیتیک به صورت بی‌مقیاس شده



شکل ۱۱- تصویر حاصل از تلفیق نتایج SAM مرتبط با آلتراسیون آرژیلیک به صورت بی‌مقیاس شده



شکل ۱۳- تصویر حاصل نسبت باندی مرتبط با آلتراسیون اکسید آهن به صورت بی‌مقیاس شده

۵- نقشه پتانسیل معدنی: در اکتشافات، شناسایی مناطقی که باید در مراحل بعدی مورد توجه قرار بگیرند با استفاده از مدل‌سازی پتانسیل معدنی صورت می‌گیرد. مدل‌سازی پتانسیل معدنی در واقع یافتن مکان‌های یا مناطقی است که یک سری معیارها و ملاک‌ها برای حضور ذخیره در آن‌ها صدق می‌کند. در فرایند اکتشافات معدنی، یکی از اهداف اصلی شناسایی مناطقی است که پتانسیل بالایی برای حضور ذخایر معدنی دارند. این شناسایی از طریق مدل‌سازی پتانسیل معدنی انجام می‌شود. در واقع، این مدل‌سازی یک ابزار پیش‌بینی است که به متخصصین اکتشاف کمک می‌کند تا مناطقی را مشخص کنند که دارای ویژگی‌ها و معیارهای لازم برای حضور ذخایر معدنی باشند [۳۵، ۳۶]. انواع روش‌ها در زیر توضیح داده شده است.

می‌شود. در نهایت، ارزیابی مدل به منظور سنجش صحت دانش تولیدشده انجام می‌گیرد تا از قابلیت کاربرد عملی آن اطمینان حاصل شود.

الگوریتم‌های یادگیری ماشین به چهار دسته اصلی تقسیم می‌شوند [۳۴]:

۱. یادگیری نظارت‌شده: در این روش از داده‌های برچسب‌دار استفاده می‌شود. متغیرهای مستقل به عنوان ورودی و متغیرهای وابسته یا هدف به عنوان خروجی مشخص هستند. هدف این الگوریتم‌ها، یافتن تابعی است که ورودی‌ها را به خروجی‌ها مرتبط کند. مثال‌ها: رگرسیون، درخت تصمیم، K نزدیک‌ترین همسایه، ماشین بردار پشتیبان (SVM).

۲. یادگیری نظارت‌نشده: در این روش برچسب داده‌ها مشخص نیست و الگوریتم سعی دارد الگوها و ساختارهای پنهان در داده‌ها را شناسایی کند. خوشه‌بندی یکی از مهم‌ترین مثال‌های یادگیری غیرنظارتی است که هدف آن گروه‌بندی داده‌های مشابه است.

۳. یادگیری تقویتی: در این روش، یک عامل با تعامل با محیط، از بازخوردهای دریافتی برای بهبود تصمیم‌گیری خود استفاده می‌کند. هر تصمیم یک امتیاز یا بازخورد دریافت می‌کند که به عامل کمک می‌کند تصمیمات بهتری در آینده بگیرد.

۴. یادگیری نیمه‌نظارتی: این روش از ترکیب داده‌های برچسب‌دار و بدون برچسب استفاده می‌کند. هدف این است که از تعداد کم داده‌های برچسب‌دار و تعداد زیاد داده‌های بدون برچسب، الگوریتم بهتری برای طبقه‌بندی ایجاد شود. یادگیری فعال یکی از روش‌های این نوع یادگیری است.

در این تحقیق برای ارزیابی مدل‌های دسته‌بندی از معیارهایی مانند دقت نرخ دسته‌بندی، حساسیت، Cross Validation، نمودار ROC و نمودار میزان تغییرات پیش‌بینی - مساحت بهبودیافته استفاده شده است.

۱. دقت نرخ دسته‌بندی: برای محاسبه این معیار از ماتریس درهم‌ریختگی (Confusion Matrix) استفاده می‌شود (شکل ۱۴). این ماتریس شامل چهار مفهوم اصلی است:

الف) داده‌مبنا: روش‌های رایج در این دسته شامل تحلیل خوشه‌ای، وزن‌دهی شواهد (Weight of Evidence) و رگرسیون لجستیک هستند. این روش به طور مستقیم از داده‌های موجود برای پیش‌بینی استفاده می‌کند.

ب) دانش مبنا: روش‌های معمول شامل سیستم‌های خبره (Expert Systems) و روش‌های فازی (Fuzzy Logic) هستند که از دانش تخصصی در فرایند پیش‌بینی بهره می‌برند.

ج) ترکیبی: (Hybrid) این رویکرد ترکیبی از دو روش داده‌مبنا و دانش مبنا است. مدل‌سازی ترکیبی با استفاده از تلفیق داده‌های واقعی و دانش تخصصی، پیش‌بینی‌های دقیق‌تری ارائه می‌دهد. روش‌های رایج در این دسته شامل شبکه‌های عصبی مصنوعی (ANN) و الگوریتم‌های ماشین یادگیری مانند جنگل تصادفی (Random Forest) و ماشین بردار پشتیبان (SVM) هستند.

۵- یادگیری ماشین

در دو دهه گذشته، الگوریتم‌های هوش مصنوعی به طور گسترده‌ای در حل مسائل مرتبط با علوم مهندسی به کار رفته‌اند. با این حال، استفاده از این روش‌ها در علوم زمین به دلیل نوآوری نسبی، با محدودیت‌هایی همراه بوده است [۱۲]. الگوریتم‌های یادگیری ماشین به عنوان زیرشاخه‌ای از هوش مصنوعی، باهدف دستیابی به نتایج با درصد اطمینان قابل قبول، توسعه یافته‌اند. هدف اصلی یادگیری ماشین، بهینه‌سازی عملکرد سیستم‌ها با استفاده از تجربیات گذشته است [۳۷]. یادگیری ماشین یکی از مهم‌ترین بخش‌های هوش مصنوعی است که از علوم کامپیوتر آغاز شده و به سرعت در حال رشد است به طوری که در حوزه‌های دیگر علوم نیز کاربردهای فراوانی یافته است.

فرایند داده‌کاوی شامل سه مرحله آماده سازی داده، یادگیری مدل و ارزیابی و تفسیر مدل می‌باشد. فرایند داده‌کاوی شامل سه مرحله اصلی است: آماده‌سازی داده، یادگیری مدل و ارزیابی و تفسیر مدل. ابتدا با تحلیل اکتشافی داده‌ها، ویژگی‌های آماری آن‌ها بررسی شده و عملیات پیش‌پردازش برای آماده‌سازی داده‌ها انجام می‌شود. سپس در مرحله یادگیری مدل، با استفاده از الگوریتم‌های مختلف، الگوها و ساختارهای پنهان در داده‌ها شناسایی

در مقابل نرخ مثبت واقعی (TPR) در نقاط مختلف آستانه قرار می‌دهد. TPR همان حساسیت است که نشان می‌دهد چه تعداد از موارد مثبت به درستی تشخیص داده شده‌اند. FPR نشان می‌دهد چه تعداد از موارد منفی به اشتباه به عنوان مثبت دسته‌بندی شده‌اند. هرچه منحنی ROC به گوشه بالا سمت چپ نزدیک‌تر باشد، عملکرد مدل بهتر است. مساحت زیر منحنی ROC که به آن AUC گفته می‌شود، معیاری برای سنجش عملکرد کلی مدل است. AUC نزدیک به ۱ نشان‌دهنده مدل عالی و AUC نزدیک به ۰.۵ نشان‌دهنده عملکرد تصادفی است [۳۸].

۵. نمودار میزان پیش‌بینی - مساحت بهبودیافته (P-A): این روش به صورت خاص برای ارزیابی مدل‌های مختلف پیش‌بینی در تعیین احتمال وقوع ذخایر معدنی استفاده می‌شود. در این روش، نرخ پیش‌بینی مدل برای نقاط مثبت و منفی محاسبه می‌شود و در مقابل آن، مساحتی از منطقه که توسط این پیش‌بینی‌ها پوشش داده شده است، رسم می‌گردد. سپس منحنی این دو متغیر (نرخ پیش‌بینی و مساحت) در نمودار ترسیم می‌شود. نقطه تقاطع نمودار نرخ پیش‌بینی و نمودار مساحت منطقه در روش P-A بیانگر نقطه‌ای است که در آن مدل بهینه‌ترین حالت ممکن برای پیش‌بینی ذخایر معدنی را نشان می‌دهد. به عبارت دیگر، این نقطه جایی است که مدل حداکثر پیش‌بینی مثبت‌ها را با کمترین مساحت ممکن پوشش می‌دهد؛ بنابراین برای به دست آوردن عدد نهایی عملکرد لازم است این دو عدد از هم سر شده تا عدد نهایی عملکرد هر مدل به دست آید. این روش به طور خلاصه به بررسی این موضوع می‌پردازد که مدل مورد استفاده تا چه حد قادر است نواحی مثبت (مانند مناطق دارای مواد معدنی) را به درستی پیش‌بینی کند [۳۹-۴۱].

در این پژوهش از الگوریتم‌های نظارت شده ماشین بردار پشتیبان و جنگل تصادفی جهت تهیه نقشه پتانسیل معدنی در محدوده شادان استفاده شده است.

۵-۱- ماشین بردار پشتیبان

SVM یک الگوریتم یادگیری نظارت شده است که عمدتاً برای دسته‌بندی و رگرسیون استفاده می‌شود. هدف اصلی آن یافتن یک فرا صفحه (Hyperplane) بهینه است که بتواند داده‌ها را به دو دسته مجزا تفکیک کند. فرا صفحه به گونه‌ای انتخاب می‌شود که حاشیه (Margin)، یعنی فاصله بین نزدیک‌ترین نقاط هر دسته به فرا صفحه، به

		Actual value	
		positive	negative
Predicted value	positive	TP	FP
	negative	FN	TN

شکل ۱۴- ماتریس درهم‌ریختگی

TN (True Negative) رکوردهایی که به درستی منفی تشخیص داده شده‌اند، TP (True Positive) رکوردهایی که به درستی مثبت تشخیص داده شده‌اند، False Positive) FP رکوردهایی که به اشتباه مثبت تشخیص داده شده‌اند، False Negative) FN رکوردهایی که به اشتباه منفی تشخیص داده شده‌اند [۱۴].

دقت نرخ دسته‌بندی نشان می‌دهد چه درصدی از کل رکوردها به درستی دسته‌بندی شده‌اند و به صورت زیر محاسبه می‌شود:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

۲. حساسیت (Sensitivity): برای ارزیابی دقیق‌تر مدل در دسته‌بندی یک دسته خاص (مانند تشخیص بیماری)، از حساسیت استفاده می‌شود. حساسیت درصد موارد مثبت واقعی است که به درستی شناسایی شده‌اند و به صورت زیر محاسبه می‌شود [۱۴]:

$$Sensitivity = \frac{TP}{TP + FN}$$

۳. Cross Validation: این روش برای جلوگیری از بیش‌برازش (Overfitting) و ارزیابی دقیق‌تر مدل استفاده می‌شود. داده‌ها به K زیرمجموعه تقسیم شده و هر بار یک زیرمجموعه برای اعتبارسنجی و بقیه برای آموزش استفاده می‌شوند. این فرایند K بار تکرار می‌شود و میانگین نتایج به عنوان تخمین نهایی انتخاب می‌شود. معمولاً از 10-fold Cross Validation استفاده می‌شود. حالت خاص آن، Leave-One-Out است که در آن K برابر با تعداد رکوردها (N) است [۱۴].

۴. نمودار ROC: این نمودار یک ابزار گرافیکی است که برای ارزیابی عملکرد مدل‌های دسته‌بندی استفاده می‌شود. این نمودار منحنی‌ای است که نرخ مثبت کاذب (FPR) را

درخت تصمیم که به طور تصادفی ساخته شده‌اند، به کاهش بیش برآزش (Overfitting) کمک می‌کند. به عبارت دیگر، مدل نهایی (جنگل تصادفی) معمولاً با داده‌های جدید بهتر عمل می‌کند نسبت به هر یک از درخت‌های تصمیم به صورت جداگانه. یکی از خروجی‌های کلیدی روش جنگل تصادفی (Random Forest) است که به تحلیل و تفسیر مدل کمک می‌کند. این مفهوم نشان می‌دهد که هر ویژگی در داده‌های ورودی تا چه اندازه در فرایند تصمیم‌گیری مدل تأثیرگذار است. اهمیت ویژگی‌ها در جنگل تصادفی به ما این امکان را می‌دهد که بفهمیم کدام ویژگی‌ها برای پیش‌بینی نتیجه‌های مدل بیشترین تأثیر را دارند و کدام ویژگی‌ها کمترین تأثیر را دارند. در هر تقسیم کردن (split) درخت تصمیم، معیار Gini به کار می‌رود تا مشخص شود که تقسیم‌گیری چقدر می‌تواند انحراف کلاس‌ها را کاهش دهد. اهمیت ویژگی‌ها می‌تواند با توجه به کاهش امتیاز Gini که هر ویژگی به وجود می‌آورد، محاسبه شود [۴۳].

در هر دو روش پارامترهای بنیادی وجود دارد که می‌بایست در هنگام مدل‌سازی بهینه شوند تا مدل دقت مدل به حداکثر برسد. در روش SVM پارامتر C میزان جریمه‌ای را که به خطاهای طبقه‌بندی در هنگام آموزش داده می‌شود، تنظیم می‌کند. مقدار بزرگ C موجب می‌شود مدل به دقت بیشتری روی داده‌های آموزشی توجه کند و احتمال بیش برآزش (Overfitting) افزایش یابد. مقدار کوچک C به مدل اجازه می‌دهد تا اشتباهات بیشتری را بپذیرد و می‌تواند از بیش برآزش جلوگیری کند. γ (Gamma) در هسته RBF (Gaussian) تأثیر دامنه هر نمونه داده را تعیین می‌کند. مقدار کم برای Gamma به معنی تأثیر وسیع‌تری از هر نمونه داده است و مقدار زیاد به معنی تأثیر محدودتر می‌باشد. در هسته چندجمله‌ای پارامترهای درجه چندجمله‌ای و مقیاس ویژگی‌ها می‌توانند تنظیم شوند. همچنین انتخاب نوع هسته (مانند خطی، RBF، چندجمله‌ای یا سیگموئید) بستگی به نوع داده‌ها و پیچیدگی مسئله دارد [۴۲]. در روش RF پارامتر (n_estimators) تعداد درخت‌های تصمیمی است که در جنگل تصادفی ساخته می‌شود. تعداد بیشتر درخت‌ها معمولاً دقت را افزایش می‌دهد، اما ممکن است زمان محاسباتی را نیز بالا ببرد. پارامتر (max_depth) حداکثر عمق هر درخت تصمیم است. عمق بیشتر می‌تواند به مدل

حداکثر برسد. این نقاط مرزی که نزدیک‌ترین نمونه‌ها به فرا صفحه هستند، بردارهای پشتیبان (Support Vectors) نامیده می‌شوند SVM. به دنبال یافتن فرا صفحه‌ای است که بیشترین فاصله را از داده‌های دو دسته داشته باشد. این حاشیه بزرگ‌تر به معنی افزایش اطمینان از دسته‌بندی صحیح داده‌های جدید است. یکی از قابلیت‌هایی روش SVM قابلیت کار با داده‌های غیرخطی در بسیاری از مسائل، داده‌ها به صورت خطی قابل تفکیک نیستند. در این موارد، SVM با استفاده از توابع هسته (Kernel Functions) داده‌ها را به فضای با ابعاد بالاتر نگاشت می‌دهد، جایی که بتوان آن‌ها را به صورت خطی تفکیک کرد. متداول‌ترین توابع هسته شامل هسته خطی، هسته چندجمله‌ای و هسته تابع پایه شعاعی (RBF) هستند. پارامتر C در SVM کنترل‌کننده میزان انعطاف‌پذیری مدل است. این پارامتر تعیین می‌کند که مدل تا چه حد می‌تواند تخلفات کوچک (نقاط اشتباه دسته‌بندی‌شده) را تحمل کند. مقدار کم برای C به معنی یک مدل با حاشیه بزرگ‌تر و اشتباهات بیشتر است، در حالی که مقدار بزرگ C مدل را به سمت کمترین خطا هدایت می‌کند [۴۲].

۵-۲- جنگل تصادفی

جنگل تصادفی یکی از روش‌های تلفیقی در یادگیری ماشین است که به طور خاص برای دسته‌بندی و رگرسیون استفاده می‌شود. این روش شامل مجموعه‌ای از درخت‌های تصمیم (Decision Trees) است که به طور تصادفی و مستقل از یکدیگر آموزش داده می‌شوند. برای ساخت هر درخت تصمیم، زیرمجموعه‌ای تصادفی از داده‌های آموزشی انتخاب می‌شود. این فرایند به نام Bootstrap Aggregating (Bagging) شناخته می‌شود. در هر تقسیم‌گیری (split) درخت، از یک زیرمجموعه تصادفی از ویژگی‌ها (برخلاف استفاده از تمامی ویژگی‌ها) برای انتخاب بهترین تقسیم استفاده می‌شود. این کار به کاهش همبستگی بین درخت‌ها و افزایش تنوع آن‌ها کمک می‌کند. هر درخت تصمیم در جنگل، پیش‌بینی خود را برای داده‌های جدید انجام می‌دهد. در روش دسته‌بندی، نتایج پیش‌بینی شده توسط همه درخت‌ها جمع‌آوری می‌شود و رأی‌گیری اکثریت (Majority Voting) برای تعیین کلاس نهایی انجام می‌شود. استفاده از چندین

که در رابطه فوق A مساحت کل ناحیه مورد نظر برحسب مترمربع و n تعداد کل نمونه‌ها یا مشاهدات است. در این پروژه با توجه به مساحت تقریبی ۳ کیلومتر مربع و تعداد ۳۰۰ نمونه برداشت شده، اندازه سلول مناسب ۵ متر لحاظ شده است.

تمام نقشه‌های شاهد برای ایجاد ۴۴۷۰۸ بردار ویژگی ۱۱ بعدی باهم ترکیب شدند. از آنجاکه این عملیات در محیط GIS انجام شد، یک پایگاه داده به‌طور خودکار تولید شد که بردارهای ویژگی را ذخیره می‌کند.

از محدوده شادان تعداد ۲۷ گمانه حفر شده است که این گمانه‌ها در یک شبکه نامنظم طراحی شده‌اند. از این گمانه‌ها تعداد ۶۹۶۵ مغزه به‌دست آمده است که طول آن‌ها بین ۰.۴ متر تا ۷ متر و طول متوسط ۱ متر بوده است. مغزه‌ها برای ۴۶ عنصر با روش ICP-OES و برای طلا به روش Fire Assay آنالیز شده‌اند در بخش تحلیل داده های حفاری، جهت بررسی کانی‌سازی مس و طلا در هر گمانه، عیار مس و طلای هر مغزه در طول آن ضرب شد و سپس میانگین‌گیری در طول گمانه انجام شده است. با بی-مقیاس کردن مقادیر موجود، مقادیر مرتبط طلا و مس هر گمانه با یکدیگر جمع شده و یک عدد به گمانه نسبت داده شده است.

با توجه به اینکه طول گمانه‌ها متفاوت می‌باشند مقایسه میانگین‌های به‌دست‌آمده کمی با خطا روبروست. به عنوان مثال نمی‌توان میانگین به‌دست‌آمده از یک گمانه ۳۰۰ متری با میانگین به‌دست‌آمده از یک گمانه ۱۰۰ متری مقایسه کرد؛ بنابراین به منظور وزن دهی مقادیر به‌دست‌آمده با نرمالایز کردن طول گمانه‌ها، از این مقادیر به عنوان ضریب وزنی هر گمانه استفاده شد.

در ادامه نمودار فرکتالی مقادیر تجمعی تعداد نمونه‌ها (نمودار فرکتالی N-S) رسم گردیده است. شکل (۱۵) نشان‌دهنده نمودار تجمعی تعداد نمونه‌ها در مقابل عدد محاسبه شده می‌باشد. این نمودار دارای ۴ نقطه شکست می‌باشد که بر این اساس گمانه‌ها به پنج دسته تقسیم شدند. این دسته بندی‌ها را می‌توان به کیفیت کانی‌سازی در هر گمانه نسبت داد. به منظور تدقیق در مدلسازی کلاس ۱ با ۹ گمانه به عنوان کلاس مطلوب و ۱۸ گمانه دیگر به عنوان کلاس نامطلوب در نظر گرفته شدند.

قدرت تفکیک بیشتری بدهد، اما ممکن است به بیش برآزش منجر شود. پارامتر (max_features) تعداد ویژگی‌هایی است که در هر تقسیم‌گیری درخت تصمیم استفاده می‌شود. تنظیم مناسب این پارامتر می‌تواند تنوع درخت‌ها را افزایش دهد و عملکرد مدل را بهبود بخشد. پارامتر (min_samples_split) حداقل تعداد نمونه‌هایی که باید در هر گره داخلی وجود داشته باشد تا تقسیم کردن انجام شود. تنظیم مناسب این پارامتر می‌تواند به جلوگیری از بیش برآزش کمک کند. پارامتر (min_samples_leaf) حداقل تعداد نمونه‌هایی است که باید در هر برگ نهایی وجود داشته باشد. این پارامتر به کنترل اندازه برگ‌ها و جلوگیری از تقسیم کردن خیلی جزئی کمک می‌کند. پارامتر (splitter) نوع الگوریتمی است که برای تقسیم کردن گره‌ها استفاده می‌شود [۴۳].

برای انتخاب مقدار بهینه پارامترهای مرتبط با هر الگوریتم، از تکنیک‌های تست‌های متقابل (Cross-Validation) که در بخش قبلی معرفی شد، استفاده شده است.

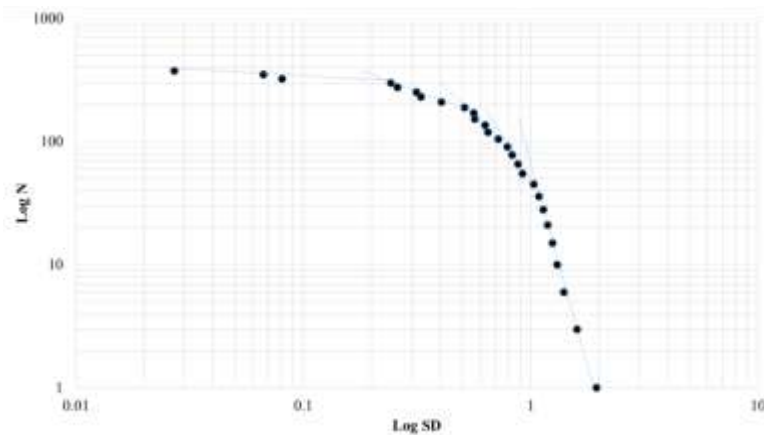
۵-۳- آماده‌سازی بردارهای ویژگی

به‌منظور اعمال الگوریتم‌های یادگیری ماشین و تهیه نقشه پتانسیل مس - طلا در محدوده شادان از ۱۱ لایه اطلاعاتی و نقشه شاهد از جمله واحدهای سنگی، تکتونیک، ژئومغناطیس، آلتراسیون‌ها و ژئوشیمی که در بخش‌های قبلی معرفی و آماده شدند، استفاده شده است. این لایه‌های اطلاعاتی حکم متغیرهای مستقل یا صفات خاصه یا ویژگی‌ها را در الگوریتم‌ها دارند.

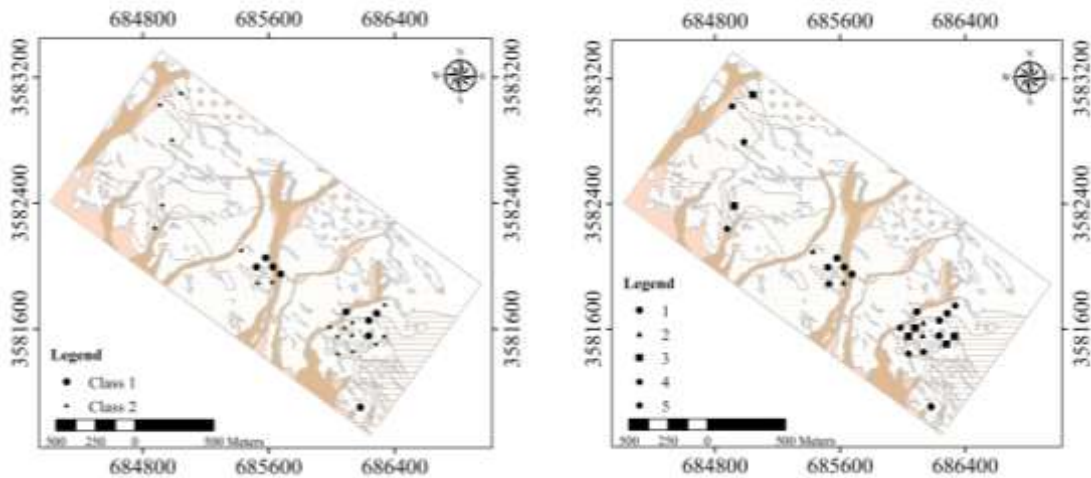
اکثر روش‌های GIS مبنا برای مدل‌سازی پتانسیل معدنی از مفهوم شبکه‌بندی واحد استفاده می‌کنند. در زمینه الگوریتم‌های یادگیری ماشین هر موقعیت، یک بردار ویژگی در نظر گرفته شده است که محتوای آن توسط ویژگی‌های نقشه‌های شاهد تعریف شده‌اند که دارای شرایط یکسانی هستند؛ بنابراین تعداد ابعاد بردارهای ویژگی برابر است با تعداد نقشه‌های شاهد می‌باشد [۴۴، ۷].

در این تحقیق اندازه سلول بهینه براساس داده‌های ژئوشیمی انتخاب شده است. جهت انتخاب اندازه سلول مناسب از رابطه پیشنهاد شده توسط هنگل [۴۵] استفاده شده است. این رابطه به‌صورت زیر تعریف می‌شود:

$$\text{Cell Size} = \sqrt{\frac{A}{n}} * 10^2 * 0.0005$$



شکل ۱۵- نمودار فرکتالی فراوانی تجمعی گمانه‌ها به منظور دسته‌بندی



شکل ۱۶- موقعیت گمانه‌ها به صورت کلاس‌بندی شده (پنج کلاسه و دو کلاسه)

کل مقادیر به صورت تصادفی به عنوان داده‌های آموزشی و ۳۰٪ باقی‌مانده به عنوان داده‌های آزمایشی برای بررسی صحت تابع به دست آمده و تعیین مقدار خطای کلاسه‌بندی استفاده شده است.

۶-۱- مدلسازی با استفاده از SVM

فرایند یادگیری در این مدل با استفاده از سه تابع کرنل خطی، RBF و چندجمله‌ای انجام شد. مدل با استفاده از داده‌های آموزشی تعلیم داده شد و از طبقه‌بندی طراحی شده برای طبقه‌بندی داده‌های آزمون استفاده شد. به منظور دستیابی به مولفه‌های بهینه مدل SVM با هر تابع کرنل از روش تست‌های متقابل استفاده شد. این فرایند به صورت تصادفی در انتخاب داده‌های آموزش و آزمون برای ده مرتبه انجام شد و طبقه‌بندی بهینه براساس معیار نرخ طبقه‌بندی صحیح برای داده‌های آزمون انتخاب شد (پارامترهای بهینه شده مدل، C برابر ۱۰ و گاما برابر با ۰/۰۱). سپس مدل ساخته شده برای تهیه مدل پتانسیل

۶- تولید مدل پتانسیل معدنی

از اطلاعات ۲۷ گمانه که طبقه‌بندی شده‌اند، جهت تعیین بردارهای آموزش و آزمایش استفاده شده است. یکی از اهداف این تحقیق این است که آیا محدوده نیاز به حفاری بیشتر دارد یا نه؛ بنابراین می‌توان از گمانه‌های قبلی برای تعیین بهترین مکان برای حفاری‌های اضافه استفاده کرد. در بخش قبلی گمانه‌ها به ۲ کلاس براساس عبار طلا و مس کلاس‌بندی شده‌اند. با توجه به موقعیت گمانه‌ها و پایگاه داده‌ای شکل‌گرفته از بردارهای ویژگی، داده‌ها براساس همپوشانی با گمانه‌ها برچسب‌گذاری شده‌اند. بردارهای ویژگی طبق دسته‌بندی گمانه‌ها، ۱ و ۲ برچسب‌گذاری شده‌اند (شکل ۱۶).

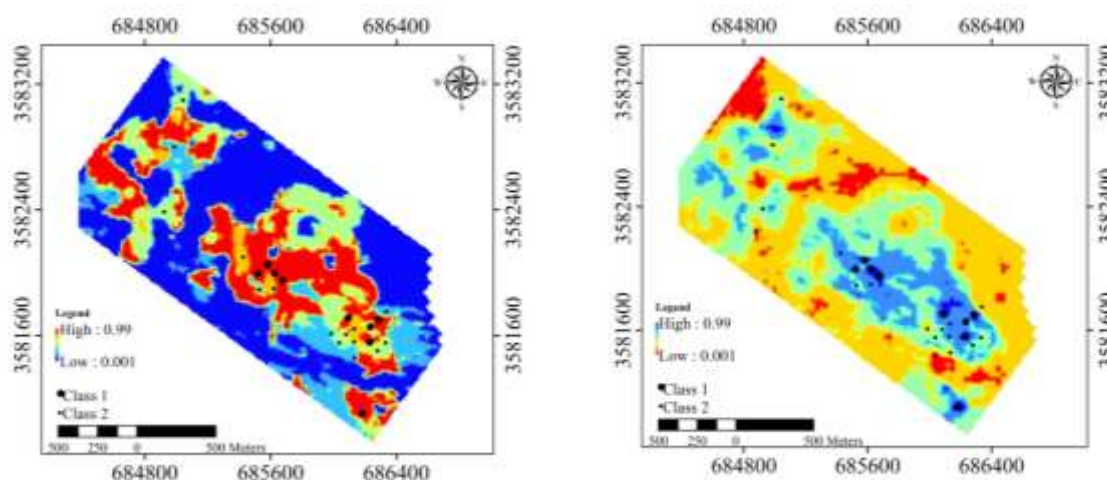
از میان ۴۴۷۰۸ بردار ویژگی، ۵۳۵ بردار ویژگی برچسب‌گذاری شده است. به منظور آموزش بهینه مدل‌ها از این تعداد داده یک مجموعه شامل ۳۷۵ داده یعنی ۰/۷٪

تست‌های متقابل با ۱۰ تکرار استفاده شد (پارامترهای بهینه شده مدل، تعداد درخت ۱۰۰۰ و تعداد نمونه در هر گره ۵). پس از ایجاد مدل و ارزیابی آن، سایر مقادیر بردارهای ویژگی با استفاده از این مدل کلاسه‌بندی شده است که شکل ۱۷ (۱۷) نقشه کلاسه‌بندی براساس الگوریتم جنگل تصادفی را نشان می‌دهد. همچنین در شکل ۲۳ (۲۳) اهمیت ویژگی‌های استفاده شده در این تحقیق را آورده شده است. کلیه فرآیند مدل‌سازی و پیش‌بینی به زبان برنامه‌نویسی پایتون و در بستر گوگل کولب انجام گرفته شد.

معدنی برای بردارهای ویژگی بدون برچسب اجرا شد. مدل اکتشافی پتانسیل مس و طلای پورفیری تولید شده با الگوریتم SVM در شکل ۱۷ (۱۷) نشان داده شده است. در این مدل اهمیت اکتشافی برای هر پیکسل به صورت درجه عضویت فازی نشان داده شده است که مقادیر نزدیک به یک نشان‌دهنده احتمال بیشتر کانی‌سازی مس و طلای پورفیری در محدوده است.

۲-۶- مدل‌سازی با استفاده از RF

مدل با استفاده از داده‌های آموزشی تعلیم داده شد و از طبقه‌بندی طراحی شده برای طبقه‌بندی داده‌های آزمون استفاده شد. به منظور بهینه‌سازی پارامترهای موثر از روش



شکل ۱۷- نقشه پتانسیل معدنی مس-طلای پورفیری محدوده شادان با بکارگیری A: روش SVM و B: روش RF

جدول ۳- شاخص‌های ارزیابی مدل‌ها

	SVM	RF
Accuracy	0.92	0.88
Sensitivity	0.88	0.79
AUC	0.93	0.88

۶- ارزیابی و مقایسه مدل‌ها

خروجی مطلوب در مدل‌سازی پتانسیل معدنی باید به‌طور دقیق نقاط با اهمیت بالای اکتشافی در منطقه مورد مطالعه را شناسایی کند. این بدان معناست که نواحی پیشنهادی باید بیشترین همبستگی را با اندیس‌های شناخته‌شده و کمترین ارتباط را با مناطق فاقد کانی‌سازی داشته باشند. برای ارزیابی عملکرد مدل به کار گرفته شده در این پژوهش، از ماتریس درهم ریختگی و شاخص‌های دقت و حساسیت استفاده شده است. ماتریس درهم ریختگی و نتایج شاخص‌های ارزیابی دقت در شکل ۱۸ (۱۸) و شکل

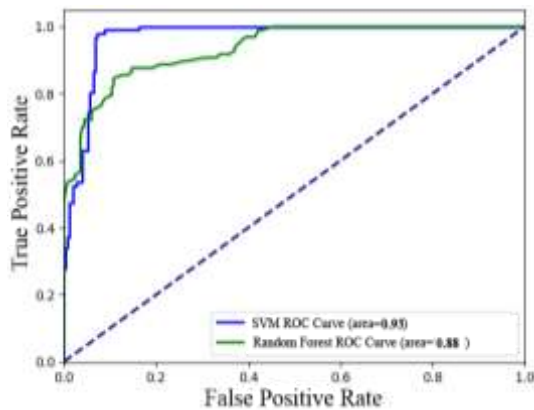
		Actual value	
		positive	negative
Predicted value	positive	95	26
	negative	18	236

شکل ۱۸- ماتریس درهم ریختگی برای روش RF

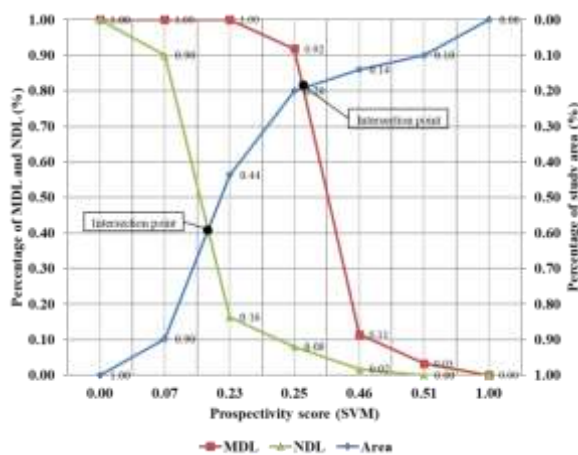
		Actual value	
		positive	negative
Predicted value	positive	106	15
	negative	14	240

شکل ۱۹- ماتریس درهم ریختگی برای روش SVM

منفی را از یکدیگر تفکیک کنند. با این حال، مدل SVM با نرخ صحت دسته‌بندی 0.92 در طبقه‌بندی داده‌های آموزشی و حساسیت 0.88 در شناسایی نواحی مطلوب، دقت بهتری نسبت به مدل RF نشان می‌دهد.



شکل ۲۰- نمودار مشخصه عملکرد سیستم (ROC) برای مدل های پتانسیل تولید شده در محدوده شادان



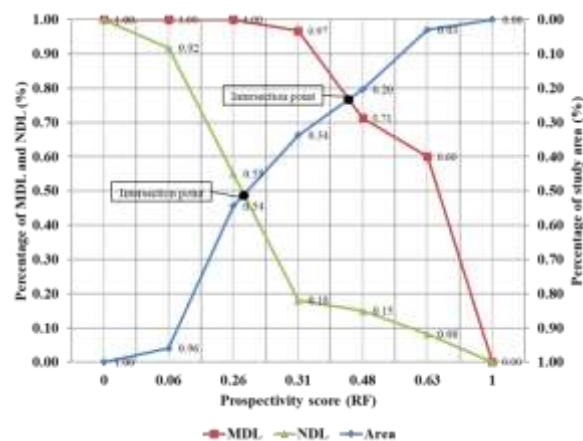
شکل ۲۱- نمودار میزان تغییرات پیش بینی - مساحت برای مدل های پتانسیل تولید شده در محدوده شادان به روش SVM:A و B: RF

موفق تر عمل کرده است. برای شناسایی مناطق هدف معدنی و برنامه‌ریزی دقیق عملیات اکتشافی، نیاز به طبقه‌بندی نقشه‌های پیوسته پتانسیل معدنی وجود دارد. در این پژوهش، از روش فرکتال عیار-مساحت برای تفکیک نواحی آنومالی از زمینه در خروجی مدل‌های مختلف بهره گرفته شده است. نتایج حاصل از این پژوهش نشان می‌دهد که هر دو روش موفق به طبقه‌بندی دقیق تمامی اندیس‌های شناخته‌شده در منطقه به‌عنوان نواحی آنومالی شده‌اند (شکل ۲۲). با این حال، تفاوت‌هایی در مساحت نواحی آنومالی شناسایی شده

(۱۹) و جدول ۳ Error! Reference source not found. ارائه شده است.

همچنین با در نظر گرفتن کلاس ۱ به عنوان کلاس مطلوب، از نمودار منحنی مشخصه عملکرد سیستم و مساحت زیر منحنی مشخصه عملکرد برای ارزیابی مدل‌ها استفاده شده است. بدین منظور از موقعیت نقاط با برچسب مطلوب و نامطلوب برای رسم نمودار منحنی مشخصه عملکرد استفاده شده است. نمودار مشخصه عملکرد سیستم برای هر دو مدل در شکل ۲۰ (۲۰) آورده شده است.

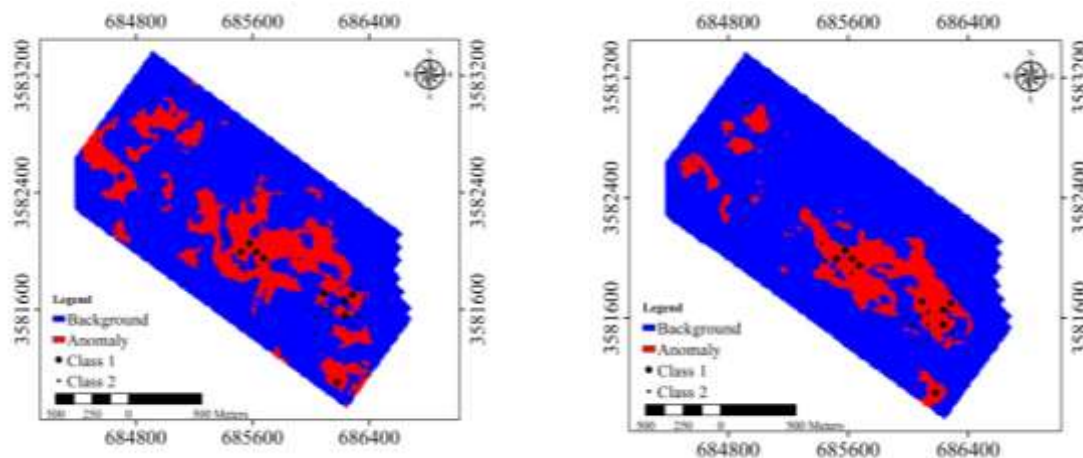
براساس ارزیابی معیارهای دقت (Accuracy)، حساسیت (Sensitivity) و مساحت زیر منحنی (ROC AUC)، نتایج نشان می‌دهند که هر دو مدل SVM و RF عملکرد مطلوبی در پیش‌بینی پتانسیل معدنی داشته‌اند. مقدار AUC برابر با 0.93 برای SVM و 0.88 برای RF نشان می‌دهد که هر دو مدل به خوبی توانسته‌اند نقاط مثبت و



براساس تحلیل نمودار میزان تغییرات پیش‌بینی - مساحت بهبود یافته (P-A) (شکل ۲۱)، عملکرد مدل‌های RF و SVM مقایسه شده‌اند. در مدل RF، نرخ پیش‌بینی نقاط مثبت برابر با 0.78 و نرخ پیش‌بینی نقاط منفی 0.49 است، در حالی که این مقادیر برای مدل SVM به ترتیب 0.81 و 0.41 هستند. با کسر این مقادیر طبق روش P-A، عملکرد نهایی مدل RF برابر با 0.29 و برای مدل SVM برابر با 0.40 محاسبه می‌شود. این تفاوت نشان می‌دهد که مدل SVM با دقت بالاتر و عملکرد بهتر نسبت به مدل RF، در شناسایی اهداف معدنی مطلوب در منطقه مورد بررسی

این تفاوت نشان‌دهنده دقت و کارایی بیشتر روش SVM در جداسازی نواحی آنومالی با مساحت کمتر و دقت بالاتر است.

وجود دارد؛ به طوری که مساحت نواحی آنومالی شناسایی شده توسط روش SVM معادل ۱۳ درصد از کل منطقه مورد بررسی بوده است، در حالی که این مقدار برای روش SVM حدود ۱۸ درصد از منطقه را شامل می‌شود.



شکل ۲۲- نمایش دو کلاسه نقشه پتانسیل معدنی مس طلای پورفیری در محدوده شادان، SVM:A و RF :B

نمی‌تواند تعمیم‌پذیری لازم را برای داده‌های جدید داشته باشد. کم‌برازش نیز زمانی به وقوع می‌پیوندد که مدل به اندازه کافی نتواند از داده‌ها الگو استخراج کند و عملکرد ضعیفی بر روی هر دو مجموعه داده آموزشی و آزمایشی ارائه دهد. با این حال، نتایج به دست آمده در این تحقیق نشان می‌دهد که الگوریتم‌های ماشین بردار پشتیبانی (SVM) و جنگل تصادفی (RF) در مواجهه با شرایط عدم توازن داده‌ها (تعداد نمونه‌های مربوط به کلاس ۲ بیشتر از کلاس ۱ است)، از نظر کارایی و دقت عملکرد مناسبی دارند. این دو روش به دلیل مکانیزم‌های داخلی خود، از جمله استفاده از تابع کرنل در SVM و ترکیب درخت‌های تصمیم‌گیری در RF، قادر به مدیریت مناسب این عدم توازن هستند و به خوبی می‌توانند به تعادل میان کلاس‌ها دست یابند [۴۶]. در مطالعه نقشه‌برداری پتانسیل معدنی ذخایر مس-طلای پورفیری با استفاده از روش جنگل تصادفی (Random Forest)، اهمیت پارامترهای ورودی مورد ارزیابی قرار گرفت. شکل ۲۳ (۲۳) نشان می‌دهد که فاکتور ۱ که از تحلیل فاکتوری مقاوم (Robust Factor Analysis) و شامل عناصر کلیدی مانند مس (Cu)، طلا (Au) و مولیبدن (Mo) است، به عنوان مهم‌ترین عامل در مدل‌سازی پتانسیل معدنی شناسایی شده است. پس از فاکتور ۱، نقشه پراکندگی عنصر مس (Cu anomaly map) و سپس نقشه

۷- بحث

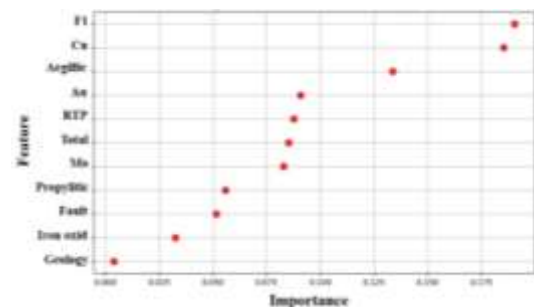
در این پژوهش، برای شناسایی اهداف اکتشافی، از روش‌های نظارت‌شده ماشین بردار پشتیبانی (SVM) و جنگل تصادفی (RF) استفاده شده است. این روش‌ها از داده‌های دارای برچسب اکتشافی برای طبقه‌بندی نواحی با پتانسیل معدنی استفاده می‌کنند. در روش SVM، مدل با استفاده از داده‌های ورودی برچسب‌دار، الگوهای مرتبط با کانی‌سازی را شناسایی کرده و نواحی مستعد برای اکتشاف را پیش‌بینی می‌کند. روش RF نیز با ایجاد مجموعه‌ای از درخت‌های تصمیم‌گیری، شناسایی نواحی معدنی با استفاده از داده‌های اکتشافی را انجام داده است.

در مواردی که داده‌های برچسب‌دار کافی برای مدل‌سازی وجود دارد، روش‌های نظارت‌شده مانند SVM و RF به عنوان ابزارهای قدرتمند برای شناسایی نواحی مستعد کانی‌سازی مورد استفاده قرار می‌گیرند. در این پژوهش، با توجه به دسترسی به داده‌های اکتشافی کافی، استفاده از روش‌های نظارت‌شده SVM و RF به ما این امکان را داده است تا با دقت بالاتری نواحی با پتانسیل بالای معدنی را شناسایی و طبقه‌بندی کنیم. در مسائل مدل‌سازی که با عدم توازن داده‌ها مواجه هستند، احتمال وقوع بیش‌برازش یا کم‌برازش وجود دارد. بیش‌برازش زمانی رخ می‌دهد که مدل به شکلی بیش از حد به داده‌های آموزشی تطبیق می‌یابد و

طراحی شده است که کلاس‌ها را در فضای ویژگی‌ها با حداکثر حاشیه از هم جدا می‌کند. این خاصیت در شرایطی که داده‌ها به خوبی تفکیک پذیر هستند یا مرزهای دقیقی بین نواحی مختلف (مانند نواحی کانی‌سازی و غیرکانی‌سازی) وجود دارد، بسیار مؤثر است. در مقابل، RF با ایجاد چندین درخت تصمیم، مرزهای احتمالی بیشتری را ایجاد می‌کند که ممکن است دقت تفکیک آن را کاهش دهد. SVM به دلیل استفاده از توابع کرنل (kernel functions) می‌تواند به صورت مؤثر در فضای ویژگی‌های با ابعاد بالا عمل کند. در پژوهش‌های معدنی، ممکن است تعداد زیادی ویژگی (مانند داده‌های ژئوفیزیکی، ژئوشیمیایی و غیره) وجود داشته باشد که SVM به دلیل قابلیت کار با داده‌های چندبعدی، توانایی بالاتری در تشخیص الگوهای پنهان و مرزهای دقیق بین نواحی مختلف دارد. در مقابل، RF نیاز به تعداد زیادی درخت برای پردازش چنین داده‌هایی دارد که ممکن است به پیچیدگی و کاهش کارایی منجر شود. SVM با استفاده از کرنل‌های مختلف (مانند کرنل‌های چندجمله‌ای یا گاوسی) قادر است مرزهای غیرخطی پیچیده‌ای را بین کلاس‌ها شناسایی کند. این ویژگی به آن امکان می‌دهد تا به طور دقیق نواحی معدنی با ویژگی‌های پیچیده را شناسایی کند. از آنجا که در مسائل اکتشافی، روابط غیرخطی بین ویژگی‌های زمین‌شناسی وجود دارد، SVM می‌تواند این پیچیدگی‌ها را بهتر مدل‌سازی کند. در حالی که RF در برخورد با این نوع روابط، با استفاده از ترکیب درخت‌های تصمیم سعی می‌کند مرزها را شناسایی کند، اما ممکن است دقت آن در تفکیک دقیق مرزهای غیرخطی پایین‌تر باشد. اگر داده‌های مورد استفاده نویز کمی داشته باشند یا به خوبی برچسب‌گذاری شده باشند، SVM به دلیل تمرکز بر حداکثرسازی حاشیه بین کلاس‌ها، می‌تواند بهتر از RF که به صورت تصادفی درخت‌های تصمیم‌گیری می‌سازد، عمل کند. SVM با این ساختار بهینه‌سازی، حساسیت بیشتری به الگوهای دقیق‌تر دارد [۴۷-۵۰].

به منظور بررسی نتایج حاصل از SVM و تحلیل نقشه پتانسیل مس-طلاهی تهیه شده، این نتایج با اطلاعات لایه-های اطلاعاتی مقایسه شده است (شکل ۲۴).

آلتراسیون آرژیلیک (Argillic alteration map) به ترتیب بیشترین تأثیر را در مدل‌سازی پتانسیل معدنی داشته‌اند. در مقابل، واحدهای زمین‌شناسی (Geological units)، با وجود حضور در مدل، کمترین نقش را در پیش‌بینی مناطق با پتانسیل معدنی بالا ایفا کرده‌اند. دلیل کم‌اهمیت بودن واحدهای زمین‌شناسی در مدل‌سازی پتانسیل معدنی مس-طلاهی پورفیری در محدوده شادان این است که بیشتر واحدهای سنگی موجود در منطقه مورد مطالعه دارای ترکیب سنگ‌شناسی مشابه هستند. این یکنواختی زمین‌شناسی باعث کاهش تأثیر این عامل در تمایز نواحی دارای پتانسیل بالای معدنی می‌شود. در مقابل، پارامترهای ژئوشیمیایی و آلتراسیون که نشان‌دهنده تغییرات موضعی در شرایط زمین‌ساختی و کانه‌زایی هستند، توانسته‌اند تفاوت‌های بیشتری در مدل‌سازی ارائه دهند و اهمیت بیشتری پیدا کنند.



شکل ۲۳- اهمیت متغیرهای مورد بررسی به دست آمده از الگوریتم جنگل تصادفی

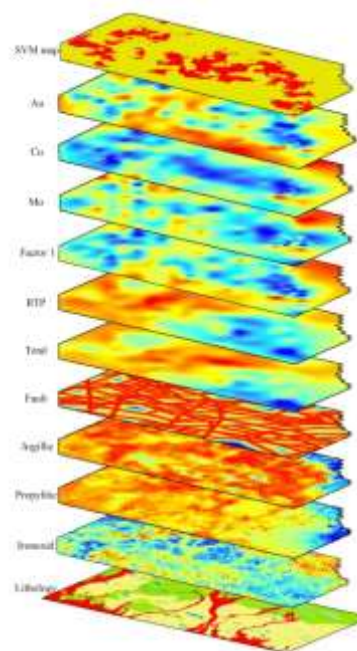
مقایسه مدل‌های پتانسیل تولید شده توسط معیارهای ارزیابی نشان دهنده عملکرد بهتر روش SVM در شناسایی اهداف اکتشافی دارد. مطالعات مشابهی با استفاده از روش SVM و RF به منظور مدل‌سازی نقشه پتانسیل معدنی انجام گرفته شده است که به برتری روش SVM در دقت اشاره کرده‌اند. یکی از این مطالعات که توسط کاردوسو و همکاران ۲۰۱۹ انجام شده است، مقایسه‌ای بین عملکرد SVM و RF در مدل‌سازی پتانسیل معدنی ارائه کرده است [۱۸]. مطالعه دیگری توسط کارانزا و همکاران نیز RF را به عنوان روشی کارآمد معرفی کرده‌اند، اما در برخی شرایط پیچیده زمین‌شناسی، SVM به دلیل تفکیک دقیق‌تر مرزهای کانی‌سازی، دقت بیشتری داشته است [۲۲].

SVM به طور خاص برای یافتن بهترین مرز (hyperplane)

با توجه به اینکه این نتایج با مدل مفهومی کانی‌سازی مس - طلای پورفیری و شرایط زمین‌شناختی مطلوب برای این نوع کانی‌سازی [۲۶، ۲۷، ۴۲] کاملاً همخوانی دارند، پیشنهاد می‌شود که برنامه‌های اکتشافی بعدی در مناطق مستعد شناسایی شده متمرکز شوند.

۸- نتیجه‌گیری

با توجه به مطالعات زمین‌شناسی انجام شده در محدوده شادان، وجود توده‌های نفوذی دیوریتی تا گرانودیوریتی با بافت پورفیری، به همراه نوع و وسعت آلتراسیون‌های مشاهده شده (شامل دگرسانی آرژیلیک، پروپلیتیک، فیلیک و حضور اکسیدهای آهن) و نوع کانی‌سازی شامل استوک ورک و افشان، همچنین برش‌های هیدروترمال و بافت جعبه‌ای ناشی از اکسیداسیون پیریت، تراکم بالای رگچه‌های سیلیسی-سولفیدی و ناهنجاری‌های قابل توجه مس و طلا، به طور کلی وجود کانی‌سازی مس-طلای پورفیری را در منطقه شادان تبیین می‌کند. در این پژوهش، اهداف اکتشافی مس و طلای پورفیری در محدوده شادان با استفاده از مدل مفهومی کانسارهای مس و طلای پورفیری و به کارگیری دو روش یادگیری ماشین، یعنی ماشین بردار پشتیبان (SVM) و جنگل تصادفی (RF) شناسایی شده‌اند. دلیل استفاده از روش نظارت شده نیز وجود ۲۷ حفاری انجام شده در محدوده است که این امکان را فراهم کرده است. در این پژوهش، اهداف اکتشافی مس و طلای پورفیری در محدوده شادان با استفاده از مدل مفهومی کانسارهای مس و طلای پورفیری و به کارگیری دو روش یادگیری ماشین، یعنی ماشین بردار پشتیبان (SVM) و جنگل تصادفی (RF) شناسایی شده‌اند. به منظور ارزیابی عملکرد مدل‌های پتانسیل تولید شده، از نمودار مشخصه عملکرد سیستم (ROC) استفاده شده است. نتایج نشان می‌دهند که مدل SVM با مقدار AUC برابر ۰.۹۳ عملکرد بهتری نسبت به روش RF داشته و توانسته است تمامی حفاری کلاس مطلوب را در ۱۹ درصد مساحت منطقه مورد بررسی به درستی توجیه کند. بنابراین، می‌توان نتیجه‌گیری کرد که مدل پتانسیل تولید شده با استفاده از روش SVM به عنوان ابزار قابل‌اعتمادتری برای هدایت فعالیت‌های اکتشافی نقاط مطلوب مس و طلای پورفیری در منطقه مورد بررسی عمل می‌کند؛ بنابراین پیشنهاد می‌شود که برنامه‌ریزی جهت عملیات اکتشافی بعدی با استفاده از یک



شکل ۲۴- نقشه پتانسیل معدنی حاصل از روش SVM و لایه-های اطلاعاتی استفاده شده

نتایج حاصل از مدل‌سازی مناطق مطلوب با استفاده از روش SVM و تحلیل داده‌های ژئوشیمیایی نشان می‌دهد که مناطق با مطلوبیت بالا لزوماً همبستگی مستقیم با عیارهای پایین فلزاتی چون مس، طلا و مولیبدن ندارند. بخش قابل توجهی از مقادیر طلا، مس و مولیبدن موجود در کلاس‌های مطلوب در محدوده‌هایی با آنومالی‌های با شدت بالا قرار گرفته‌اند. در تحلیل‌های زمین‌شناسی، بیشتر واحدهای سنگی شناسایی شده در این مناطق شامل آندزیت، ریولیت، گرانیت و شادان پورفیری بوده و تنها بخش کوچکی از توف و رسوبات کواترنری در این دسته‌بندی جای می‌گیرند. این نشان‌دهنده نقش قابل توجه سنگ‌های آذرین در پیدایش مناطق مطلوب است. از لحاظ آلتراسیون، مطابق انتظار، سطح بالایی از آلتراسیون‌های تشخیص داده شده توسط تصاویر ASTER ارتباط نزدیکی با مناطق مطلوب نشان می‌دهد. در این میان، آلتراسیون‌های آرژیلیک و فیلیک بیشترین همپوشانی را با این مناطق دارند که به‌ویژه در سامانه‌های پورفیری بسیار مورد توجه قرار می‌گیرند. بررسی لایه‌های گسلی نشان می‌دهد که مناطق مطلوب به شدت با نواحی دارای تکتونیزه بالا مرتبط هستند که نشان‌دهنده نقش فعال شکستگی‌ها و گسل‌ها در فرآیندهای کانی‌سازی است. بررسی داده‌های ژئومغناطیسی نیز نشان می‌دهد که احتمال وقوع کانی‌سازی مس و طلا در این مناطق کاملاً با آنومالی‌های مغناطیسی مشاهده شده مطابقت دارد.

تأییدیه اخلاقی

نویسندگان متعهد می‌شوند که مطالب ارائه‌شده در این مقاله حاصل کار پژوهشی آنان بوده و پیش‌تر در هیچ مجله، همایش یا نشریه دیگری به چاپ نرسیده یا هم‌زمان برای انتشار به مجله دیگری ارسال نشده است.

مشارکت‌های نویسندگان

حسن حسین‌زاده: تحلیل، تفسیر نتایج و نگارش مقاله
معصومه دادپور: تحلیل، تفسیر نتایج و نگارش مقاله
آرش گورابجیری پور: تهیه داده‌ها و مشارکت در نظارت بر فرایند تحلیل‌ها

غلامرضا نوروزی: نظارت عالی بر کل فرایند پژوهش، شامل تهیه داده‌ها، کنترل روند تحلیل‌ها، تفسیر نتایج و نگارش نهایی مقاله
 همه نویسندگان نسخه نهایی مقاله را مطالعه و تأیید کرده‌اند.

منابع مالی

این پژوهش بدون دریافت هرگونه حمایت مالی از سازمان‌ها، مؤسسات دولتی یا خصوصی، شرکت‌ها یا نهادهای تأمین‌کننده مالی انجام شده است.

شبکه حفاری منظم در مناطق مستعد شناسایی شده توسط مدل SVM صورت گیرد. اولویت باید به انجام عملیات حفاری در بخش شمال‌غربی محدوده اختصاص یابد، زیرا تاکنون در این ناحیه هیچ‌گونه حفاری انجام نشده است. به‌منظور بهینه‌سازی الگوریتم‌های یادگیری ماشین و بهبود نقشه پتانسیل معدنی به‌دست‌آمده، همچنین برای طراحی مدل پویای اکتشافی، پیشنهاد می‌شود که پس از انجام هر فعالیت اکتشافی در محدوده، نتایج به‌دست‌آمده برای آموزش مجدد الگوریتم‌ها مورد استفاده قرار گیرد. این فرآیند به بهبود مستمر دقت مدل‌ها و ارتقاء کیفیت پیش‌بینی‌ها کمک خواهد کرد.

قدردانی

نویسندگان از شرکت معادن و صنایع معدنی کارند صدر جهان برای در اختیار گذاشتن داده‌های اکتشافی قدردانی می‌نمایند.

تعارض منافع

نویسندگان اعلام می‌کنند که در خصوص انتشار این مقاله هیچ‌گونه تعارض منافع مالی، مشاوره‌ای، نهادی، شخصی یا سایر روابطی که بتواند بر نتایج، تفسیرها یا فرایند انتشار مقاله اثرگذار باشد، وجود ندارد.

مراجع

- [1] Sun, Tao, Fei Chen, Lianxiang Zhong, Weiming Liu, and Yun Wang. "GIS-based mineral prospectivity mapping using machine learning methods: A case study from Tongling ore district, eastern China." *Ore Geology Reviews* 109 (2019): 26-49.
- [2] Liu, Haiming, Jeff Harris, Ross Sherlock, Poursan Behnia, Eric Grunsky, Mostafa Naghizadeh, Kate Rubingh, Györgyi Tuba, Eric Roots, and Graham Hill. "Mineral prospectivity mapping using machine learning techniques for gold exploration in the Larder Lake area, Ontario, Canada." *Journal of Geochemical Exploration* 253 (2023): 107279.
- [3] Amponsah, Prince Ofori, and Eric Dominic Forson. "Geospatial modelling of mineral potential zones using data-driven based weighting factor and statistical index techniques." *Journal of African Earth Sciences* 206 (2023): 105020.
- [4] Geranian, Hamid, and Malieh Nakhaei. "Application of multivariate decision-making algorithms in the mineral potential mapping; Case study: West Basiran, South Khorasan Province." vol 1 (2023): 13-21.
- [5] Roshanravan, Bijan, Oliver P Kreuzer, Amanda Buckingham, Majid Keykhay-Hosseinpour, and Edward Keys. "Mineral potential modelling of orogenic gold systems in the granites-tanami Orogen, Northern Territory, Australia: A multi-technique approach." *Ore Geology Reviews* 152 (2023): 105224.
- [6] Keykhay-Hosseinpour, Majid, Alok Porwal, and Kalimuthu Rajendran. "Targeting porphyry Cu deposits in the Chahargonbad region of Iran: A joint application of deep belief networks and random forest techniques." *Geochemistry* (2024): 126155.
- [7] Carranza, EJM, T Woldai, and EM Chikambwe. "Application of data-driven evidential belief functions to prospectivity mapping for aquamarine-bearing pegmatites, Lundazi district, Zambia." *Natural Resources Research* 14 (2005): 47-63.

- [8] Peng, Qingming, Zhongzheng Wang, Gongwen Wang, Wengao Zhang, Zhengle Chen, and Xiaoning Liu. "3D Mineral Prospectivity Mapping from 3D Geological Models Using Return-Risk Analysis and Machine Learning on Imbalance Data." *Minerals* 13, no. 11 (2023): 1384.
- [9] Keykhay-Hosseinpoor, Majid, Amir-Hossein Kohsary, Amin Hossein-Morshedy, and Alok Porwal. "A machine learning-based approach to exploration targeting of porphyry Cu-Au deposits in the Dehsalm district, eastern Iran." *Ore Geology Reviews* 116 (2020): 103234.
- [10] Senanayake, Indishe P, Anthony S Kiem, Gregory R Hancock, Václav Metelka, Chris B Folkes, Phillip L Blevin, and Anthony R Budd. "A spatial data-driven approach for mineral prospectivity mapping." *Remote Sensing* 15, no. 16 (2023): 4074.
- [11] Carranza, Emmanuel John Muico. *Geochemical Anomaly and Mineral Prospectivity Mapping in GIS*. Elsevier, 2008.
- [12] Lary, David J, Amir H Alavi, Amir H Gandomi, and Annette L Walker. "Machine learning in geosciences and remote sensing." *Geoscience Frontiers* 7, no. 1 (2016): 3-10.
- [13] Zhang, Zhiqiang, Gongwen Wang, Emmanuel John M Carranza, Jingguo Du, Yingjie Li, Xinxing Liu, and Yongjun Su. "An Uncertainty-Quantification Machine Learning Framework for Data-Driven Three-Dimensional Mineral Prospectivity Mapping." *Natural Resources Research* (2024): 1-19.
- [14] Mitchell, Tom M, and Tom M Mitchell. *Machine learning*. Vol. 1: McGraw-hill New York, 1997.
- [15] Granek, Justin. "Application of machine learning algorithms to mineral prospectivity mapping." University of British Columbia, 2016.
- [16] Lachaud, Alix, Marcus Adam, and Ilija Mišković. "Comparative study of random forest and support vector machine algorithms in mineral prospectivity mapping with limited training data." *Minerals* 13, no. 8 (2023): 1073.
- [17] Zheng, Chaojie, Feng Yuan, Xianrong Luo, Xiaohui Li, Panfeng Liu, Meilan Wen, Zesu Chen, and Stefano Albanese. "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* (2023): 105567.
- [18] Cardoso-Fernandes, J, AC Teodoro, Alexandre Lima, and E Roda-Robles. 2019. "Evaluating the performance of support vector machines (SVMs) and random forest (RF) in Li-pegmatite mapping: Preliminary results." *Earth Resources and Environmental Remote Sensing/GIS Applications X*.
- [19] McKay, G, and JR Harris. "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." *Natural Resources Research* 25, no. 2 (2016): 125-143.
- [20] Rodriguez-Galiano, Victor, Manuel Sanchez-Castillo, M Chica-Olmo, and MJOGR Chica-Rivas. "Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines." *Ore Geology Reviews* 71 (2015): 804-818.
- [21] Harris, JR, E Grunsky, P Behnia, and D Corrigan. "Data-and knowledge-driven mineral prospectivity maps for Canada's North." *Ore Geology Reviews* 71 (2015): 788-803.
- [22] Carranza, Emmanuel John M, and Alice G Laborte. "Data-driven predictive mapping of gold prospectivity, Baguio district, Philippines: Application of Random Forests algorithm." *Ore Geology Reviews* 71 (2015): 777-787.
- [23] Rodriguez-Galiano, VF, Mario Chica-Olmo, and Mario 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." *International Journal of Geographical Information Science* 28, no. 7 (2014): 1336-1354.
- [24] Abedi, Maysam, Gholam-Hossain Norouzi, and Abbas Bahroudi. "Support vector machine for multi-classification of mineral prospectivity areas." *Computers & Geosciences* 46 (2012): 272-283.
- [25] Zuo, Renguang, and Emmanuel John M Carranza. "Support vector machine: A tool for mapping mineral prospectivity." *Computers & Geosciences* 37, no. 12 (2011): 1967-1975.
- [26] Sillitoe, Richard H. "Porphyry copper systems." *Economic geology* 105, no. 1 (2010): 3-41.
- [27] Sinclair, WD. "Porphyry deposits." *Mineral deposits of Canada: A synthesis of major deposit-types, district metallogeny, the evolution of geological provinces, and exploration methods: Geological Association of Canada, Mineral Deposits Division, Special Publication 5* (2007): 223-243.

- [28] Singer, Donald A, Vladimir I Berger, and Barry C Moring. "Porphyry copper deposits of the world: Database and grade and tonnage models, 2008." US Geological Survey, 2008.
- [29] Cooke, David R, Peter Hollings, and John L Walshe. "Giant porphyry deposits: characteristics, distribution, and tectonic controls." *Economic Geology* 100, no. 5 (2005): 801-818.
- [30] Eppinger, Robert G, Karen D Kelley, David L Fey, Stuart A Giles, Burke J Minsley, and Steven M Smith. USGS Exploration Geochemistry Studies at the Pebble Porphyry Cu-Au-Mo Deposit, Alaska: Pdf of Presentation. US Department of the Interior, Geological Survey, 2010.
- [31] Behn, Gerardo, Francisco Camus, Pedro Carrasco, and Hunter Ware. "Aeromagnetic signature of porphyry copper systems in northern Chile and its geologic implications." *Economic Geology* 96, no. 2 (2001): 239-248.
- [32] Fatehi, Moslem, and Hooshang H Asadi. "Data integration modeling applied to drill hole planning through semi-supervised learning: A case study from the Dalli Cu-Au porphyry deposit in the central Iran." *Journal of African Earth Sciences* 128 (2017): 147-160.
- [33] Mining, What Is Data. Introduction to data mining. Springer, 2006.
- [34] Jiawei, Han, and Kamber Micheline. Data mining: concepts and techniques. Morgan kaufmann, 2006.
- [35] Porwal, Alok Kumar. "Mineral potential mapping with mathematical geological models." (2006).
- [36] Bonham-Carter, Graeme. Geographic information systems for geoscientists: modelling with GIS. Elsevier, 1994.
- [37] Kanevski, Mikhail, Vadim Timonin, and Alexi Pozdnukhov. Machine learning for spatial environmental data: theory, applications, and software. EPFL press, 2009.
- [38] Chen, Yongliang, and Wei Wu. "A prospecting cost-benefit strategy for mineral potential mapping based on ROC curve analysis." *Ore Geology Reviews* 74 (2016): 26-38.
- [39] Yousefi, Mahyar, and Emmanuel John M Carranza. "Prediction–area (P–A) plot and C–A fractal analysis to classify and evaluate evidential maps for mineral prospectivity modeling." *Computers & Geosciences* 79 (2015): 69-81.
- [40] Roshanravan, Bijan, Hamid Aghajani, Mahyar Yousefi, and Oliver Kreuzer. "An improved prediction-area plot for prospectivity analysis of mineral deposits." *Natural Resources Research* 28 (2019): 1089-1105.
- [41] Keykhay-Hosseinpour, Majid, Amir Hossein Kouhsari, Amin Hossein Morshedy, and Alok Porwal. "Porphyry Cu-Au prospectivity modelling using semi-supervised learning algorithm in Dehsalm district, eastern Iran." *Journal of Economic Geology* 13, no. 1 (2021): 193-213.
- [42] Stitson, MO, JAE Weston, A Gammernan, V Vovk, and V Vapnik. "Theory of support vector machines." *University of London* 117, no. 827 (1996): 188-191.
- [43] Breiman, Leo. "Random forests." *Machine Learning* 45 (2001): 5-32.
- [44] Porwal, Alok, EJM Carranza, and M Hale. "A hybrid neuro-fuzzy model for mineral potential mapping." *Mathematical Geology* 36, no. 7 (2004): 803-826.
- [45] Hengl, Tomislav. "Finding the right pixel size." *Computers & Geosciences* 32, no. 9 (2006): 1283-1298.
- [46] Fernández, Alberto, Salvador García, Mikel Galar, Ronaldo C Prati, Bartosz Krawczyk, and Francisco Herrera. *Learning from Imbalanced Data Sets*. Vol. 10: Springer, 2018.
- [47] Jakkula, Vikramaditya. "Tutorial on support vector machine (svm)." *School of EECS, Washington State University* 37, no. 2.5 (2006): 3.
- [48] Yue, Shihong, Ping Li, and Peiyi Hao. "SVM classification: Its contents and challenges." *Applied Mathematics-A Journal of Chinese Universities* 18 (2003): 332-342.
- [49] Oshiro, Thais Mayumi, Pedro Santoro Perez, and José Augusto Baranauskas. 2012. "How many trees in a random forest?" *Machine Learning and Data Mining in Pattern Recognition: 8th International Conference, MLDM 2012, Berlin, Germany, July 13-20, 2012. Proceedings* 8.
- [50] Biau, Gérard, and Erwan Scornet. "A random forest guided tour." *Test* 25 (2016): 197-227.