



**Digital mapping of soil classes using satellite images and environmental data in Badr watershed, Kurdistan province**

Moslem Zarinibahador<sup>1</sup>✉

1- PhD graduate in soil science, Soil and Water Research Expert, Agricultural and Natural Resources Research and Training Center, Tehran Province, Agricultural Research, Education and Extension Organization (TAT), Tehran, Iran.

Email: [moslem.zarini@ymail.com](mailto:moslem.zarini@ymail.com)

---

**Article Info**

**Article type:** Research Article

**Article history:**

**Received:** May 31, 2025

**Revised:** July 05, 2025

**Accepted:** July 16, 2025

**Published:** Summer, 2025

**Keywords:**

Auxiliary variable,  
Digital soil mapping,  
Soil classification,  
ANN.

---

**ABSTRACT**

Digital Soil Mapping (DSM), as a modern approach to soil map production, offers higher accuracy and efficiency compared to traditional methods by utilizing advanced algorithms and diverse environmental data. This study aims to compare the performance of several machine learning models—including Artificial Neural Networks (ANN), k-Nearest Neighbors (KNN), Discriminant Analysis (DA), Multiple Logistic Regression (MLR), Decision Tree Analysis (DTA), and Random Forest (RF)—in spatial prediction of soil classification classes, from the order to family level, within the DSM framework. The input data were derived from satellite imagery and environmental covariates related to the Badr watershed, located in Kurdistan Province, Iran. Additionally, an ensemble logistic regression model was employed to integrate the outputs of the aforementioned models. Results of 10-fold spatial cross-validation showed that the ensemble logistic regression model achieved the highest accuracy in predicting most soil classification levels. The Random Forest model consistently ranked second in prediction performance, while KNN showed the lowest accuracy. Among the environmental covariates, geomorphology, topography, and vegetation cover were identified as the most influential factors in predicting soil classes from order to subgroup levels. However, vegetation cover was less significant at the family level. The findings of this research highlight the potential of machine learning models to enhance the accuracy of soil maps and contribute to sustainable natural resource management.

---

**Cite this article:** Zarinibahador, M., (2025), Digital mapping of soil classes using satellite images and environmental data in Badr watershed, Kurdistan province, *Soil and Sustainable Development*, 1 (1), 47-73.

**DOI:** <https://doi.org/10.22034/ssd.2025.527346.1006>

© The Author(s). Publisher: The University of Jiroft Press

---



## نقشه‌برداری رقومی کلاس‌های خاک با استفاده از تصاویر ماهواره‌ای و داده‌های محیطی در حوضه آبخیز بدر، استان کردستان

مسلم زرینی بهادر<sup>✉</sup>

۱- فارغ التحصیل دکتری خاکشناسی، کارشناس تحقیقات خاک و آب مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان تهران، سازمان تحقیقات، آموزش و ترویج کشاورزی (تات)، تهران، ایران، ایمیل: [moslem.zarini@ymail.com](mailto:moslem.zarini@ymail.com)

اطلاعات مقاله	چکیده
نوع مقاله: پژوهشی	نقشه‌برداری رقومی خاک (DSM) به‌عنوان رویکردی نوین در تهیه نقشه‌های خاک، با بهره‌گیری از الگوریتم‌های پیشرفته و داده‌های محیطی متنوع، دقت و سرعت بالاتری را نسبت به روش‌های سنتی فراهم می‌آورد. هدف این پژوهش، مقایسه عملکرد چندین مدل یادگیری ماشین شامل شبکه عصبی مصنوعی (ANN)، نزدیک‌ترین همسایه (KNN)، تحلیل تشخیصی (DA)، رگرسیون لجستیک چندگانه (MLR)، درخت تصمیم (DTA) و جنگل تصادفی (RF) در پیش‌بینی مکانی کلاس‌های مختلف خاک از سطح رده تا فامیل در حوضه آبخیز بدر، واقع در استان کردستان می‌باشد. داده‌های کمکی مورد استفاده، از تصاویر ماهواره‌ای و متغیرهای محیطی مربوط به حوضه آبخیز تهیه شد. همچنین از مدل رگرسیون لجستیک ترکیبی برای تلفیق خروجی مدل‌های مختلف استفاده گردید. نتایج اعتبارسنجی با روش کافلد ۱۰ مکانی نشان داد که مدل MLR، بهترین عملکرد را در پیش‌بینی اغلب سطوح طبقه‌بندی خاک دارد. مدل RF در بیشتر موارد رتبه دوم را به خود اختصاص داد، در حالی که مدل KNN ضعیف‌ترین عملکرد را نشان داد. متغیرهای کمکی مهم در پیش‌بینی رده، زیررده، گروه بزرگ و زیرگروه خاک به ترتیب اهمیت عبارت‌اند از ژئومورفولوژی، پستی‌وبلندی و پوشش گیاهی. همچنین پارامترهای ژئومورفولوژی و پستی‌وبلندی بیشترین تأثیر در پیش‌بینی فامیل خاک را داشتند. نتایج کلی این پژوهش نشان داد که نقشه‌برداری رقومی خاک نقش بسزایی در بهبود دقت نقشه‌های خاک و مدیریت پایدار منابع طبیعی دارد.
تاریخ دریافت: ۱۰ خرداد ۱۴۰۴ تاریخ بازنگری: ۱۴ تیر ۱۴۰۴ تاریخ پذیرش: ۲۵ تیر ۱۴۰۴ تاریخ انتشار: تابستان ۱۴۰۴	
واژه‌های کلیدی: متغیرهای کمکی، نقشه‌برداری رقومی خاک، رده‌بندی خاک، شبکه عصبی مصنوعی.	
ارجاع به مقاله: زرینی بهادر، مسلم، (۱۴۰۴)، نقشه‌برداری رقومی کلاس‌های خاک با استفاده از تصاویر ماهواره‌ای و داده‌های محیطی در حوضه آبخیز بدر، استان کردستان، خاک و توسعه پایدار، ۱ (۱)، ۴۷-۷۳.	
© نویسندگان، ناشر: انتشارات دانشگاه جیرفت	
DOI: <a href="https://doi.org/10.22034/ssd.2025.527346.1006">https://doi.org/10.22034/ssd.2025.527346.1006</a>	



## مقدمه

نقشه‌بردای رقومی خاک (DSM) شامل روش‌های ریاضی برای ایجاد مدل‌هایی برای پیش‌بینی خصوصیات پیوسته و طبقه‌بندی خاک با استفاده از ویژگی‌های محیطی به عنوان پیش‌بینی‌کننده است. روش مدل‌سازی را می‌توان با استفاده از چارچوب DSM (McBratney et al., 2003) برای ارتباط متغیرهای کمکی محیطی با متغیر خاک هدف پیاده‌سازی کرد. از آنجایی که نقشه‌برداری معمولی خاک نیاز به نمونه‌برداری متراکم از خاک دارد، به دست آوردن اطلاعات جامع خاک به زمان و هزینه بیشتری نیاز دارد. در مقابل، DSM می‌تواند بر این محدودیت‌ها غلبه کند و نقشه‌های خاک را سریع‌تر و اقتصادی‌تر در بیشتر نقاط جهان پیش‌بینی کند. برای دستیابی به این هدف، می‌توان با استفاده از یک منبع اطلاعاتی جامع و با دسترسی آسان، مانند مدل رقومی ارتفاع (DEM) و تصاویر سنجنش از دور، نقشه‌هایی از ویژگی‌های مختلف خاک و کلاس‌های خاک ایجاد کرد (Grinand et al., 2008).

خاک یکی از بنیادی‌ترین عناصر اکوسیستم به‌شمار می‌رود و نقش حیاتی متعددی از جمله تولید زیست‌توده و غذا، ترسیب کربن، حفظ تنوع زیستی خاک، تصفیه آب و ارائه خدمات اجتماعی و فرهنگی ایفا می‌کند. نقشه‌های خاک قادرند عملکردها و کارکردهای مختلف خاک را نمایش داده و توصیف کنند؛ از جمله می‌توانند برای ارزیابی قابلیت‌های عملکردی خاک مانند ظرفیت نگهداری آب، مواد مغذی، کربن و سایر ویژگی‌ها مورد استفاده قرار گیرند (Adhikari and Hartemink, 2016). نقشه‌برداری از عملکردهای خاک زمانی به‌صورت مؤثر امکان‌پذیر است که نوع خاک منطقه به‌خوبی شناسایی شده باشد. با ترکیب این اطلاعات با داده‌های محیطی و بهره‌گیری از توابع انتقالی، نقشه‌برداری رقومی خاک می‌تواند برای ارزیابی قابلیت‌های خاص هر طبقه از خاک به کار گرفته شود (Dominati et al., 2010). نقشه‌نهایی خاک باید شامل اطلاعاتی باشد که ویژگی‌ها و کیفیت زمین را به‌طور دقیق توصیف کند؛ اطلاعاتی که می‌توانند مبنایی برای توسعه راهبردهای مدیریت پایدار زمین و حفاظت از خاک در برابر تغییرات کاربری زمین فراهم آورند.

DSM پتانسیل زیادی در غلبه بر برخی از ناسازگاری‌های نقشه‌برداری سنتی خاک نشان می‌دهد (Caubet et al., 2019; Ma et al., 2019). این امر به این دلیل است که DSM می‌تواند مشاهدات مورفولوژیکی مبتنی بر میدان، اندازه‌گیری‌های آزمایشگاهی یا خاک طبقه‌بندی‌شده را با منابع مختلف غنی از داده‌های محیطی جغرافیایی مبتنی بر شبکه که فاکتورهای SCORPAN را نشان می‌دهند، ادغام کند (McBratney et al., 2003). این امر با استفاده از ایجاد نقشه‌های

خاک مبتنی بر پیکسل توسط مدل‌های ریاضی یا آماری که متغیرهای کمکی محیطی را با هرگونه اطلاعات خاک مرتبط می‌کند، فعال می‌شود. توابع مکانی خاک را می‌توان با برازش یک مدل با استفاده از روش‌های زمین آماری و یا آماری که توابع پیش‌بینی مکانی خاک برای پیش‌بینی اطلاعات خاک در مکان‌های نمونه‌برداری نشده استفاده می‌شوند، کمی‌سازی کرد (Heung et al., 2016; Minasny and McBratney, 2016).

طبقات خاک توسط چندین نوع گسترده از روش‌های آماری و داده‌کاوی از طریق DSM پیش‌بینی شده است. این شامل روش‌ها رگرسیون لجستیک (LR) (Zeraatpisheh et al., 2017)، جنگل‌های تصادفی (RF) (Pahlavan-Rad et al., 2016)، درختان طبقه‌بندی (CT) (Adhikari et al., 2014)، ماشین‌های بردار پشتیبانی (SVM) (Kovačević et al., 2010)، و شبکه‌های عصبی مصنوعی (ANN) (Behrens et al., 2005)، است. رویکرد معمول تحقیق ارزیابی چندین مدل و انتخاب بهترین عملکرد است (Brungard et al., 2015; Taghizadeh-Mehrjardi et al., 2015). با این حال، انتخاب یک طبقه‌بندی‌کننده مناسب به طور بالقوه مشکل‌ساز است (Górecki and Krzyśko, 2015). به این دلیل است که هر طبقه‌بندی‌کننده در یک موقعیت خاص، مزایا و معایب خاص خود را دارد، بنابراین یک طبقه‌بندی‌کننده ممکن است در یک موقعیت معین دقت بیشتری نسبت به سایر طبقه‌بندی‌کننده‌ها داشته‌باشد و بالعکس (Górecki and Krzyśko, 2016; Swiderski et al., 2015). مجموعه چندین مدل آموزش دیده جایگزینی است که به ترکیب دانش و اطلاعات به دست آمده از طبقه‌بندی‌کننده‌های مختلف کمک می‌کند که منجر به دقت بیشتر طبقه‌بندی می‌شود (Swiderski et al., 2016). علاوه بر این، مدل‌های ترکیبی به طور بالقوه منجر به پیش‌بینی‌های بهتر و پایدارتر و همچنین کاهش خطر انتخاب طبقه‌بندی‌کننده‌های اشتباه می‌شوند (Górecki and Krzyśko, 2015).

علاقه فزاینده‌ای به DSM برای استفاده از همه مدل‌ها از طریق مدل‌سازی گروهی وجود دارد (Román Dobarco et al., 2009; Diks and Vrugt, 2010; Wasson and Hartemink, 2017). ایده انگیزشی این است که مدل‌های مختلف رقیب اما مرتبط را می‌توان ترکیب کرد و در نتیجه کیفیت وظایف شناسایی و طبقه‌بندی را با یکپارچه‌سازی دانش کسب‌شده افزایش داد (Wasson and Hartemink, 2009). روش‌های ترکیبی در نقشه‌برداری رقومی خاک، ویژگی‌های پیوسته خاک را مورد بررسی قرار داده‌اند. این ویژگی‌ها شامل، اما نه محدود به pH خاک، بافت خاک و ظرفیت آب موجود خاک است (Román Dobarco et al., 2017; Caubet et al., 2019; Malone et al., 2014; Padarian et al., 2014).

یکی از کاربردهای مهم و بسیار مفید DSM می‌تواند نقشه برداری طبقات خاک برای به دست آوردن اطلاعات برای استراتژی‌های خاص مدیریت خاک با توجه به قابلیت‌ها و محدودیت‌های واحدهای مختلف نقشه باشد. در چارچوب DSM، مفهوم خاک یکنواخت (Mallavan et al., 2010) می‌تواند برای نقشه برداری و به دست آوردن اطلاعات خاک در مناطقی که نمونه برداری انجام نشده است بسیار مفید باشد. این مفهوم بر تئوری تشکیل خاک تکیه دارد (Jenny, 1994) که می‌تواند با یافتن رابطه بین مناطق با شرایط تشکیل خاک مشابه در DSM گنجانده شود. بنابراین، می‌توانیم منطقه‌ای بدون داده‌های خاک (به نام سایت گیرنده) با استفاده از رابطه عوامل تشکیل دهنده خاک با داده‌های خاک یک منطقه نمونه برداری شده (به نام سایت اهداکننده) نقشه برداری کنیم. فرآیند نقشه برداری از یک سایت نمونه برداری نشده با استفاده از الگوریتمی که متغیر هدف را با عوامل تشکیل دهنده خاک در منطقه نمونه برداری شده مرتبط می‌کند، برون‌یابی مکانی نامیده می‌شود. این نوع نقشه برداری صرفه جویی در زمان و از نظر اقتصادی ارزان است. بنابراین، به دست آوردن نقشه یک منطقه بازدید نشده برای برنامه‌ریزی کاربری اراضی و تصمیم‌گیری استراتژیک در مورد منابع خاک، یک رویکرد سودمند و بسیار مفید است.

مطالعات اندکی از برون‌یابی برای نقشه برداری رقومی خاک با روش‌هایی مانند طبقه‌بندی درختان (Lemerrier et al., 2012)، شبیه‌سازی و رگرسیون (Malone et al., 2016) و رگرسیون لجستیک چندجمله‌ای (Abbaszadeh Afshar et al., 2018) استفاده کرده‌اند. اولین مطالعات برون‌یابی توسط بوئی و موران (Bui and Moran, 2001) و اسکال و همکاران انجام شد (Scully et al., 2005). با این حال، مقایسه الگوریتم‌های مختلف DSM برای نقشه برداری نوع خاک با استفاده از رویکرد برون‌یابی بررسی نشده است. بنابراین، در مطالعه حاضر، چندین الگوریتم یادگیری ماشین را برای برون‌یابی زیرگروه‌های خاک مقایسه می‌کنیم. انتخاب انواع الگوریتم‌ها بر اساس نوعی طبقه‌بندی ارائه شده با توجه به پیچیدگی مدل‌ها در مطالعات برون‌یابی و همکاران بود (Brungard et al., 2015). این تنوع شامل درخت‌های طبقه‌بندی، رگرسیون لجستیک، بردارهای پشتیبانی، شبکه‌های عصبی مصنوعی و سایر الگوریتم‌ها است که در بخش مواد و روش‌ها ارائه شده‌اند.

موضوع مشاهدات طبقات خاک نامتعادل منجر به کاهش دقت الگوریتم‌های یادگیری ماشینی می‌شود و اغلب منجر به از دست دادن یک یا چند کلاس با تعداد مشاهدات بسیار کمتر (به نام طبقات اقلیت) در نقشه تولید شده می‌شود (He

است (Zhu et al., 2017). برای غلبه بر این مشکل، آموزش کلاس خاک اقلیت با روش نمونه برداری بیش از حد بر روی الگوریتم‌های انتخابی DSM آزمایش شد تا تأثیر آن بر دقت برون‌یابی مشاهده شود (Sharififar et al., 2019a). به نظر می‌رسد این راه حل ساده‌ترین راه برای حل مشکل باشد و همچنین توسط نویسندگان برای مطالعاتی غیر از خاکشناسی پیشنهاد شده است (Abdi and Hashemi, 2015; Chawla et al., 2002).

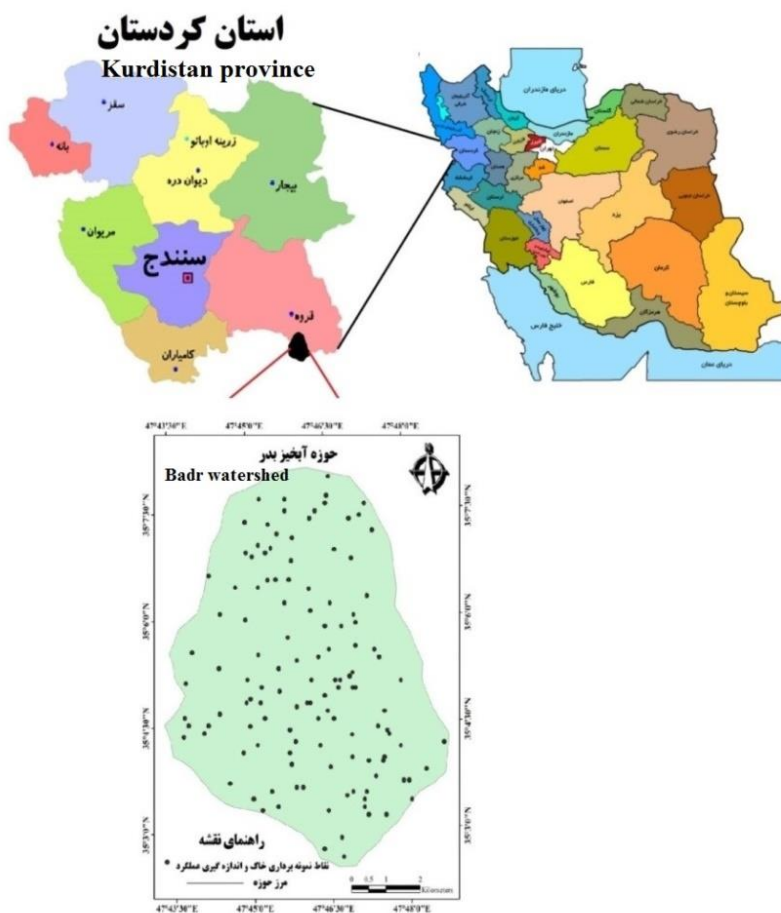
نقشه‌برداری دقیق و به‌روز از کلاس‌های خاک در مقیاس‌های محلی و ناحیه‌ای برای مدیریت پایدار منابع طبیعی، کشاورزی دقیق، ارزیابی ریسک فرسایش و برنامه‌ریزی کاربری اراضی ضروری است. روش‌های سنتی نقشه‌برداری خاک، وقت‌گیر، پرهزینه و محدود در پوشش مکانی هستند. در مقابل، نقشه‌برداری رقومی خاک با استفاده از مدل‌های یادگیری ماشین و داده‌های ماهواره‌ای، امکان تولید نقشه‌های دقیق و گسترده‌تری را با صرف زمان و هزینه کمتر فراهم می‌کند. لذا این پژوهش می‌کوشد با به‌کارگیری مدل‌های پیشرفته، عملکرد و قابلیت این روش‌ها را در پیش‌بینی کلاس‌های مختلف خاک ارزیابی کند. و هدف صرفاً مدل‌سازی کلاس‌های خاک نیست، بلکه مقایسه عملکرد مدل‌های مختلف یادگیری ماشین در سطوح مختلف طبقه‌بندی خاک (رده، زیررده، گروه بزرگ، زیرگروه، فامیل) و تحلیل متغیرهای مؤثر در پیش‌بینی آن‌ها نیز از جمله اهداف اصلی پژوهش است.

### منطقه مطالعاتی

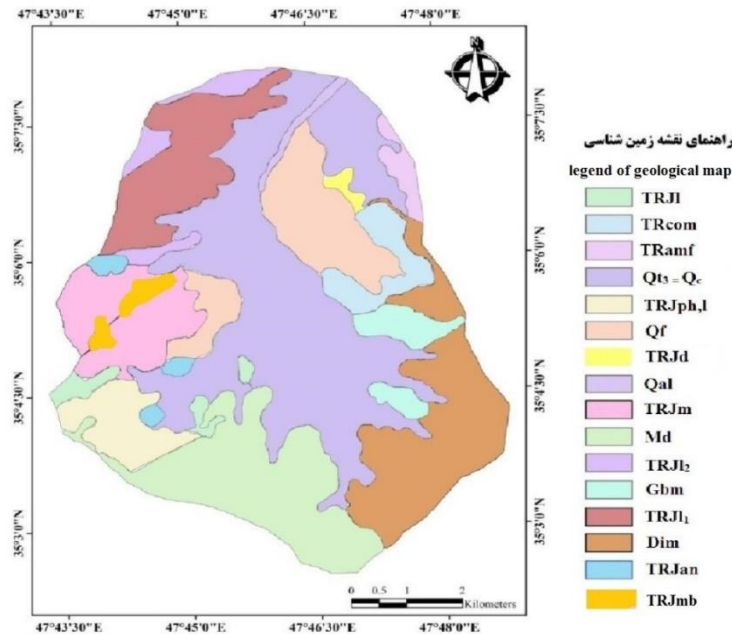
منطقه مورد مطالعه با وسعتی حدود ۶۷۰۰ هکتار در جنوب شهرستان قروه و بین مختصات جغرافیائی ۴۷ درجه و ۴۳ دقیقه و ۴۷ درجه و ۴۹ دقیقه طول شرقی و ۳۵ درجه و ۲ دقیقه و ۳۵ درجه و ۸ دقیقه عرض شمالی واقع شده است (شکل ۱). مطابق با آمار طولانی‌مدت هواشناسی، متوسط درجه حرارت سالیانه‌ی هوا ۱۲/۰۷ درجه سانتی‌گراد و میانگین بارندگی سالیانه ۳۴۵/۸ میلی‌متر است. الگوی بارش، مدیترانه‌ای بوده و بخش اعظم بارندگی در فصول سرد سال اتفاق می‌افتد. بر اساس اقلیم نمای آمبرژه، اقلیم این ناحیه نیمه‌خشک و سرد است. در منطقه مورد مطالعه، رژیم رطوبتی خاک زریک (Xeric) و رژیم حرارتی آن مزیک (Mesic) تعیین شده است.

قدیمی‌ترین واحد چینه‌شناسی در منطقه، تریاس است. رسوبات تریاس از دولومیت کاملاً بلورین و خردشده تشکیل شده است. دگرگونی در این منطقه در حد رخساره آمفیبولیت است و دگرگونی مجاورتی حاصل از نفوذ مجموعه پلوتونیک

قروه در سنگ‌هایی که قبلاً در مزوزوئیک متحمل دگرگونی ناحیه‌ای شده بودند تا رخساره پیروکسن هورنفلس پیش رفته است (حریری، ۱۳۷۴). اصلی‌ترین رخداد کوهزایی در این منطقه مربوط به کوهزایی کیمرین میانی است که سبب دگرگونی سنگ‌های رسوبی- آتشفشانی (تریاس-ژوراسیک) گردیده و حاصل آن تشکیل مرمر، آمفیبولیت، شیست و فیلیت است (شکل ۲) (حسینی، ۱۳۷۶). از لحاظ ژئومرفولوژی، حوزه‌ی آبخیز مورد مطالعه شامل کوه و اراضی پای کوهی (پیدمنت) است. کاربری‌های رایج اراضی کشاورزی، گندم دیم و آبی، جو دیم و آبی، نخود، لوبیا، یونجه، شبدر و... است.



شکل ۱- موقعیت منطقه مورد مطالعه و پراکنش مکانی نقاط نمونه برداری



شکل ۲- نقشه زمین‌شناسی حوضه آبخیز بدر شهرستان قروه، استان کردستان

## مواد و روش‌ها

برای اجرای این پژوهش مراحل زیر، طبق شکل ۴ انجام گرفت:

جمع‌آوری داده‌های متغیرهای کمکی، پردازش و فرآیند انتخاب مهم‌ترین آن‌ها

در این پژوهش، به‌منظور تولید متغیرهای کمکی مورد استفاده در مدل‌سازی، از سه منبع اصلی داده بهره گرفته شد: تصاویر ماهواره‌ای، مدل رقومی ارتفاع (DEM) و نقشه ژئومورفولوژی. مبنای اصلی برای استخراج متغیرهای توپوگرافی، مدل رقومی ارتفاع با قدرت تفکیک مکانی ۱۰ متر بود.

در گام نخست، داده‌های اولیه شامل تصاویر ماهواره‌ای لندست ۸ (سال ۲۰۱۳) و مدل رقومی ارتفاع منطقه از سازمان نقشه‌برداری کشور تهیه گردید. همچنین، نقشه زمین‌شناسی منطقه قروه از سازمان زمین‌شناسی کشور دریافت و با برش آن در محدوده حوضه آبخیز بدر، نقشه زمین‌شناسی اختصاصی این حوضه استخراج شد. این نقشه در محیط سامانه اطلاعات جغرافیایی (GIS) رقومی‌سازی و پردازش گردید.

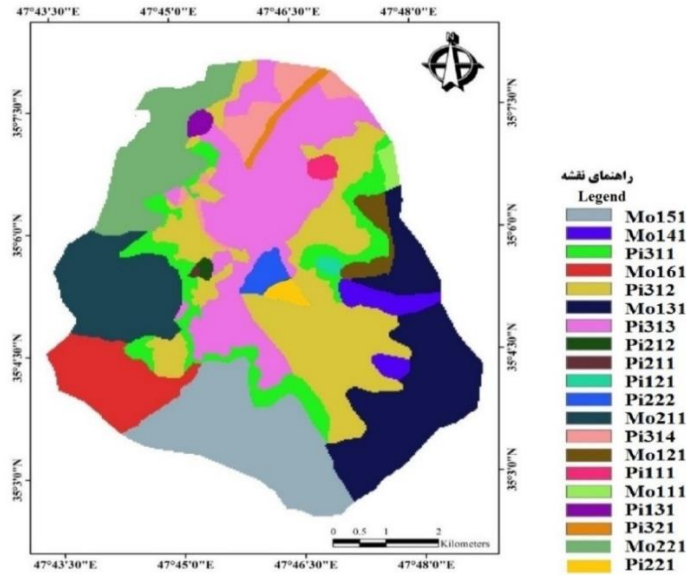
بر اساس داده‌های زمین‌شناسی و با استفاده از روش ژئوپدولوژیک زینک (Zinck, 1989)، نقشه ژئومورفولوژی حوضه ترسیم شد. در ادامه، با استفاده از مدل رقومی ارتفاع و ابزارهای تحلیلی نرم‌افزارهای ArcGIS 10 و SAGA، مشتقات

مختلف از مرتبه اول و دوم مدل ارتفاعی (مانند شیب، جهت شیب، انحناى پلان و انحناى پروفیل، شاخص رطوبت توپوگرافی، شاخص موقعیت ارتفاعی و شاخص جریان تجمعی) استخراج گردید. در پایان این مرحله، مجموعه‌ای از متغیرهای کمکی در قالب لایه‌های اطلاعاتی پیوسته تولید شد. به منظور کاهش تعداد متغیرها و شناسایی مؤثرترین عوامل در پیش‌بینی کلاس‌های خاک، تحلیل همبستگی و معیارهای آماری دیگر نظیر تحلیل مؤلفه‌های اصلی (PCA) یا اهمیت متغیرها در مدل‌های یادگیری ماشین مانند جنگل تصادفی به کار گرفته شد.

تفکیک واحدهای ژئومورفیک موجود در منطقه مطالعاتی با توجه به تفاوت رنگ، الگوی آبراه‌های، شیب، شکل شیب و ارتفاع که در تصاویر سه‌بعدی گوگل ارث قابل مشاهده هستند، براساس نقشه زمین‌شناسی و سامانه سلسله مراتبی زینک (Zinck, 1989) انجام شد. این سامانه از بالاترین تا پایین‌ترین سطح طبقه‌بندی شامل زمین‌نما (Landscape)، پستی‌وبلندی (Relief) و یا قالب (Molding)، سنگ و مواد مادری (Lithology) و شکل زمین (Landform) است (جدول ۲). به منظور نشان دادن رابطه بین زمین ریخت (Geoform) و خاک، خاک‌های) مربوط به هر شکل زمین نیز معرفی گردید. در پایان، نقشه ژئومورفولوژی در سامانه اطلاعات جغرافیائی (GIS) رقومی گردید. در مرحله دوم، محل نقاط مشاهداتی تعیین، شناسایی خاک‌ها در صحرا (Schoeneberger et al., 2012)، نمونه‌برداری از لایه‌های مختلف خاک، اندازه‌گیری‌های فیزیکی و شیمیائی و کانی‌شناسی خاک‌ها انجام و خاک‌ها طبقه‌بندی شدند (Soil Survey Staff, 2014). در مرحله سوم، مدل‌سازی انجام، نقشه‌ی ویژگی‌های خاک تهیه و ارزیابی مدل‌ها صورت گرفت.

### جدول ۱- راهنمای نقشه زمین‌شناسی

علامت روی نقشه زمین‌شناسی	جنس سنگ‌ها و رسوبات
TRJl	Recrystallized limestone and dolomite
TRcom	Metamorphic complex: orthogneiss, amphibolite, mica schist, metagabbro-diorite, tactite and scapolite marble
Tramf	Alteration of marble and amphibolite
Qt3 = Qc	Lowest alluvial plain deposits
TRJph,l	Association of micaschist, phyllite, slate and crystalline limestone and dolomite
Qf	Alluvial fan
TRJd	Crystalline dolomite and limestone
Qal	Recent alluvium in stream channel
TRJm	Alteration of schist, quartzite, meta-volcanic rocks and marble
Md	Monzodiorite, quartz monzonite, syenite quartz diorite
TRJ2	Grey to black marble
Gbm	Foliated and brecciated gabbro and gabbro-diorite
TRJ1	White to cream marble
Dim	Foliated and brecciated diorite and gabbro-diorite, diorite
TRJan	Andalusite, cordierite schist
TRJmb	Marble with tremolite and phlogopite



شکل ۳- نقشه ژئودولوژی حوضه آبخیز بدر شهرستان قروه، استان کردستان

جدول ۲- تحلیل زمین نما در حوضه آبخیز بدر

Landscape <sup>5</sup>	Relief <sup>3</sup> /Molding <sup>4</sup>	Lithology <sup>2</sup>	Landform <sup>1</sup>	
(Mo) Mountain	High Mountain(Mo1)	(Mo11) TRamf	(Mo111) Slope facet complex	
		TRcom (Mo12)	Slope facet complex (Mo121)	
		Dim (Mo13)	(Mo131) Slope facet complex	
		(Mo14) Gbm	(Mo141) Slope facet complex	
		(Mo15) Md	(Mo151) Slope facet complex	
		(Mo16) TRJph,l	(Mo161) Slope facet complex	
		(Mo21) TRJm	(Mo211) Slope facet complex	
(Pi) Piedmont	(Pi1) Hill	TRJl(Mo22)	Slope facet complex(Mo221)	
		TRJd (Pi11)	Slope facet complex (Pi111)	
		(Pi12) TRcom	Slope facet complex (Pi121)	
	Alluvial fan (Pi2)	(Pi13) Qt1	Slope facet complex (Pi131)	
		Qf1(Pi21)	Apical (Pi211)	
	Glacis <sup>6</sup> (Pi3)	Qf2(Pi22)		Distal (Pi212)
				Apical (Pi221)
Qf3 (Pi31)			Distal (Pi222)	
			Slope > 20% (Pi311)	
Qal (Pi32)		20% > Slope > 10% (Pi312)		
		10% > Slope > 5% (Pi313)		
		5% > Slope > 0% (Pi314)		
		Slope facet complex (Pi321)		

<sup>1</sup> آخرین، همگونترین و کوچکترین واحد ژئودولوژیکی است که با شکل، دینامیک و سن مشخص می شود (زینک، ۱۹۸۹)

<sup>2</sup> به جنس سنگ های سخت و رخساره تشکیلات پوششی نرم اشاره دارد (زینک، ۱۹۸۹)

<sup>3</sup> شکلی از زمین است که تعیین کننده آن مجموعه توپوگرافی و زمین شناسی است (زینک، ۱۹۸۹)

<sup>4</sup> شکلی از زمین است که تعیین کننده آن فرآیندهای مورفونتیکی یا شرایط مورفواقلیمی است (زینک، ۱۹۸۹)

<sup>5</sup> قسمت بزرگی از زمین که در آن یا ریلیف های مشابه تکرار شده اند یا مجموعه ای از ریلیف های غیر مشابه دیده می شوند (زینک، ۱۹۸۹).

<sup>6</sup> مترادف پدیمنت و ترجمه آن به فارسی "دشت سر" است. محیط تشکیل فن ها اغلب رسوبی است، ولی دشت سرها در محیط های رسوبی-فرسایشی بوجود می آیند. سنگ بنای دشت سر توسط حرکت تکتونیکی زمین گذاشته می شود (زینک، ۱۹۸۹).

## مدلسازی مکانی

در این تحقیق از رویکردهای شبکه عصبی مصنوعی، نزدیک ترین همسایه K، تحلیل تشخیصی، رگرسیون لجستیک چندگانه، درخت تصمیم و جنگل تصادفی برای مدل سازی مکانی کلاس های مختلف خاک از سطح رده تا فامیل استفاده شد.

## اعتبارسنجی مدل ها

در این پژوهش، جهت ارزیابی مدل های یادگیری ماشین از دو روش بهره گرفته شد: ارزیابی متقابل تصادفی و ارزیابی تصادفی مکانی. در این نوع اعتبارسنجی ها، داده ها به K زیرمجموعه دسته بندی شدند. از این K زیرمجموعه، هر بار یکی برای اعتبارسنجی و K-1 تای دیگر برای آموزش بکار رفتند. این روال، K بار تکرار شد و همه داده ها دقیقاً یکبار برای آموزش و یکبار برای اعتبارسنجی بکار رفتند. در نهایت، میانگین نتیجه این K بار اعتبارسنجی به عنوان یک تخمین نهایی برگزیده شد. لازم به ذکر است که تفاوت ارزیابی متقابل تصادفی و ارزیابی تصادفی مکانی، در انتخاب مجموعه داده های خاک جهت آموزش و ارزیابی است. در روش ارزیابی متقابل تصادفی، داده های به طور تصادفی به دو دسته تقسیم می شوند: داده هایی که برای آموزش به کار می روند و داده هایی که برای اعتبارسنجی یا آزمون مورد استفاده قرار می گیرند. در چنین شرایطی ممکن است داده های آموزش و آزمون از لحاظ مکانی در کنار یکدیگر در فاصله بسیار نزدیک قرار گیرند. اما در روش ارزیابی متقابل مکانی، در ابتدا منطقه مطالعاتی به تعدادی قطعه، تقسیم بندی شده و سپس داده های قرار گرفته در یک قطعه جهت ارزیابی مدل استفاده شده و بقیه قطعه ها برای آموزش مدل به کار می روند. در چنین شرایطی، می توان مطمئن شد که مدل آموزش دیده، در کل منطقه مورد ارزیابی قرار گرفته است (Valavi et al., 2018).

در این تحقیق برای ارزیابی مدل های یادگیری ماشین، از دو معیار اعتبارسنجی صحت کلی و شاخص کاپا استفاده شد. صحت کلی (Overall Accuracy) یکی از شاخص های مهم و رایج در ارزیابی مدل های طبقه بندی و سیستم های تحلیل داده است. این معیار نشان می دهد که چه درصدی از نمونه های ورودی به درستی توسط مدل طبقه بندی شده اند. به عبارت دیگر، صحت کلی میزان موفقیت مدل در پیش بینی صحیح نمونه ها را بر اساس تعداد کل نمونه ها اندازه گیری می کند و به صورت معادله ۱ تعریف می شود:

$$\text{معادله ۱} \quad \text{نمونه درست طبقه بندی شده} = \frac{\text{صحت کلی}}{\text{کل نمونه ها}}$$

این شاخص، به عنوان یک معیار ساده و مستقیم، در بسیاری از مطالعات و پروژه‌های کاربردی در حوزه‌های مختلف مانند سنجش از دور، یادگیری ماشین، و تحلیل داده‌ها مورد استفاده قرار می‌گیرد. صحت کلی به طور خاص در مواردی که داده‌ها متعادل هستند، می‌تواند تصویر دقیقی از عملکرد کلی مدل ارائه دهد. با این حال، در مواردی که داده‌ها نابرابر هستند و تعداد نمونه‌های هر کلاس متفاوت است، استفاده صرف از صحت کلی ممکن است گمراه‌کننده باشد، زیرا ممکن است مدل صرفاً بر اساس پیش‌بینی کلاس‌های غالب، نمره بالایی کسب کند در حالی که در پیش‌بینی کلاس‌های نادر، عملکرد ضعیفی داشته باشد. در چنین مواردی، معیارهای دیگری مانند شاخص کاپا (Kappa)، دقت (Precision)، حساسیت (Recall)، و شاخص F1 بهتر می‌توانند تصویر جامع‌تری از عملکرد مدل ارائه دهند (Powers, 2020). در نتیجه، صحت کلی به عنوان یک شاخص اولیه و ساده در ارزیابی مدل‌های طبقه‌بندی کاربرد دارد، اما برای تحلیل دقیق‌تر و جامع‌تر، باید همراه با معیارهای دیگر مورد استفاده قرار گیرد.

شاخص کاپا (Kappa Index) یک معیار آماری است که برای ارزیابی میزان توافق بین دو ارزیاب یا دو روش طبقه‌بندی استفاده می‌شود، به ویژه زمانی که ارزیابی‌ها شامل دسته‌بندی‌های کیفی (مانند طبقه‌بندی کلاس‌های خاک) باشند. این شاخص، میزان توافق فراتر از توافق تصادفی را نشان می‌دهد و مقداری بین -۱ و +۱ دارد. هرچه مقدار کاپا به ۱ نزدیک‌تر باشد، توافق بین ارزیاب‌ها بیشتر است و مقدار ۱ نشان‌دهنده توافق کامل است (Viera and Garrett, 2005). فرمول شاخص کاپا به شکل معادله ۲ است:

$$\text{معادله ۲} \quad K = \frac{\text{Pr}(a) - \text{Pr}(e)}{1 - \text{Pr}(e)}$$

- $\text{Pr}(a)$ : قرارداد نسبی مشاهده شده بین مقیاس‌ها یا ارزیاب‌ها (توافق بین دو کدگذار)
- $\text{Pr}(e)$ : درصد توافق مورد انتظار (احتمال فرضی قرارداد شانس)

## نتایج و بحث

متغیرهای کمکی مهم در پیش‌بینی کلاس‌های خاک

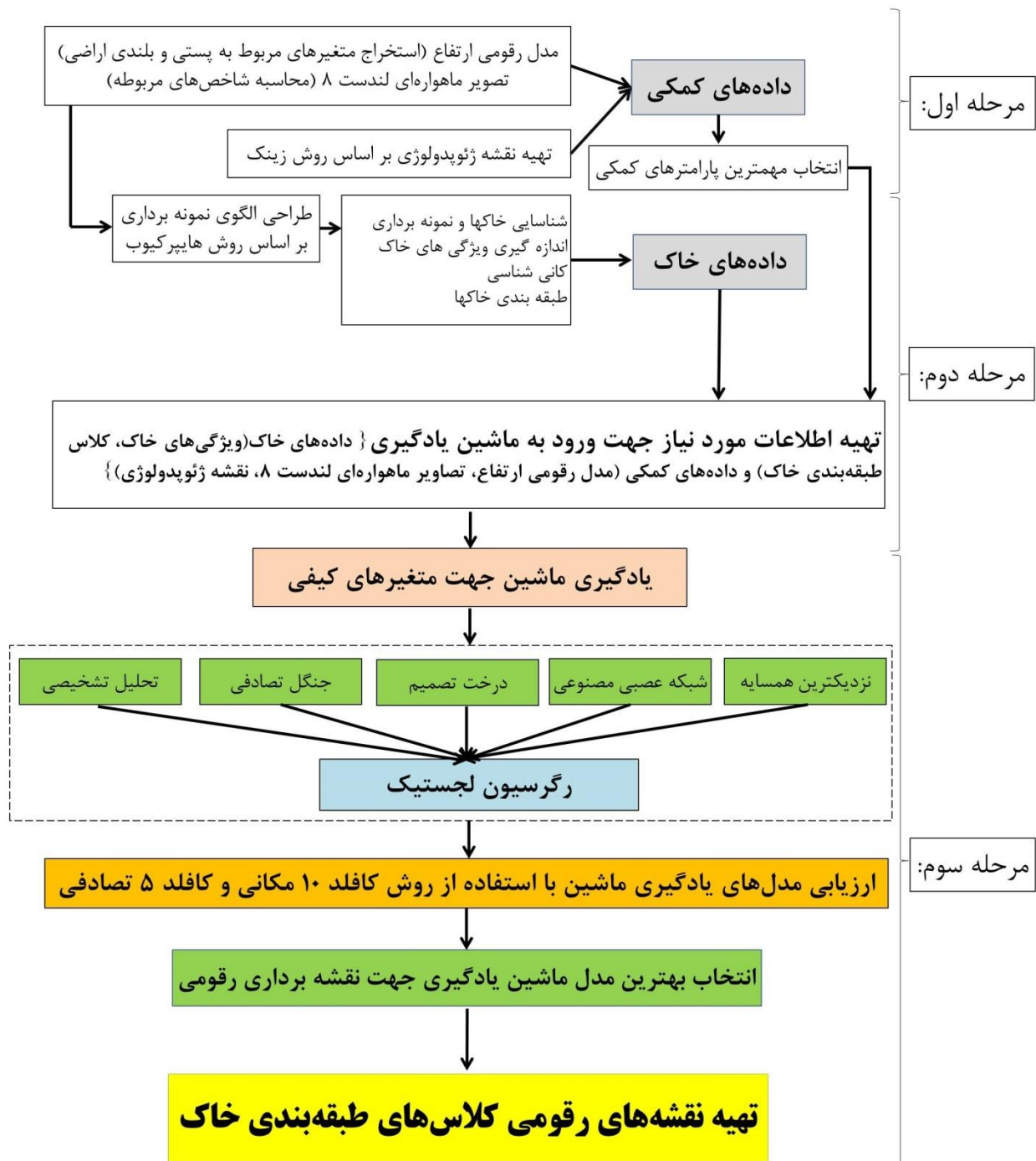
متغیرهای کمکی مهم در پیش‌بینی کلاس‌های طبقه‌بندی خاک، در منطقه مورد مطالعه به ترتیب اهمیت در جدول ۳ ارائه شده‌اند. درجه‌بندی اهمیت با به‌کارگیری الگوریتم ReliefAttributeEval نرم‌افزار WEKA مشخص گردید.

جدول ۳- متغیرهای کمکی مهم در پیش‌بینی کلاس‌های طبقه‌بندی خاک، به ترتیب اهمیت. برای هر کلاس، از چپ به راست، از اهمیت متغیر کمکی کاسته می‌شود

کلاس‌ها	متغیرهای کمکی مهم به ترتیب اهمیت
Order	Geomorphology, Slope, Topographic wetness index, Vertical distance to channel network, DEM, Valley depth, Catchment slope, SAVI, NDVI, MRRTF
Suborder	Geomorphology, Slope, Topographic wetness index, Vertical distance to channel network, DEM, Valley depth, Catchment slope, SAVI, NDVI, MRRTF
Great Group	Geomorphology, Relative slope position, Topographic wetness index, Vertical distance to channel network, DEM, Catchment slope, SAVI, NDVI, MRRTF
Subgroup	Geomorphology, Relative slope position, Channel network base level, Slope, DEM, Topographic wetness index, LS Factor, Catchment Slope, SAVI, NDVI
Family	Geomorphology, Channel network base level, Topographic wetness index, Vertical distance to channel network

نتایج پیش‌بینی سطوح طبقه‌بندی خاک‌های مورد مطالعه تا سطح فامیل

نتایج ارزیابی صحت پیش‌بینی سطوح طبقه‌بندی خاک مورد مطالعه، با استفاده از روش اعتبارسنجی کافلد ۱۰ مکانی در جدول ۴ ارائه شده است. این پیش‌بینی توسط مدل‌های نزدیک‌ترین همسایه K (KNN)، تحلیل درخت تصمیم (DTA)، شبکه عصبی مصنوعی (ANN)، جنگل تصادفی (RF) و رگرسیون لجستیک ترکیبی (MLR) صورت گرفت. در روش استفاده از رگرسیون لجستیک ترکیبی، پیش‌بینی توسط ترکیبی از چهار مدل قبل انجام گرفت (شکل ۴).



شکل ۴- مراحل اجرای پژوهش

جدول ۴- نتایج ارزیابی صحت پیش‌بینی سطوح طبقه‌بندی خاک مورد مطالعه، با استفاده از روش اعتبارسنجی کافلد ۱۰ مکانی

کلاس	شاخص صحت	KNN	DTA	ANN	RF	MLR
رده	صحت کلی	۰/۷۸۰	۰/۷۸۸	۰/۸۰۱	۰/۸۴۰	۰/۸۴۴
	ضریب کاپا	۰/۵۹۱	۰/۶۰۲	۰/۶۱۶	۰/۷۱۹	۰/۷۲۳
زیررده	صحت کلی	۰/۷۹۴	۰/۸۰۴	۰/۸۱۱	۰/۸۱۳	۰/۸۳۵
	ضریب کاپا	۰/۶۲۸	۰/۶۴۹	۰/۶۵۵	۰/۶۵۸	۰/۷۰۲
گروه بزرگ	صحت کلی	۰/۷۵۰	۰/۷۵۷	۰/۷۶۰	۰/۷۷۲	۰/۷۷۸
	ضریب کاپا	۰/۵۶۷	۰/۵۸۰	۰/۵۸۱	۰/۵۹۴	۰/۶۰۶
زیرگروه	صحت کلی	۰/۵۳۸	۰/۵۷۲	۰/۵۸۴	۰/۶۱۸	۰/۶۳۳
	ضریب کاپا	۰/۳۷۰	۰/۳۷۱	۰/۴۱۳	۰/۴۲۸	۰/۴۵۱
فامیل	صحت کلی	۰/۳۶۰	۰/۳۸۶	۰/۴۰۴	۰/۴۳۳	۰/۴۷۷
	ضریب کاپا	۰/۲۸۳	۰/۳۰۳	۰/۳۲۴	۰/۳۳۰	۰/۳۴۰

برای پیش‌بینی رده، زیررده، گروه بزرگ، زیرگروه و فامیل خاک، مدل رگرسیون لجستیک ترکیبی، بیشترین مقادیر شاخص‌های صحت کلی و کاپا را نشان داد. به عبارت دیگر، پیش‌بینی این مدل در مورد پنج سطح طبقه‌بندی خاک مورد اشاره، نسبت به سایر مدل‌ها بهترین است. زیرا در این روش، پیش‌بینی توسط ترکیبی از چهار مدل نزدیک‌ترین همسایه، تحلیل درخت تصمیم، شبکه عصبی مصنوعی و جنگل تصادفی صورت گرفت (جدول ۴). صحت و دقت پیش‌بینی پنج مدلی که بکار رفتند، به ترتیبی که در جدول ۴ نوشته شده‌اند، از چپ به راست افزایش نشان داد.

شاخص‌های صحت کلی و کاپا برای مدل‌های رگرسیون لجستیک ترکیبی و جنگل تصادفی، به ترتیب از رده به طرف فامیل به طور منظم سیر نزولی دارند. در مورد سه مدل دیگر، مقادیر این شاخص‌ها از رده به طرف زیررده، افزایش و از زیررده تا فامیل بطور منظم، کاهش نشان می‌دهند. بطور کلی، علت کاهش صحت پیش‌بینی سطوح مختلف طبقه‌بندی خاک از رده به طرف فامیل این است که تعداد ویژگی‌های خاک که جهت طبقه‌بندی از رده تا فامیل بکار می‌روند، به سمت سطوح پائین‌تر افزایش می‌یابد. این افزایش، پیش‌بینی را مشکل‌تر می‌سازد.

نتایج ارزیابی صحت پیش‌بینی سطوح طبقه‌بندی خاک مورد مطالعه، با استفاده از روش اعتبارسنجی کافلد ۵ تصادفی در جدول ۵ ارائه شده است. این پیش‌بینی توسط مدل‌های شبکه عصبی مصنوعی (ANN)، تحلیل درخت تصمیم (DTA)، آنالیز تشخیصی (DA) و رگرسیون لجستیک ترکیبی (MLR) صورت گرفت.

در روش اعتبارسنجی کافلد ۵ تصادفی، برای پیش‌بینی رده‌های خاک، مدل‌های شبکه‌ی عصبی مصنوعی و مدل رگرسیون لجستیک ترکیبی، نتایج مشابهی را ارائه نموده و مقادیر شاخص‌های صحت کلی و کاپای برابری را نشان دادند. پیش‌بینی این مدل‌ها به ترتیب بهتر از مدل‌های تحلیل درخت تصمیم و آنالیز تشخیصی بود. برای پیش‌بینی زیررده‌ها و گروه‌های بزرگ خاک، مدل رگرسیون لجستیک ترکیبی (MLR) بالاترین مقادیر صحت کلی و شاخص کاپا را نشان داد و مدل‌های شبکه عصبی مصنوعی، تحلیل درخت تصمیم و آنالیز تشخیصی به ترتیب در رتبه‌های بعدی قرار گرفتند. برای پیش‌بینی زیرگروه‌های خاک، مدل شبکه عصبی مصنوعی بالاترین مقادیر دقت کلی و شاخص کاپا را نشان داد و مدل‌های رگرسیون لجستیک ترکیبی (MLR)، تحلیل درخت تصمیم و آنالیز تشخیصی در رتبه‌های بعدی واقع شدند. برای پیش‌بینی فامیل‌های خاک، مدل رگرسیون لجستیک ترکیبی (MLR) بالاترین مقادیر دقت کلی (۰/۹۸) و شاخص کاپا (۰/۹۷) را نشان داد و مدل‌های شبکه عصبی مصنوعی، آنالیز تشخیصی و تحلیل درخت تصمیم به ترتیب در رتبه‌های بعدی قرار گرفتند (جدول ۳).

با وجودیکه مقادیر OA و K برای کافلد ۵ تصادفی بالاتر از کافلد ۱۰ مکانی است، بکار بردن روش ارزیابی کافلد ۱۰ مکانی بخاطر مکانی بودنش بر استفاده از روش ارزیابی کافلد ۵ تصادفی که مکانی نیست، ارجحیت دارد (Valavi et al., 2018). روش ارزیابی کافلد ۱۰ مکانی، روش دقیقی است که قابلیت تعمیم مدل‌های پیش‌بینی کننده را بهتر مورد ارزیابی قرار می‌دهد.

دقت پایین مدل آنالیز تشخیصی (DA) در روش اعتبارسنجی کافلد ۵ (جدول ۵) به این دلیل است که این مدل، نقشه ژئومورفولوژی را به‌عنوان ورودی قبول نمی‌کند. مهم‌ترین متغیر کمکی برای پیش‌بینی کلاس‌های رده تا فامیل خاک، نقشه ژئومورفولوژی است (جدول ۳) که با دقت بسیار بالایی نقشه آن تهیه شد، بنابراین اگر این متغیر وارد مدل نشود، مطمئناً دقت پیش‌بینی پایین می‌آید.

متغیرهای کمکی مهم در پیش‌بینی رده‌ی خاک به ترتیب اهمیت در جدول ۳ ملاحظه می‌گردند. این متغیرها به ترتیب اهمیت عبارت‌اند از: ژئومورفولوژی، شیب، شاخص خیزی، فاصله عمودی شبکه آبراهه، مدل رقومی ارتفاع، نسبت پهنای دره به عمق آن، شیب حوضه، شاخص گیاهی تعدیل‌کننده اثر خاک، شاخص گیاهی تفاضلی نرمال شده و شاخص همواری قله با درجه تفکیک بالا.

در منطقه مورد مطالعه، مهم‌ترین عامل محیطی تأثیرگذار بر پیش‌بینی رده خاک، نقشه ژئومورفولوژی است (جدول ۳). اینسپتی‌سول‌ها و مالی‌سول‌ها در زمین‌نمای پای کوه و انتی‌سول‌ها در زمین‌نمای کوه تشکیل شده‌اند (شکل ۵).

جدول ۵- نتایج ارزیابی صحت پیش‌بینی سطوح طبقه‌بندی خاک مورد مطالعه، با استفاده از روش اعتبارسنجی کافلد ۵ تصادفی

کلاس	شاخص صحت	DA	DTA	ANN	MLR
رده	صحت کلی	۰/۴۸	۰/۹۱۲	۰/۹۸۴	۰/۹۸۴
	ضریب کاپا	۰/۴۵۸	۰/۸۴۱	۰/۹۷۱	۰/۹۷۱
زیررده	صحت کلی	۰/۸۲۴	۰/۹۱۲	۰/۹۷۶	۰/۹۹۲
	ضریب کاپا	۰/۷	۰/۸۴۱	۰/۹۵۷	۰/۹۸۶
گروه بزرگ	صحت کلی	۰/۷۹۲	۰/۸۸	۰/۹۷۶	۰/۹۸۴
	ضریب کاپا	۰/۶۷۵	۰/۸۰۸	۰/۹۵۹	۰/۹۷۳
زیرگروه	صحت کلی	۰/۷۵۲	۰/۷۶۸	۰/۹۶۸	۰/۸۷۲
	ضریب کاپا	۰/۶۷۹	۰/۶۶۷	۰/۸۶	۰/۸۲۷
فامیل	صحت کلی	۰/۷۷۶	۰/۶۷۲	۰/۸۸	۰/۹۸۴
	ضریب کاپا	۰/۷۵	۰/۶	۰/۸۶	۰/۹۷

جعفری و همکاران (۲۰۱۲) معتقدند که ژئوپدولوژی یک منبع مفید اطلاعات برای سنجش مواد مادری و سایر اطلاعات خاک بخصوص در مناطق خشک است (Jafari et al., 2012). این محققین نقشه ژئومورفولوژی را مهم‌ترین تخمین‌گر در پیش‌بینی کلاس‌های خاک در خاک‌های منطقه زرنند کرمان یافتند. اختلاف در مواد مادری و تکامل خاک به‌طور عمده‌ای به‌وسیله واحدهای ژئومورفولوژی مختلف نشان داده می‌شود (Stoorvogel et al., 2009). تقی زاده مهرجردی و همکاران (۲۰۱۴) از نقشه ژئومورفولوژی در تهیه نقشه شوری خاک به‌عنوان متغیر کمکی استفاده نمودند (Taghizadeh-Mehrjardi et al., 2014). کمپن و همکاران (۲۰۰۹) واحدهای ژئومورفولوژی را به عنوان داده‌های متغیر کمکی برای به‌روز کردن نقشه‌های خاک ۱:۵۰,۰۰۰ در هلند بکار بردند (Kempen et al., 2009).

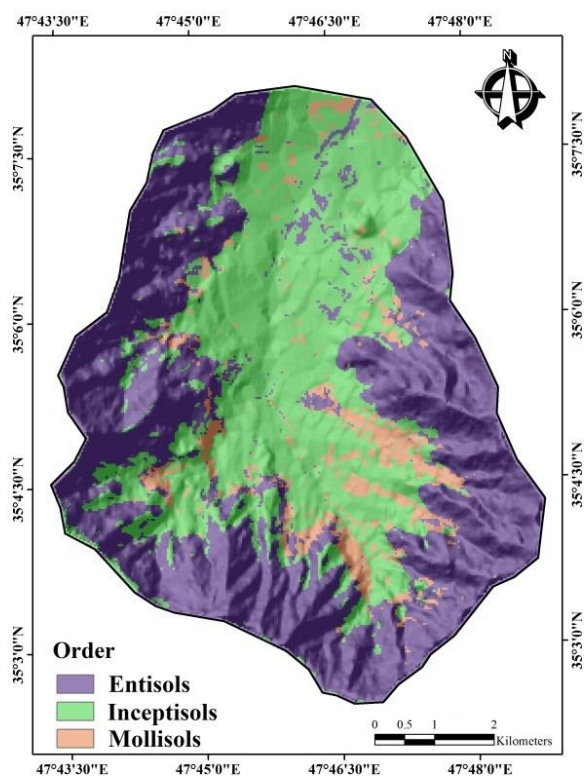
بعد از زمین‌نما، مهم‌ترین عامل محیطی مؤثر بر رده خاک، شیب است (جدول ۳). در زمین‌نمای پای کوه، به خاطر شیب کمتر، آب بیشتری در خاک نفوذ می‌کند و در نتیجه شاخص خیسی در این زمین‌نما بیشتر می‌شود. بعد از زمین‌نما و شیب، مهم‌ترین متغیر اثرگذار بر رده خاک، شاخص خیسی است (جدول ۳). در نقاطی که خاک‌های رده‌های اینسپتی‌سولز و مالی‌سولز وجود دارند، فاصله عمودی شبکه آبراهه کمتر و هر جا که خاک‌های رده انتی‌سولز دیده می‌شوند، این فاصله بیشتر است. فاصله عمودی (ارتفاع سطح پایه) شبکه آبراهه بعد از شاخص خیسی، مهم‌ترین عامل محیطی مؤثر بر رده خاک است (جدول ۳). ارتفاع (DEM) خاک‌های رده‌های اینسپتی‌سولز و مالی‌سولز، کمتر و ارتفاع خاک‌های رده انتی‌سولز از سطح دریا بیشتر است. بعد از فاصله عمودی شبکه آبراهه، مهم‌ترین متغیر اثرگذار بر رده خاک، مدل رقومی ارتفاع است (جدول ۱). هر جا که اینسپتی‌سولز و مالی‌سولز مشاهده می‌شود، نسبت عرض دره به عمق آن بیشتر و برعکس، در محل انتی‌سولها، این نسبت کمتر است. عمق دره بعد از مدل رقومی ارتفاع بااهمیت‌ترین متغیر کمکی در پیش‌بینی رده خاک است. شیب حوضه در زمین‌نمایی که اینسپتی‌سولز و مالی‌سولز قرار دارند (پای کوه)، کمتر و جایی که انتی‌سول وجود دارد (کوه)، بیشتر است. شیب حوضه بعد از عمق دره، مهم‌ترین عامل محیطی مؤثر بر رده خاک است (جدول ۱). اینسپتی‌سولز و مالی‌سولز در زمین‌نمایی (پای کوه) تشکیل شده‌اند که پوشش گیاهی (SAVI و NDVI) متراکم‌تر است. در زمین‌نمای انتی‌سولها (کوه)، پوشش گیاهی کمتر است. همواری قله بلندی‌ها با درجه تفکیک بالا کم‌اهمیت‌ترین متغیر تأثیرگذار بر رده خاک تعیین شده است (جدول ۳). هر جا که اینسپتی‌سول و مالی‌سول تشکیل شده‌اند، این همواری بیشتر (پای کوه) و در محل تشکیل انتی‌سولها، همواری مورد اشاره کمتر است (کوه).

به جز پوشش گیاهی، بقیه متغیرهای کمکی تأثیرگذار بر پیش‌بینی رده خاک از نوع پستی‌وبلندی هستند (جدول ۳). مک براتنی و همکاران (۲۰۰۳) گزارش کردند، در مطالعات انجام‌شده در مورد پیش‌بینی رده‌های خاک، مهم‌ترین متغیر محیطی بکار گرفته‌شده، پستی‌وبلندی بوده است (McBratney et al., 2003). این مشخصه زمین، توزیع آب در عمق و سطح یک زمین‌نما و در نتیجه تشکیل خاک را تحت تأثیر قرار می‌دهد (Moore et al., 1993).

متغیرهای کمکی مهم در پیش‌بینی زیررده‌ی خاک، همان متغیرهای کمکی مهم در پیش‌بینی رده‌های خاک هستند (جدول ۳) علت این امر این است که محل وجود رده و زیررده در منطقه مطالعاتی یکسان است.

نقشه های رقومی رده ها و زیررده های خاک های منطقه مطالعاتی، به ترتیب در شکل های ۵ و ۶ نشان داده شده اند. در تهیه این نقشه ها و نقشه های سایر کلاس های رده بندی خاک، فقط از مدل رگرسیون لجستیک ترکیبی که به عنوان بهترین مدل جهت پیش بینی کلاس های طبقه بندی خاک شناخته شد، استفاده گردید.

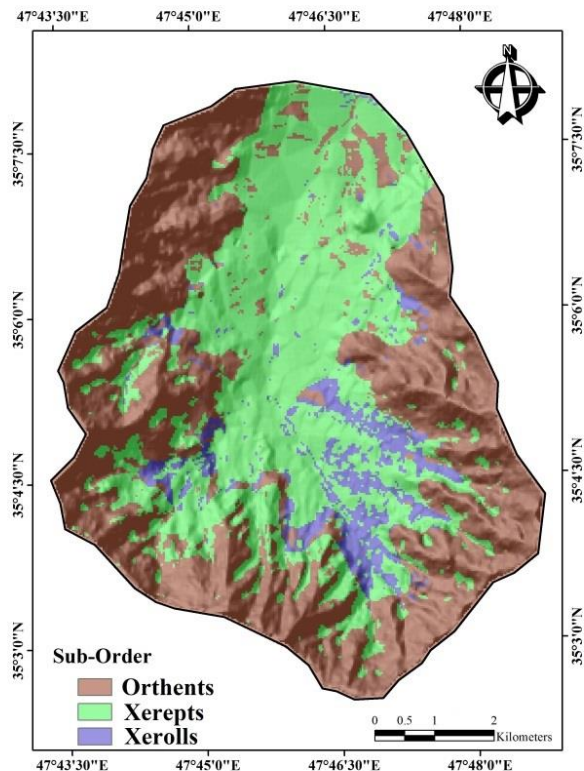
خاک های زمین نمای کوه (شکل ۳) در غرب، جنوب و شرق حوضه آبخیز مورد مطالعه در رده انتی سولز قرار گرفته اند. با نزدیک شدن به مرکز و شمال حوضه، رده خاک ها از انتی سولز به اینسپتی سولز در زمین نمای پای کوه (شکل ۳) تغییر می یابد. مالی سولز ها اغلب در جنوب شرقی حوضه در اراضی پای کوهی دیده می شوند (شکل ۵).



شکل ۵- ترسیم نقشه پیش بینی رده های خاک با استفاده از مدل رگرسیون لجستیک ترکیبی

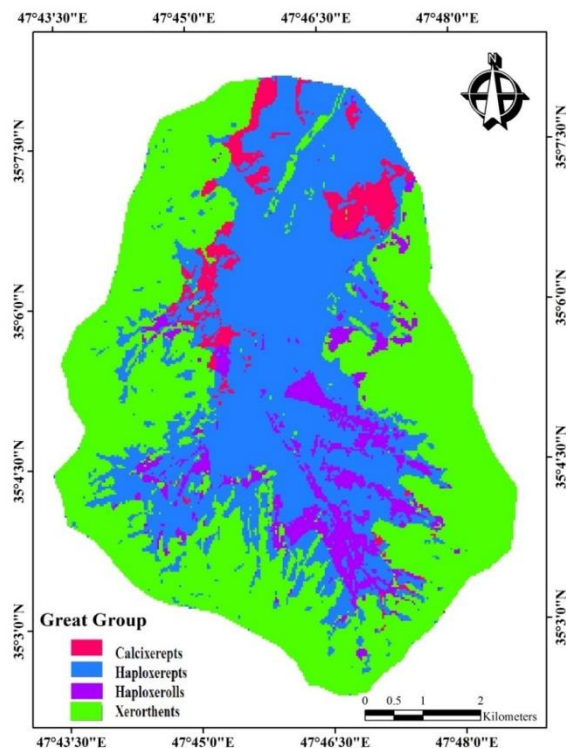
زیررده های خاک ها در همان محل های رده هایشان قرار دارند (شکل ۶).

متغیرهای کمکی مهم در پیش بینی گروه های بزرگ خاک به ترتیب اهمیت عبارتند از: ژئومورفولوژی، موقعیت نسبی شیب، شاخص خیزی، فاصله عمودی شبکه آبراهه، مدل رقومی ارتفاع، شیب حوضه، شاخص گیاهی تعدیل کننده اثر خاک، شاخص گیاهی تفاضلی نرمال شده و شاخص همواری قله با درجه تفکیک بالا (جدول ۳).



شکل ۶- ترسیم نقشه پیش‌بینی زیررده‌های خاک با استفاده از مدل رگرسیون لجستیک ترکیبی

نقشه رقومی گروه بزرگ خاک‌های منطقه مطالعاتی، در شکل ۷ نشان داده شده است. در شمال غرب و غرب منطقه مطالعاتی، گروه بزرگ Calcixerepts در پای کوه‌هایی قرار گرفته است که جنس سنگ آن‌ها مرمر سفید تا کرم (TRJ11) و مرمر خاکستری تا سیاه (TRJ11) است (شکل ۱). در شمال شرق، جنس سنگ کوه‌های بالادست این گروه بزرگ، تناوب مرمر و آمفیبولیت (TRamf)، مجموعه‌ای از سنگ‌های دگرگونی و آذرین، مشتمل بر سنگ مرمر (TRcom) و مجموعه دولومیت و سنگ‌آهک (TRJd) است. این ارتباط حاکی از آن است که آهکی بودن و تشکیل افق کلسیک در این خاک‌ها، متأثر از وجود سنگ‌های آهکی، به‌ویژه مرمر (سنگ‌آهک دگرگون‌شده) در کوه‌های بالای سر این خاک‌هاست؛ بنابراین بایستی یکی از متغیرهای کمکی تأثیرگذار بر تشکیل این گروه بزرگ، سنگ مادر آهکی (زمین‌شناسی (سنگ‌شناسی) باشد. این در حالی است که در تحقیق حاضر، نوع سنگ مادر به‌عنوان یک متغیر کمکی در مدل‌های پیش‌بینی کننده وارد نشده است (جدول ۳).



شکل ۷- ترسیم نقشه پیش بینی گروه های بزرگ خاک با استفاده از مدل رگرسیون لجستیک ترکیبی

با عنایت به اینکه Calcixerept ها برخلاف Xerorthent ها در پای کوه (در جدول ۳، مهم ترین متغیر کمکی مؤثر در تشکیل خاک در سطح گروه بزرگ، ژئومورفولوژی است) واقع شده اند و در این موقعیت، شیب زمین کمتر از شیب کوه است (در جدول ۳، بعد از ژئومورفولوژی، مهم ترین متغیر کمکی تأثیرگذار بر گروه بزرگ، موقعیت نسبی شیب است)، لذا، آب بیشتری در خاک نفوذ می کند و در نتیجه با شسته شدن کربنات کلسیم از افق های فوقانی به طرف افق های تحتانی، افق کلسیک در عمق کمتر از ۱۰۰ سانتیمتر تشکیل و Calcixerept به وجود می آید.

Haploxerept ها نیز در پای کوه (در جدول ۳، مهم ترین متغیر کمکی مؤثر در تشکیل خاک در سطح گروه بزرگ، ژئومورفولوژی است) قرار گرفته اند و در این موقعیت، شیب زمین کمتر از شیب کوه است (در جدول ۳، بعد از ژئومورفولوژی، مهم ترین متغیر کمکی تأثیرگذار بر گروه بزرگ، موقعیت نسبی شیب است)، لذا، آب بیشتری در خاک نفوذ می کند و در نتیجه با انجام تخریب شیمیایی بیشتر، افق کمبیک تشکیل و Haploxerept حادث می شود. وسعت خاک های Haploxerept خیلی بیشتر از خاک های Calcixerept است؛ زیرا همان گونه که اشاره شد، تشکیل Calcixerept

ها، نیاز به وجود منبع کربنات کلسیم دارد که در منطقه مورد مطالعه اغلب سنگ مرمر است. نظر به اینکه این سنگ فقط در کوه‌های شمال غرب و شمال شرق حوضه آبخیز بدر وجود دارند، خاک‌های Calcixercept فقط در پای این کوه‌ها دیده می‌شوند. در اراضی پای کوه، هر جا که سنگ مادر آهکی در کوه‌های مشرف بر خاک‌ها نباشد، گروه‌های بزرگ Haploxerept و Haploxeroll تشکیل می‌شوند. وسعت خاک اول از خاک دوم بیشتر است.

Haploxeroll ها اغلب در جنوب شرقی منطقه مطالعاتی مشاهده می‌شوند. به نظر می‌رسد این خاک‌ها در نقاطی تشکیل شده‌اند که پوشش گیاهی متراکم‌تر است. پوشش گیاهی بیشتر، مواد آلی بیشتر و تشکیل افق مالیک را باعث می‌شود. شاخص گیاهی تعدیل‌کننده اثر خاک و شاخص گیاهی تفاضلی نرمال شده جزو متغیرهای کمکی مهم تأثیرگذار بر تشکیل چهار گروه بزرگ موجود در منطقه مطالعاتی قلمداد شده ولی تقریباً کمترین درجه اهمیت را نسبت به سایر متغیرها دارند (جدول ۳). به نظر می‌رسد که این متغیرها باید برای Haploxeroll ها از اهمیت بیشتری برخوردار باشند. همان‌گونه که در جدول ۳ ملاحظه می‌گردد، چنین نیست و آن‌ها برای همه چهار گروه بزرگ از درجه اهمیت یکسانی برخوردار می‌باشند.

با توجه به اینکه در همه محدوده مطالعاتی، رژیم رطوبتی خاک، زریک است، بنابراین گروه بزرگ Xerorthents در همان محل زیررده Orthents و رده Entisols، یعنی در زمین‌نمای کوه دیده می‌شود (شکل ۶).

متغیرهای کمکی مهم در پیش‌بینی زیرگروه خاک به ترتیب اهمیت عبارت‌اند از: ژئومورفولوژی، موقعیت نسبی شیب، سطح‌مبنای شبکه آبراهه (ارتفاع سطح پایه شبکه آبراهه)، شیب، مدل رقومی ارتفاع، شاخص خیسی، عامل طول و تندی شیب، شیب حوضه، شاخص گیاهی تعدیل‌کننده اثر خاک و شاخص گیاهی تفاضلی نرمال شده (جدول ۳).

نقشه رقومی زیرگروه خاک‌های منطقه مطالعاتی، در شکل ۸ نشان داده شده است. با توجه به اینکه خاک‌های زیرگروه Lithic Xerorthents در زمین‌نمای کوه واقع شده‌اند، در عمق کمتر از ۵۰ سانتیمتر به سنگ می‌رسند. این خاک‌ها ۴۳/۲ درصد مساحت منطقه مورد مطالعه را شامل می‌شوند. خاک‌های Typic Xerorthents که ۱/۶٪ مساحت منطقه مطالعاتی را تشکیل می‌دهند، از عمق بیشتری نسبت به خاک‌های Lithic Xerorthents برخوردار بوده و اغلب در شمال و به میزان کمتر در غرب و جنوب غرب حوضه وجود دارند.

در شمال، مرکز و شمال شرقی منطقه مطالعاتی، به ترتیب فراوانی، زیرگروه‌های Calcic Haploxerepts (۴/۶٪)، Typic Calcixerepts (۳/۲٪) و Vertic Calcixerepts (۰/۸٪) در پای کوه‌های متشکل از سنگ‌های آهکی قرار گرفته‌اند (شکل ۸). همان‌گونه که در بخش پیش‌بینی گروه‌های بزرگ خاک نیز ذکر گردید، سنگ مادر جزو متغیرهای تأثیرگذار بر تشکیل خاک‌های فوق‌الذکر منظور نشده است (جدول ۳).

بعد از خاک‌های Lithic Xerorthents، فراوان‌ترین خاک‌ها در منطقه مطالعاتی، خاک‌های Typic Haploxerepts هستند که ۲۱/۶٪ خاک‌های منطقه را تشکیل می‌دهند. این خاک‌ها که اغلب اراضی پای کوهی کوه‌های شرقی را اشغال کرده‌اند، در پای کوه‌های با سنگ غیر آهکی مشاهده می‌شوند. به این دلیل، بجای افق کلسیک دارای افق کمبیک می‌باشند. خاک‌های Fluventic Haploxerepts که ۱۲٪ سطح منطقه مورد مطالعه را پوشش می‌دهند، در پائین‌ترین دشت‌های آبرفتی (Qt3 = Qc)، مخروط افکنه‌ها (Qf) و آبرفت‌های جدید در کنار آبراهه‌ها (Qal) وجود دارند (شکل ۱ و شکل ۳). آبرفت‌ها که بخشی از تشکیلات زمین‌شناسی منطقه به حساب می‌آیند، به‌عنوان عامل مؤثر در تشکیل این خاک‌ها، جزو متغیرهای کمکی مهم در پیش‌بینی کلاس این خاک‌ها در جدول ۳ منظور شده‌اند. به خاطر مواد مادری این خاک‌ها که آبرفتی است و کاهش نامنظم ماده آلی از عمق ۲۵ تا ۱۲۵ سانتی‌متری یا تا سنگ بستر، برای طبقه‌بندی آن‌ها در سطح زیرگروه از پیشوند Fluventic استفاده شده است.

۳/۲٪ خاک‌ها Vertic Haploxerepts هستند. این خاک‌ها که درصد رس بیشتری دارند و دارای شکاف، خاکدانه گوه‌ای و یا اسلیکنساید می‌باشند، در اراضی پای کوه‌های جنوبی، جایی که این کوه‌ها از سنگ‌های دیوریت و گابرو دیوریت دگرگون‌شده (dim)؛ مونوزدیوریت، کوآرتز مونزونیت و سینیت کوآرتز دیوریت (md) و میکا شیست، فیلیت، اسلیت و سنگ‌آهک و دولومیت (TRJph, l) تشکیل شده‌اند؛ به چشم می‌خورند. این سنگ‌ها به خاطر مقاومت کم (مثل دیوریت و گابرو) تا متوسطی (مثل مونزونیت) که نسبت به تخریب دارند و یا به علت ریزدانه بودن بافتشان (فیلیت و اسلیت)، در اثر تخریب، تولید رس می‌کنند.

### نتیجه‌گیری

یادگیری ماشین ابزار قدرتمندی برای مدل‌سازی دقیق و سریع کلاس‌های خاک است که با بهره‌گیری از داده‌های متنوع محیطی، اطلاعات مفیدی برای مدیریت زمین، کشاورزی، منابع طبیعی و حفاظت محیط زیست فراهم می‌کند. این

روش‌ها در کنار سیستم‌های اطلاعات جغرافیایی (GIS) و سنجش‌ازدور، آینده مطالعات خاک را به شدت متحول کرده‌اند. به جای صرف هزینه و زمان بالا برای نمونه‌برداری فیزیکی، می‌توان با کمک مدل‌های یادگیری ماشین، کلاس‌های خاک را در مناطق نمونه‌برداری نشده پیش‌بینی کرد. برخی کلاس‌های خاک که با روش‌های سنتی به سختی تفکیک‌پذیر هستند، با یادگیری ماشین و داده‌های ترکیبی با دقت بالاتر قابل شناسایی هستند. در برنامه‌ریزی کشاورزی، شناخت کلاس خاک به انتخاب نوع محصول، کوددهی و آبیاری کمک می‌کند و با پیش‌بینی دقیق کلاس خاک، تصمیم‌گیری‌های کشاورزی بهینه‌تر می‌شود.

جهت برآورد کلاس‌های خاک از سطح رده تا فامیل خاک، در مطالعه حاضر از مدل‌های تحلیل درخت تصمیم، آنالیز تشخیصی، شبکه عصبی مصنوعی، نزدیک‌ترین همسایه و جنگل تصادفی بهره گرفته شد. همچنین، از مدل رگرسیون لجستیک جهت ترکیب کردن نتایج سایر مدل‌ها استفاده گردید. مدل‌های بکارگرفته شده توسط دو روش ارزیابی کافلد تصادفی و کافلد مکانی مورد ارزیابی قرار گرفتند. نتایج به طور کلی نشان داد که:

(۱) متغیرهای کمکی مهم و تأثیرگذار در پیش‌بینی رده، زیررده، گروه بزرگ و زیرگروه خاک به ترتیب اهمیت عبارت‌اند از سنگ مادر و ژئومورفولوژی، پستی و بلندی و پوشش گیاهی. در پیش‌بینی فامیل خاک، همه این متغیرها با همان ترتیب اهمیت نقش دارند، به جز پوشش گیاهی. با وجودی که متغیر کمکی سنگ مادر در مدل‌های پیش‌بینی کننده وارد نشده است، ولی در نقشه‌های سطوح مختلف طبقه‌بندی خاک، اثر خود را در پیش‌بینی این سطوح به خوبی نشان می‌دهد. اگر این متغیر نیز در مدل‌ها دخالت داده می‌شد، درجه اهمیت آن بهتر مشخص می‌گردید. روش اعتبارسنجی کافلد ۱۰ مکانی، بخاطر مکانی بودنش بر روش‌های اعتبارسنجی غیر مکانی ارجحیت دارد.

(۲) بر اساس روش اعتبارسنجی کافلد ۱۰ مکانی، بهترین مدل برای پیش‌بینی اغلب سطوح طبقه‌بندی خاک، رگرسیون لجستیک ترکیبی است. برای همه سطوح طبقه‌بندی خاک، مدل جنگل تصادفی، از نظر قدرت پیش‌بینی، رتبه دوم را به خود اختصاص می‌دهد. به طور کلی، نزدیک‌ترین همسایه K، پائین‌ترین عملکرد را در پیش‌بینی سطوح طبقه‌بندی خاک دارا بوده است.

۳) بر اساس روش اعتبارسنجی کافلد ۱۰ مکانی، دقت و صحت مدل‌های رگرسیون لجستیک ترکیبی و جنگل تصادفی، به ترتیب از رده به طرف فامیل به طور منظم سیر نزولی دارند. در مورد سه مدل دیگر، این دقت و صحت، از رده به طرف زیر رده، افزایش و از زیر رده تا فامیل بطور منظم، کاهش نشان می‌دهند.

### منابع مورد استفاده

- حریری، ع. ۱۳۷۴. نگرشی بر خاستگاه گروهی از سنگ‌های دگرگونه گستره قروه. پایان‌نامه کارشناسی ارشد، دانشگاه شهید بهشتی، تهران.
- حسینی، م. ۱۳۷۶. شرح نقشه زمین‌شناسی ۱:۱۰۰۰۰۰ چهار گوش قروه (پیوست نقشه)، سازمان زمین‌شناسی و اکتشاف معدنی کشور.
- Abdi, L., & Hashemi, S. (2015). To combat multi-class imbalanced problems by means of over-sampling and boosting techniques. *Soft Computing*, 19, 3369-3385. <https://doi.org/10.1007/s00500-014-1291-z>.
- Adhikari, K. & Hartemink, A.E. (2016). Linking soils to ecosystem services-A global review. *Geoderma* 2016, 262, 101-111. <https://doi.org/10.1016/j.geoderma.2015.08.009>.
- Adhikari, K., Minasny, B., Greve, M.B., & Greve, M.H. (2014). Constructing a soil class map of Denmark based on the FAO legend using digital techniques. *Geoderma*, 214, 101-113. <https://doi.org/10.1016/j.geoderma.2013.09.023>.
- Abbaszadeh Afshar, F., Ayoubi, S., & Jafari, A. (2018). The extrapolation of soil great groups using multinomial logistic regression at regional scale in arid regions of Iran. *Geoderma*, 315, 36-48. <https://doi.org/10.1016/j.geoderma.2017.11.030>.
- Behrens, T., Förster, H., Scholten, T., Steinrücken, U., Spies, E.D., & Goldschmitt, M. (2005). Digital soil mapping using artificial neural networks. *J. Plant Nutr. Soil Sci.* 2005, 168, 21-33. <https://doi.org/10.1002/jpln.200421414>.
- Brungard, C.W., Boettinger, J.L., Duniway, M.C., Wills, S.A., & Edwards, T.C. (2015). Machine learning for predicting soil classes in three semi-arid landscapes. *Geoderma* 2015, 239, 68-83. <https://doi.org/10.1016/j.geoderma.2014.09.019>.
- Bui, E.N., & Moran, C.J. (2001). Disaggregation of polygons of surficial geology and soil maps using spatial modelling and legacy data. *Geoderma*, 103(1-2), 79-94. [https://doi.org/10.1016/S0016-7061\(01\)00070-2](https://doi.org/10.1016/S0016-7061(01)00070-2).
- Caubet, M., Dobarco, M.R., Arrouays, D., Minasny, B., & Saby, N.P. (2019). Merging country, continental and global predictions of soil texture: Lessons from ensemble modelling in France. *Geoderma*, 337, 99-110. <https://doi.org/10.1016/j.geoderma.2018.09.007>.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321-357. <https://doi.org/10.1613/jair.953>.
- Diks, C.G.H., & Vrugt, J.A. (2010). Comparison of point forecast accuracy of model averaging methods in hydrologic applications. *Stoch. Environ. Res. Risk Assess.*, 24, 809-820. <https://doi.org/10.1007/s00477-010-0378-z>.
- Dominati, E., Patterson, M., & Mackay, A. (2010). A framework for classifying and quantifying the natural capital and ecosystem services of soils. *Ecol. Econ.*, 69, 1858-1868. <https://doi.org/10.1016/j.ecolecon.2010.05.002>.
- Górecki, T., & Krzyśko, M. (2015). Regression methods for combining multiple classifiers. *Commun. Stat.-Simul. C.* 2015, 44, 739-755. <https://doi.org/10.1080/03610918.2013.794286>.
- Grinand, C., Arrouays, D., Laroche, B., & Martin, M.P. (2008). Extrapolating regional soil landscapes from an existing soil map: Sampling intensity, validation procedures, and integration of spatial context. *Geoderma*, 143(1-2), 180-190. <https://doi.org/10.1016/j.geoderma.2007.11.004>.
- He, H., & Garcia, E.A. (2009). Learning from imbalanced data. *IEEE Transactions on knowledge and data engineering*, 21(9), 1263-1284. <https://doi.org/10.1109/TKDE.2008.239>.
- Heung, B., Ho, H.C., Zhang, J., Knudby, A., Bulmer, C.E., & Schmidt, M.G. (2016). An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma* 2016, 265, 62-77. <https://doi.org/10.1016/j.geoderma.2015.11.014>.
- Jafari, A. Finke, P.A. de Wauw, J.V. Ayoubi, S. & Khademi, H. (2012). Spatial prediction of USDA- great soil groups in the arid Zard region, Iran: comparing logistic regression approaches to predict diagnostic horizons and soil types. *E. J. Soil Sci*, 63: 284-298. <https://doi.org/10.1111/j.1365-2389.2012.01425.x>.
- Jenny, H. (1994). *Factors of soil formation: a system of quantitative pedology*. Courier Corporation.

- Kempen, B. Brus, D.J. Heuvelink, G.B.M., & Stoorvogel, J.J. (2009). Updating the 1:50000 Dutch soil map using legacy soil data: A multinomial logistic regression approach. *Geoderma* 151, 311-326. <https://doi.org/10.1016/j.geoderma.2009.04.023>.
- Kovačević, M., Bajat, B., & Gajić, B. (2010). Soil type classification and estimation of soil properties using support vector machines. *Geoderma*, 154, 340-347. <https://doi.org/10.1016/j.geoderma.2009.11.005>.
- Lemercier, B., Lacoste, M., Loum, M., & Walter, C. (2012). Extrapolation at regional scale of local soil knowledge using boosted classification trees: A two-step approach. *Geoderma*, 171, 75-84. <https://doi.org/10.1016/j.geoderma.2011.03.010>.
- Ma, Y.X., Minasny, B., Malone, B.P., & McBratney, A.B. (2019). Pedology and digital soil mapping (DSM). *Eur. J. Soil Sci.* 2019, 70, 216-235. <https://doi.org/10.1111/ejss.12790>.
- Mallavan, B. P., Minasny, B., & McBratney, A.B. (2010). Homosoil, a methodology for quantitative extrapolation of soil information across the globe. *Digital soil mapping: bridging research, environmental application, and operation*, 137-150. [https://doi.org/10.1007/978-90-481-8863-5\\_12](https://doi.org/10.1007/978-90-481-8863-5_12).
- Malone, B. P., Jha, S. K., Minasny, B., & McBratney, A.B. (2016). Comparing regression-based digital soil mapping and multiple-point geostatistics for the spatial extrapolation of soil data. *Geoderma*, 262, 243-253. <https://doi.org/10.1016/j.geoderma.2015.08.037>.
- Malone, B.P., Minasny, B., Odgers, N.P., & McBratney, A.B. (2014). Using model averaging to combine soil property rasters from legacy soil maps and from point data. *Geoderma* 2014, 232, 34-44. <https://doi.org/10.1016/j.geoderma.2014.04.033>.
- McBratney, A.B., Santos, M.M., & Minasny, B. (2003). On digital soil mapping. *Geoderma* 2003, 117, 3-52. [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4).
- Minasny, B., & McBratney, A.B. (2016). Digital soil mapping: A brief history and some lessons. *Geoderma* 2016, 264, 301-311. <https://doi.org/10.1016/j.geoderma.2015.07.017>.
- Moore, I.D. Gessler, P. Nielsen, G.A. & Peterson, G.A. (1993). Soil attributes prediction using terrain analysis. *Soil Science Society of America Journal* 57: 443-452. <https://doi.org/10.2136/sssaj1993.03615995005700020026x>.
- Padarian, J., Minasny, B., McBratney, A.B., & Dalgliesh, N. (2014). Predicting and mapping the soil available water capacity of Australian wheatbelt. *Geoderma Reg.* 2014, 2, 110-118. <https://doi.org/10.1016/j.geodrs.2014.09.005>.
- Pahlavan-Rad, M.R., Khormali, F., Toomanian, N., Brungard, C.W., Kiani, F., Komaki, C.B., & Bogaert, P. (2016). Legacy soil maps as a covariate in digital soil mapping: A case study from Northern Iran. *Geoderma* 2016, 279, 141-148. <https://doi.org/10.1016/j.geoderma.2016.05.014>.
- Powers, D. M. (2020). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. arXiv preprint arXiv: 2010.16061.
- Román Dobarco, M., Arrouays, D., Lagacherie, P., Ciampalini, R., & Saby, N.P.A. (2017). Prediction of topsoil texture for Region Centre (France) applying model ensemble methods. *Geoderma* 2017, 298, 67-77. <https://doi.org/10.1016/j.geoderma.2017.03.015>.
- Schoeneberger, P.J. Wysocki, D.A. Benham, E.C. Soil Survey Staff. (2012). Field book for describing and sampling soils. 3rd Version. Natural Resources Conservation Service. National Soil Survey Center. Lincoln, NE.
- Scull, P., Franklin, J., & Chadwick, O.A. (2005). The application of classification of tree analysis to soil type prediction in a desert landscape. *Eco. Model.* 181: 1-15. <https://doi.org/10.1016/j.ecolmodel.2004.06.036>.
- Shariffar, A., Sarmadian, F., Alikhani, H., Keshavarzi, A., Asghari, O., & Malone, B. P. (2019a). Lateral and vertical variations of soil organic and inorganic carbon content in Aridisols and Entisols of a rangeland. *Eurasian Soil Science*, 52, 1051-1062. <https://doi.org/10.1134/S1064229319090084>.
- Shariffar, A., Sarmadian, F., Malone, B. P., & Minasny, B. (2019b). Addressing the issue of digital mapping of soil classes with imbalanced class observations. *Geoderma*, 350, 84-92. <https://doi.org/10.1016/j.geoderma.2019.05.016>.
- Soil Survey Staff. (2014). *Soil Taxonomy: A basic system of soil classification for making and interpreting soil surveys*. 12th Edition. NRCS. USDA. doi:10.1016/S0016-7061(00)00097-5
- Stoorvogel, J., Kempen, J., Heuvelink, B., & Bruin, S. (2009). Implementation and evaluation of existing knowledge for digital soil mapping in Senegal. *Geoderma* 149:161-170. <https://doi.org/10.1016/j.geoderma.2008.11.039>.
- Swiderski, B., Osowski, S., Kruk, M., & Barhoumi, W. (2016). Aggregation of classifiers ensemble using local discriminatory power and quantiles. *Expert Syst. Appl.* 2016, 46, 316-323. <https://doi.org/10.1016/j.eswa.2015.10.038>.
- Taghizadeh-Mehrjardi, R., Sarmadian, F., Minasny, B., Triantafyllis, J. & Omid, M. (2014). Digital mapping of soil classes using decision tree and auxiliary data in the Ardakan region, Iran. *Arid Land Research and Management*, 28(2), 147-168. <https://doi.org/10.1080/15324982.2013.828801>.

- Taghizadeh-Mehrjardi, R., Nabiollahi, K., Minasny, B., & Triantafyllis, J. (2015). Comparing data mining classifiers to predict spatial distribution of USDA-family soil groups in Baneh region, Iran. *Geoderma* 2015, 253, 67–77. <https://doi.org/10.1016/j.geoderma.2015.04.008>.
- Valavi, R., Elith, J., José, J., Lahoz, M., & Gurutzeta, G. (2018). Block CV: an R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Biorxiv*. <https://doi.org/10.1101/2041-210X.13107>.
- Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. *Fam med*, 37(5), 360-363. <https://doi.org/10.1370/afm.1182>.
- Wasson, T., & Hartemink, A.J. (2009). An ensemble model of competitive multi-factor binding of the genome. *Genome Res.*, 19, 2101–2112. <https://doi.org/10.1101/gr.093450.109>.
- Zeraatpisheh, M., Ayoubi, S., Jafari, A., & Finke, P. (2017) Comparing the efficiency of digital and conventional soil mapping to predict soil types in a semi-arid region in Iran. *Geomorphology* 2017, 285, 186–204. <https://doi.org/10.1016/j.geomorph.2017.02.015>.
- Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision* (pp. 2223-2232).
- Zinck, J.K. (1989). *Physiography and Soils: Soil Survey Courses Subject Matter K6*. Soil Division. ITC. Enschede, The Netherlands. 161 p.