



## ارزیابی تعمیم‌پذیری مدل جنگل تصادفی برای پیش‌بینی کلاس‌های خاک در سطح زیر گروه

محمد جمشیدی<sup>۱</sup>، \* محمدامیر دلاور<sup>۲</sup>، روح‌اله تقی‌زاده مهرجردی<sup>۳</sup> و کلبی برانگارد<sup>۴</sup>

<sup>۱</sup>دانشجوی دکتری گروه علوم خاک، دانشگاه زنجان و مربی پژوهش مؤسسه تحقیقات خاک و آب، سازمان تحقیقات، آموزش و ترویج کشاورزی،

<sup>۲</sup>دانشیار گروه علوم خاک، دانشگاه زنجان، <sup>۳</sup>استادیار گروه علوم خاک، دانشگاه اردکان، <sup>۴</sup>استادیار گروه علوم محیطی، دانشگاه نیومکزیکو

تاریخ دریافت: ۹۷/۸/۲؛ تاریخ پذیرش: ۹۷/۱۰/۲۲

### چکیده

**سابقه و هدف:** در ایران بیش‌تر نقشه‌های خاکی که طی شش دهه گذشته با روش سنتی تهیه شده‌اند، به دلیل نیاز به هزینه و زمان زیاد به روز نشده‌اند. در سال‌های اخیر روش نقشه‌برداری رقومی خاک با هدف تهیه نقشه‌های دقیق بر مبنای مدل‌سازی کمی روابط بین کلاس‌ها یا ویژگی‌های خاک و داده‌های کمکی ارزان نماینده عوامل خاک‌سازی به‌عنوان روش جایگزین روش‌های سنتی معرفی شده است. در این رابطه یک روش استفاده از مدل‌سازی داده‌های موجود خاک در یک منطقه برای پیش‌بینی کلاس‌های خاک در یک منطقه دیگر فاقد نقشه خاک است. این مطالعه با هدف ارزیابی تعمیم‌پذیری مدل جنگل تصادفی مستخرج از منطقه دهنده اطلاعات برای تهیه نقشه کلاس‌های خاک در سطح زیرگروه برای منطقه گیرنده انجام شد.

**مواد و روش‌ها:** دشت سعادت‌شهر استان فارس به‌عنوان منطقه دهنده یا مرجع و دشت سیدان در مجاور آن به‌عنوان منطقه گیرنده اطلاعات با توجه به شباهت متغیرهای محیطی در دو منطقه انتخاب شدند. در منطقه دهنده، موقعیت ۸۲ بدون مشاهده‌ای بر اساس روش مربعات لاتین تعیین و طی مطالعه خاکشناسی مطابق سامانه رده‌بندی خاک آمریکایی (۲۰۱۴) طبقه‌بندی شدند. ۲۵ متغیر کمکی پستی و بلندی و داده‌های سنجش از دور با قدرت تفکیک ۳۰ متر در مطالعه استفاده شدند. در منطقه دهنده با کاربرد روش جنگل تصادفی برای مدل‌سازی روابط بین کلاس‌های خاک و متغیرهای کمکی مهم، ۷۰ درصد بدون‌ها برای آموزش مدل و ۳۰ درصد برای آزمون به‌کار رفتند. در منطقه گیرنده با کاربرد متغیرهای کمکی مشابه در مدل‌های استخراج شده از منطقه دهنده (یکی با ۷۰ درصد و دیگری با ۱۰۰ درصد داده‌ها)، کلاس‌های پیش‌بینی شده با نتایج طبقه‌بندی ۲۷ خاک‌رخ مطالعه شده در این منطقه مقایسه و صحت مدل‌ها در تعمیم داده‌ها مورد ارزیابی قرار گرفت. توانایی مدل در پیش‌بینی کلاس‌های خاک در هر دو منطقه بر اساس نتایج آماره‌های صحت کاربر، صحت تولیدکننده، صحت کلی و ضریب کاپا مورد ارزیابی قرار گرفت.

**یافته‌ها:** نتایج نشان داد از میان تمامی عوامل محیطی استفاده شده، شیب، شاخص همواری دره با درجه تفکیک بالا، شاخص ناهمواری‌های توپوگرافی، شاخص خیسی توپوگرافی و مساحت حوزه اصلاح شده بیش‌ترین تأثیر را در پیش‌بینی کلاس‌های خاک در سطح زیرگروه داشته‌اند. دستیابی به مقادیر ۷۲ درصد برای صحت کلی و ۰/۵۹ برای

\* مسئول مکاتبه: [amir-delavar@znu.ac.ir](mailto:amir-delavar@znu.ac.ir)

ضریب کاپا در پیش‌بینی کلاس‌های خاک بیانگر ارتباط خوب بین داده‌های مشاهده‌ای و پیش‌بینی در منطقه مرجع بود. در منطقه تعمیم، با استفاده از ۷۰ درصد داده‌های آموزشی منطقه دهنده، صحت کلی و ضریب کاپا به ترتیب ۴۵ درصد و ۰/۲۷ و با کاربرد ۱۰۰ درصد داده‌ها، علاوه بر پیش‌بینی یک کلاس بیش‌تر در منطقه گیرنده، صحت کلی و ضریب کاپا به ترتیب تا ۵۲ درصد و ۰/۳۸ بهبود نشان داد. در بین زیرگروه‌های خاک، بهترین پیش‌بینی مربوط به خاک غالب تیپیک کلسی‌زرتیز و تیپیک زراورتنز و ضعیف‌ترین آن مربوط به خاک‌های مشابه کلاس‌های غالب بود. علاوه بر آن نتایج گویای آن است که مدل قادر به پیش‌بینی کلاس‌های دارای فراوانی ناچیز در هر دو منطقه گیرنده و دهنده نبود.

**نتیجه‌گیری:** این پژوهش نشان داد در مناطقی از ایران که فاقد نقشه خاک بوده یا نقشه‌های موجود به روز نشده است، انتقال مدل‌های ساخته شده بر پایه نقشه‌برداری رقومی در مناطق مشابه دارای داده‌های کافی می‌تواند ابزاری کارآمد برای تهیه نقشه خاک در این مناطق باشد. صرفه‌جویی در هزینه و زمان و دقت قابل قبول، می‌تواند مشوق‌های اصلی استفاده از این روش توسط خاک‌شناسان باشد.

**واژه‌های کلیدی:** روش مربعات لاتین، عوامل خاک‌سازی، نقشه‌برداری رقومی خاک

#### مقدمه

مطالعات معمول یا سنتی خاک‌شناسی، مجموعه‌ای از عملیات سیستماتیک میدانی و ستادی است که به منظور فهرست‌برداری و تعیین پراکنش انواع خاک‌ها بر مبنای تجزیه و تحلیل روابط خاک-زمین‌نما در مناطق مطالعاتی انجام می‌شود و اطلاعات و داده‌های آن در قالب نقشه خاک و گزارش مطالعه، منتشر می‌گردد (۳۵). نقشه‌های خاک در این نوع مطالعات با توجه به این‌که بر اساس مدل ذهنی خاک‌شناس و به صورت پولیگونی<sup>۱</sup> تهیه می‌شوند، برای کاربرانی که نیاز به اطلاعات دقیق از خاک‌ها دارند از دقت کافی برخوردار نیست (۳۶). از طرف دیگر انجام این مطالعات به دلیل هزینه‌بر و زمان‌بر بودن (۲۵) همواره با محدودیت روبرو است.

برای فائق آمدن بر محدودیت‌های نقشه‌های سنتی خاک، در چند دهه اخیر پیشرفت‌های فناوری سامانه‌های اطلاعات مکانی، سنجش از دور و نرم‌افزارهای رایانه‌ای، فرصتی برای خاک‌شناسان

فراهم آورده‌اند تا با استفاده از روش‌های داده‌کاوی، مفاهیم تئوریک تشکیل خاک‌ها (۱۵) را کمی نموده و بر اساس آن خصوصیات یا کلاس‌های خاک را پیش‌بینی نمایند. این روش‌ها در سال ۲۰۰۳ تحت عنوان روش‌های نقشه‌برداری رقومی خاک<sup>۲</sup> معرفی شدند (۲۱).

در بین مطالعات بسیاری که تاکنون بر مبنای روش‌های نقشه‌برداری رقومی خاک با هدف پیش‌بینی خصوصیات یا کلاس‌های خاک انجام شده، مواردی از آن در ارتباط با ارزیابی تعمیم‌پذیری<sup>۳</sup> مدل‌ها و اطلاعات خاک از یک منطقه به منطقه فاقد داده‌های خاک بوده است (۱، ۱۱، ۱۹ و ۲۰). این روش مبتنی بر ساخت مدل خاک-زمین‌نما در منطقه دهنده و تعمیم نتایج مدل به منطقه دیگر برای تولید نقشه توزیع واحدهای خاک است (۱۱). بر اساس پژوهش‌های لاگاچری (۲۰۰۲) نقشه‌های پیش‌بینی خاک-زمین‌نما را می‌توان با دو روش انجام داد:

2- Digital soil mapping, DSM

3- Extrapolation

1- Polygon-based

است (۳۴). نتایج مطالعه گریناند و همکاران (۲۰۰۸) نشان داد که صحت نقشه‌های تولیدشده با مدل‌سازی<sup>۵</sup> داده‌های منطقه مرجع به مراتب بیش‌تر از نقشه‌های تولیدشده در منطقه تعمیم<sup>۶</sup> با استفاده از همان مدل بوده است. این کاهش دقت مدل‌ها در پیش‌بینی کلاس‌های خاک در منطقه تعمیم حتی در استفاده از داده‌های مشاهداتی جدید، در سطوح پایین‌تر طبقه‌بندی به‌ویژه از زیرگروه تا سطح فامیل شدیدتر بود. کاهش دقت از آن‌جا منشأ می‌گیرد که مدل‌سازی ارتباط بین خصوصیات جداکننده خاک‌ها در این سطوح و عوامل محیطی دشوار است. همچنین از آن‌جا که خاک‌ها در مناطق جغرافیایی مختلف دارای تأثیر متفاوت عوامل خاک‌سازی بوده و از پیچیدگی ویژگی‌های مکانی برخوردار هستند، این کاهش دقت دور از انتظار نیست (۱۱). نتایج اغلب مطالعات مربوط به تعمیم داده‌های خاک، نیاز به انجام پژوهش‌های بیش‌تر در رابطه با فراهم آوردن روش‌ها و تکنیک‌های کاربردی‌تر برای ارتقاء قابلیت تعمیم‌پذیری اطلاعات و مدل‌های ساخته‌شده از روابط خاک- زمین‌نما را برجسته نموده است (۱۱). در سال ۲۰۱۰، روش خاک‌های هم‌سنخ<sup>۷</sup> برای تهیه نقشه خاک در مناطق فاقد داده و مشابه منطقه ساخت مدل پیشنهاد شد (۱۹). اساس این مطالعات بر این فرضیه استوار است که مناطق دارای عوامل محیطی مشابه و عوامل خاک‌سازی مشابه، احتمالاً دارای خاک‌های مشابه نیز خواهند بود. بنابراین گام نخست، تعیین میزان شباهت بین خصوصیات محیطی (نمایندگان عوامل خاک‌سازی) در منطقه دهنده و گیرنده اطلاعات است. در عین حال که ایده خاک هم‌سنخ توانسته است بستری برای تعمیم اطلاعات و مشاهدات در دسترس از مناطق دارای اطلاعات خاک

(۱) طبقه‌بندی نظارت‌شده<sup>۱</sup> به مفهوم طراحی و ساخت مدل پیش‌بینی بر اساس مشاهدات انواع خاک‌های شناسایی‌شده<sup>۲</sup> و (۲) طبقه‌بندی نظارت‌نشده<sup>۳</sup> به معنی گروه‌بندی مشاهدات (پیکسل‌های) نمایش‌دهنده ویژگی‌های همگن بدون دانش قبلی. به‌هرحال، طبقه‌بندی نظارت‌شده معمولاً منجر به نتایج بهتری می‌شود و نقشه‌های تولید شده قابلیت بیش‌تری برای تفسیر دارند (۱۶).

موضوع کلیدی در روش‌های تعمیم داده‌های خاک، داشتن یک مدل حاصل از داده‌های جدید خاک یا حاصل از داده‌های نقشه‌های تفصیلی موجود برای دستیابی به یک پیش‌بینی مناسب از مدل خاک- زمین‌نما در منطقه دهنده و در نتیجه تخمین دقیق از کاربرد مدل در منطقه تعمیم است (۸). در روشی که در آن از داده‌های نقشه‌های موجود در منطقه دهنده استفاده می‌شود نیاز به بازیابی مدل ذهنی مطالعه‌کننده قبلی است (۴ و ۱۷).

تامپسون و همکاران (۲۰۰۶) با هدف بررسی کارایی تعمیم مدل‌های ساخته شده برای پیش‌بینی ویژگی‌های خاک در مناطق خارج از محل استخراج مدل، به ارزیابی روابط بین مقادیر شن، سیلت و کربن آلی در افق سطحی با متغیرهای اولیه و ثانویه پستی و بلندی در سه منطقه مختلف پرداختند. ایشان در این مطالعه برای دستیابی به توانایی مدل خاک- زمین‌نما، از تجزیه رگرسیون ساده، خطای میانگین و خطای ریشه میانگین مربعات<sup>۴</sup> استفاده نمودند. نتایج آن‌ها نشان داد در عین شباهت عوامل خاک‌سازی، برای تعمیم نتایج از یک منطقه به منطقه دیگر، نیاز به استفاده از سطوح دقت مناسب در استفاده از داده‌های سنجش از دور و استفاده از نقاط مشاهداتی بیش‌تر

5- Donor area  
6- Recipient area  
7- Homosoi

1- Supervised classification  
2- Recognized soil types  
3- Unsupervised classification  
4- Root mean square error, RMSE

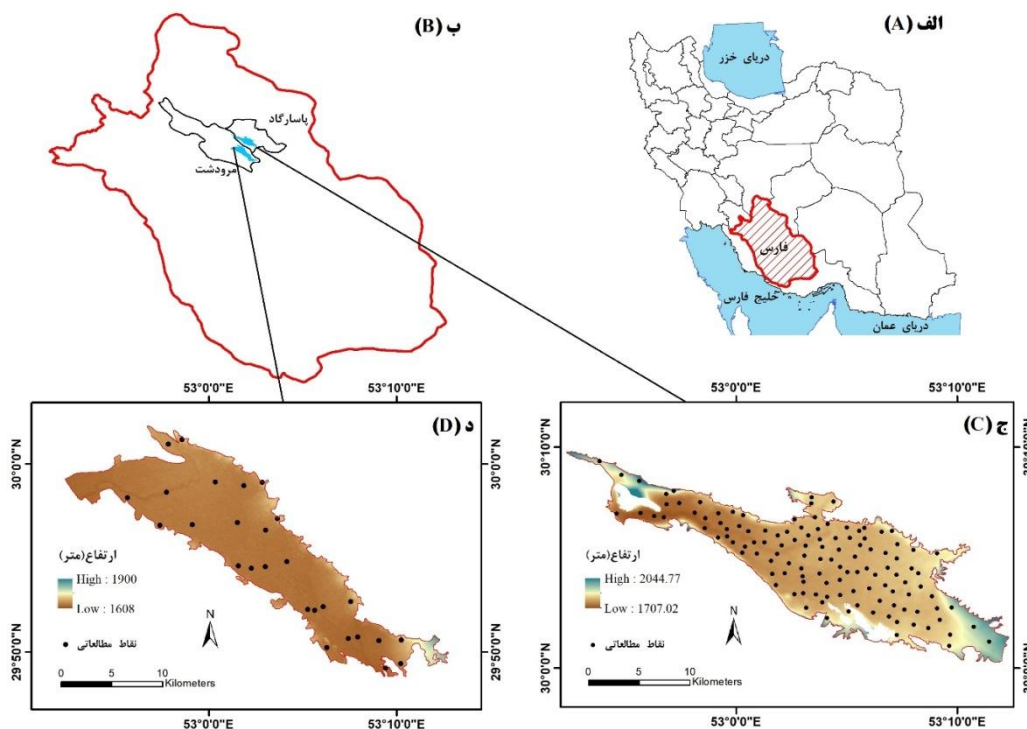
کاربران این نوع مطالعات بوده و از طرفی با توجه به محدودیت‌های روش‌های مرسوم، این مطالعه به منظور ارزیابی روش‌های جایگزین مطالعات سنتی با فرضیاتی مبتنی بر تولید داده‌های دقیق، ارزان و سریع برای مدیریت منابع خاک اجرا شده است. از این‌رو، اهداف این مطالعه عبارت هستند از: (۱) ارزیابی دقت روش جنگل تصادفی<sup>۲</sup> در تهیه نقشه کلاس‌های خاک در سطح زیرگروه در منطقه دارای داده‌های جدید خاک (منطقه مرجع) و (۲) ارزیابی توانایی تعمیم مدل به‌دست آمده در منطقه دارای داده‌های خاک به منطقه خارج از محل ساخت مدل (منطقه تعمیم).

### مواد و روش‌ها

**منطقه مورد مطالعه:** در این مطالعه دو منطقه دهنده و گیرنده اطلاعات به‌عنوان محدوده مطالعاتی انتخاب شدند (شکل ۱). منطقه دهنده دشت سعادت‌شهر واقع در شهرستان پاسارگاد در استان فارس با وسعت حدود ۱۶ هزار هکتار است. دشت سعادت‌شهر در فاصله ۱۲۰ کیلومتری شمال شهر شیراز مرکز استان فارس در مختصات جغرافیایی بین  $52^{\circ} 52' 40''$  تا  $53^{\circ} 11' 54''$  طول شرقی و  $30^{\circ} 09' 45''$  تا  $30^{\circ} 11' 28''$  عرض شمالی قرار دارد. دشت سیدان، منطقه گیرنده اطلاعات در مجاورت دشت سعادت‌شهر در شهرستان مرودشت واقع شده است. وسعت دشت سیدان حدود ۲۰ هزار هکتار و در مختصات جغرافیایی بین  $52^{\circ} 52' 45''$  تا  $53^{\circ} 12' 8''$  طول شرقی و  $30^{\circ} 00' 48''$  تا  $30^{\circ} 02' 18''$  عرض شمالی واقع شده است.

به مناطق مشابه فاقد اطلاعات فراهم آورد، اما به استناد منابع علمی این تلاش به‌عنوان یکی از عرصه‌های فعالیت مطالعات نقشه رقومی خاک، در موارد معدودی مورد توجه قرار گرفته است (۲۰ و ۳۷). در چنین مطالعاتی برای هر عامل پس از تعیین متغیرهای محیطی دخیل، شاخص تشابه<sup>۱</sup> (آستانه شباهت) تعریف و بر اساس آن مناطق مشابه مشخص می‌گردد. برای تعیین شاخص تشابه در بعضی مطالعات، فواصل بین ویژگی‌ها یا کلاس‌های خاک (۵ و ۲۴) و در برخی دیگر فواصل بین متغیرهای کمی محیطی (۱۹ و ۳۷) ملاک قرار گرفته است.

در ایران طی شش دهه گذشته، مطالعات خاک‌شناسی با سطوح دقت مختلف تهیه و داده‌های آن در زمان تولید به‌عنوان اطلاعات پایه در پروژه‌های عمرانی و توسعه‌ای مورد استفاده قرار گرفته است. نقشه‌های خاک در این مطالعات به دلیل نیاز به هزینه بسیار زیاد، به روز نشده‌اند. از طرف دیگر، وسعت کمی از این مطالعات با سطح دقت تفصیلی انجام شده است؛ بنابراین، به نظر می‌رسد تنها راه باقی‌مانده استفاده از روش‌هایی است که به یمن پیشرفت‌های علمی و تکنیکی می‌توانند به صورت موفقیت‌آمیزی در تولید نقشه‌های دقیق در مقایسه با نقشه‌های موجود یا تولید نقشه‌های جدید با هزینه‌های بسیار کم‌تر به کار گرفته شوند. یک روش پیشنهادی تعمیم‌پذیری مدل‌های نقشه‌برداری رقومی است. از مطالعاتی که در قالب نقشه‌برداری رقومی خاک در ایران انجام شده (۱، ۲۶ و ۳۳) در تعداد کمی از آن به موضوع تعمیم‌پذیری مدل‌های پیش‌بینی‌کننده کلاس‌های خاک پرداخته شده است (۱). از آن‌جا که دغدغه خاک‌شناسان دستیابی به اطلاعات و داده‌های دقیق خاک در دامنه جغرافیایی وسیع مثل عرصه‌های کشاورزی به‌منظور برآوردن نیازها و انتظارات جدید



شکل ۱- موقعیت مناطق مطالعاتی، الف) استان فارس، ب) شهرستان‌های پاسارگاد و مرودشت، ج) دشت سعادت‌شهر، منطقه دهنده، د) دشت سیدان، منطقه گیرنده.

Figure 1. Study areas location: A) Fars Province, B) Pasargad and Marvdasht counties, C) Saadat Shahr plain, donor area, D) Seidan plain, recipient area.

رژیم حرارتی خاک ترمیک<sup>۱</sup> و رژیم رطوبتی خاک زیریک<sup>۲</sup> است (۳۱).

بر اساس تقسیم‌بندی ماہلر (۱۸)، دشت سعادت‌شهر دارای سه واحد اصلی فیزیوگرافی شامل مخروط افکنه‌های آبرفتی و واریزه‌ای<sup>۳</sup>، دشت‌های دامنه‌ای<sup>۴</sup> و اراضی پست<sup>۵</sup> است. در دشت سیدان واحد اراضی پست وجود ندارد و بخش‌هایی از آن تحت پوشش واحد فلات‌های<sup>۶</sup> مرتفع است (۳۱). سازندهای زمین‌شناسی در منطقه گسترش دو دشت شامل سازند سورمه (آهک دولومیتیک توده‌ای و تیره رنگ)،

در دو دشت مطالعاتی پارامترهای محیطی تقریباً مشابه هستند. در دشت سعادت‌شهر ارتفاع منطقه از حدود ۱۷۰۷ تا ۲۰۴۵ متر و در دشت سیدان بین حدود ۱۶۰۸ تا ۱۹۰۰ متر متغیر است. این دو دشت با گسترش شمال‌غرب- جنوب شرق تقریباً از همه طرف به ارتفاعات محدود می‌شوند. در دشت سعادت‌شهر متوسط دمای سالانه ۱۸/۳، متوسط حداقل دمای سالانه ۱۱/۷ و متوسط حداکثر دمای سالانه ۲۴/۸ درجه سلسیوس و متوسط بارندگی ۲۳۲ میلی‌متر در سال است. در دشت سیدان دمای متوسط سالانه، متوسط حداقل و متوسط حداکثر به ترتیب ۱۷/۳، ۹/۰ و ۲۵/۶ درجه سلسیوس و متوسط بارندگی سالانه ۳۰۲/۶ میلی‌متر است. در هر دو دشت

- 1- Thermic
- 2- Xeric
- 3- Alluvio-colluvial fans
- 4- Piedmont plain
- 5- Low lands
- 6- Plateau

متغیرهای کمکی: در این مطالعه از ۲۵ متغیر کمکی پستی و بلندی و داده‌های سنجش از دور استفاده شد. متغیرهای محیطی پستی و بلندی (۹) بر مبنای مدل رقومی ارتفاع<sup>۲</sup> با قدرت تفکیک ۳۰ متر متعلق به سنجنده آستر با استفاده از نرم‌افزار ساگا جی‌آی‌اس (۲۸) تهیه شدند. از تصویر ماهواره‌ای سنجنده لندست هشت برای تهیه نقشه کاربری اراضی و شاخص‌های پوشش گیاهی استفاده شد. تصاویر موردنظر مربوط به سال میلادی ۲۰۱۷ از سایت USGS (<https://earthexplorer.usgs.gov>) استخراج و لایه‌های فوق در نرم‌افزار ENVI 5.3 پس از انجام تصحیحات مکانی و اتمسفری تولید شدند. شاخص‌های پوشش گیاهی شامل شاخص تفاضلی نرمال شده پوشش گیاهی<sup>۳</sup> و شاخص تفاضلی نرمال شده رطوبتی<sup>۴</sup> بر اساس رابطه‌های ۱ و ۲ تهیه شدند:

$$NDVI = \frac{(\rho_{NIR} - \rho_R)}{(\rho_{NIR} + \rho_R)} \quad (1)$$

$$NDMI = \frac{(\rho_{NIR} - \rho_{MIR})}{(\rho_{NIR} + \rho_{MIR})} \quad (2)$$

که در آن‌ها،  $\rho_{NIR}$ ،  $\rho_R$  و  $\rho_{MIR}$  به ترتیب بازتاب طیفی باندهای چهار، پنج و شش مربوط به سنجنده لندست هشت است.

همچنین از تصاویر ماهواره لندست هشت مربوط به سال ۲۰۱۷ میلادی برای تهیه نقشه دمای سطح زمین<sup>۵</sup> استفاده شد. نقشه دمای سطح زمین از رابطه ۳ بر اساس قانون پلانک (۱۴) محاسبه شد:

$$LST = \frac{K_2}{\ln\left(\frac{K_1}{B(LST)} + 1\right)} \quad (3)$$

- 2- Digital elevation model, DEM  
3- Normalized difference vegetation index, NDVI  
4- Normalized difference moisture index, NDMI  
5- Land surface temperature, LST

سازندهای داریان، فهلیان (آهک اوربیتولین‌دار خاکستری متمایل به سیاه و آهک دولومیتیک)، سازند کژدمی (آهک خاکستری و آهک مارنی)، سازند سروک (آهک خاکستری متراکم) و سازند بختیاری (کنگومرا، ماسه‌سنگ دانه‌درشت) است (۳۱). رودخانه سیوند از هر دو دشت سعادت‌شهر و دشت سیدان عبور می‌نماید. در هر دو دشت کشاورزی دیم و آبی انجام می‌گیرد. جنگل‌ها و مراتع در اراضی شیب‌دار اطراف دو دشت گسترش یافته‌اند و نواحی مرکزی تحت کشت محصولات کشاورزی شامل گندم، جو، چغندرقد، صیفی‌جات و برنج است.

**مطالعات صحرائی و آزمایشگاهی:** در این مطالعه، در مجموع ۱۰۹ نقطه مطالعاتی شامل ۸۲ نقطه در منطقه دهنده اطلاعات و ۲۷ نقطه در منطقه گیرنده مشخص (شکل ۱) و طی مطالعات میدانی، حفر، تشریح و نمونه‌برداری شدند (۲۹). حفر و مطالعه خاکرخ‌ها در منطقه گیرنده تنها برای ارزیابی صحت مدل به کار رفت و ساخت مدل در این منطقه با این فرض انجام شد که هیچ‌گونه اطلاعات و داده خاک در دسترس نیست. تعداد نقاط مطالعاتی در منطقه گیرنده با توجه به اختلاف کم مساحت دو منطقه، نزدیک به تعداد نقاط آزمون در منطقه دهنده تعیین شد تا مقایسه صحت مدل در دو منطقه با مجموعه داده‌های تقریباً یکسان انجام گیرد. نمونه‌برداری در منطقه دهنده با استفاده از روش مربعات لاتین<sup>۱</sup> (۲۳) و در منطقه گیرنده بر اساس الگوی فیزیوگرافی منطقه بر روی چند مقطع عرضی با فواصل منظم انجام شد. با توجه به بعضی محدودیت‌ها و نظر کارشناسی، مقداری تغییر جزئی در موقعیت بعضی نقاط مشاهداتی انجام شد. بر اساس نتایج تشریح خاکرخ‌ها و تجزیه‌های آزمایشگاهی نمونه‌های خاک، خاک‌ها بر اساس سیستم جامع رده‌بندی خاک (۳۲)، طبقه‌بندی شدند.

- 1- Latin hypercube method

شدند. مقدار هر متغیر کمکی در نقاط مطالعاتی از لایه‌های شبکه‌ای مربوطه استخراج و در یک پایگاه داده ذخیره شد تا در مدل مورد استفاده به کار روند. جدول ۱ لایه‌های کمکی اولیه مورد استفاده در این مطالعه را نشان می‌دهد.

که در آن،  $K_1$  و  $K_2$  ضرایب واسنجی باندهای حرارتی است که از لندست هشت دریافت شدند. B(LST) رادیانس قابل محاسبه است (۶). تمامی متغیرهای کمکی مورد استفاده، در قدرت تفکیک ۳۰ متر در نرم‌افزار ArcGIS, 10.4 هم‌مقیاس

جدول ۱- متغیرهای کمکی مورد استفاده در این مطالعه.

**Table 1. Environmental covariates that used in this study.**

مأخذ Reference	متغیر محیطی کمکی Environmental covariate
DEM	ارتفاع Elevation
DEM	تحدب Convexity
DEM	جهت شیب Aspect
DEM	شیب Slope
DEM	ارتفاع شیب Slope height
DEM	طول شیب Slope length
DEM	عمق دره Valley depth
DEM	شاخص قدرت جریان Stream power index
DEM	شاخص خیس‌توپوگرافی Topographic wetness index
DEM	مساحت حوزه اصلاح شده Modified catchment area
DEM	شاخص همواری دره با درجه تفکیک بالا Multiresolution Index of Valley Bottom Flatness
DEM	متوسط موقعیت شیب Mid slope position
DEM	ارتفاع نرمال شده Normalize height
DEM	موقعیت نسبی شیب Relative slope position

ادامه جدول ۱-

Continue Table 1.

مأخذ Reference	متغیر محیطی کمکی Environmental covariate
DEM	انحنای کل Total curvature
DEM	انحنای سطحی Plan curvature
DEM	انحنای نیم‌رخ Profile curvature
DEM	ناهمواری‌های توپوگرافی Terrain ruggedness index
DEM	فاکتور نمای توپوگرافی Terrain view factor
DEM	انحنای مماسی Tangential curvature
DEM	آسمان قابل روئیت Visible sky
Landsat 8	نقشه کاربری اراضی Land use map
Landsat 8	شاخص گیاهی تفاضلی نرمال شده Normalized difference vegetation index
Landsat 8	شاخص رطوبتی تفاضلی نرمال شده Normalized difference moisture index
Landsat 8	دمای سطح زمین Land Surface Temperature

تقسیم‌بندی گره‌ها استفاده می‌شود، به صورت تصادفی انجام می‌گیرد. در روش جنگل تصادفی دو ویژگی *mtry* و *ntree* به ترتیب برای تعداد متغیرهای کمکی مورد استفاده در هر زیر مجموعه و تعداد درختان مورد استفاده در جنگل تعیین می‌گردد. تعداد متغیرها می‌تواند از یک تا تعداد کل متغیرهای کمکی را در برگیرد. تعداد درختان توسط کاربر و معمولاً از ۵۰۰ تا ۱۰۰۰ انتخاب می‌گردد (۳).

در الگوریتم جنگل تصادفی، امکان انتخاب مهم‌ترین و مؤثرترین متغیرهای کمکی امکان‌پذیر است (۳). تخمین اهمیت یک متغیر با ردیف کردن

**مفاهیم مدل جنگل تصادفی:** روش جنگل تصادفی در بین روش‌های درختی، شیوه نسبتاً پیچیده‌ای است که به منظور افزایش دقت مدل در آن چندین درخت تصمیم آموزش داده می‌شود. نتیجه حاصل پیش‌بینی گروهی از درختان تصمیم‌گیری است (۲). این روش به کرات برای تهیه نقشه کلاس‌های خاک استفاده شده است (۱۰، ۱۲ و ۱۳). در روش یادگیری جنگل تصادفی هر درخت تصمیم‌گیری با استفاده از یک نمونه تصادفی که از مجموعه داده‌های آموزشی انتخاب شده است، آموزش می‌بیند. انتخاب مجموعه‌ای از متغیرهای پیش‌بینی‌کننده نیز که برای



بود، برای پیش‌بینی کلاس‌های خاک در این منطقه استفاده شد. دقت مدل در منطقه دهنده و گیرنده مورد ارزیابی و مقایسه قرار گرفت.

ارزیابی صحت مدل: در فرایند مدل‌سازی در منطقه دهنده، با انتخاب تصادفی، ۷۰ درصد داده‌های جمع‌آوری شده به‌عنوان مجموعه نقاط آموزش و ۳۰ درصد آن‌ها به‌منظور اعتبارسنجی استفاده شدند (۳). پیش‌بینی کلاس‌های خاک در منطقه گیرنده به دو صورت: (۱) استفاده از ۷۰ درصد نقاط مطالعاتی منطقه دهنده و (۲) استفاده از تمامی نقاط مطالعاتی در منطقه دهنده، به‌عنوان مجموعه داده‌های آموزش انجام شد. در این پیش‌بینی از نتایج مطالعه ۲۷ نقطه مشاهداتی در منطقه گیرنده به‌عنوان داده‌های آزمون در هر دو صورت استفاده شد. سپس، برای ارزیابی صحت مدل، از شاخص‌های صحت کلی<sup>۴</sup>، صحت کاربر<sup>۵</sup>، صحت تولیدکننده<sup>۶</sup> و ضریب کاپا<sup>۷</sup> (۷) استفاده شد. ضریب کاپا از رابطه‌های ۴، ۵ و ۶ قابل محاسبه است (۳۰):

$$K = \frac{p_o - p_c}{1 - p_c} \quad (4)$$

$$P_o = \frac{np}{no} \times 100\% \quad (5)$$

$$P_c = \sum_{i=1}^c \frac{n_{ip} \times n_{io}}{no} \quad (6)$$

که در آن‌ها،  $P_o$  سازگاری مشاهده‌ای،  $P_c$  سازگاری شانسی،  $C$  تعداد سلسله گروه‌های نمونه‌های آزمایشی،  $no$  تعداد کل نمونه‌های آزمایشی،  $np$  تعداد کل نمونه‌های آزمایشی که به درستی پیش‌بینی شده‌اند،  $n_{ip}$  تعداد کل نمونه‌های آزمایشی که پیش‌بینی می‌شود در طبقه  $i$  ام قرار گیرند،  $n_{io}$  تعداد کل نمونه‌های آزمایشی که حقیقتاً در طبقه  $i$  ام قرار دارند.

مقادیر هر متغیر در نمونه‌های خارج از سبد<sup>۱</sup> که معمولاً ۱/۳ مشاهده‌ها بوده و در مدل به‌کار نرفته‌اند و با طبقه‌بندی مجدد نمونه‌های خارج از سبد با استفاده از متغیر ردیف شده، انجام می‌گیرد. تغییر در خطای خارج از سبد به‌عنوان معیاری برای اهمیت متغیر به‌کار می‌رود. متغیرهایی که حذف آن‌ها منجر به افزایش نسبتاً بالاتری در خطای خارج از سبد می‌شود، متغیرهای مهم‌تر تلقی می‌شوند. یکی از شاخص‌هایی که در مدل جنگل تصادفی، اهمیت متغیرهای کمکی را برای تخمین ویژگی هدف نمایش می‌دهد، شاخص میانگین کاهش جینی<sup>۲</sup> است که در این مطالعه استفاده شد. برای یک متغیر، این شاخص معیاری از میانگین کاهش کلی ناخالصی و یا همگنی در گره‌ها و برگ‌های انتهایی همه درختان تصمیم به‌واسطه استفاده از آن متغیر بوده و بین صفر تا یک متغیر است. متغیرهایی که استفاده از آن‌ها باعث خلوص بالاتری در گره‌های انتهایی شد، مقادیر بزرگ‌تری از شاخص اهمیت جینی را به خود اختصاص داده و به‌عنوان متغیرهای مهم در مدل نهایی استفاده شدند. متغیرهای مهم عیناً برای منطقه گیرنده با درجه وضوح مشابه (۳۰ متر  $\times$  ۳۰ متر) تهیه و در مدل انتقال یافته به‌کار گرفته شدند.

در پژوهش حاضر، برای تعیین مهم‌ترین متغیرهای کمکی و مدل‌سازی رابطه خاک-سرزمین از مدل جنگل تصادفی بسته جنگل تصادفی<sup>۳</sup> موجود در نرم‌افزار R (۲۷) استفاده و کدهای تعریف شده در آن اجرا شد. برای تعداد درختان در مدل، مقدار ۱۰۰۰ انتخاب شد. در فرایند پیش‌بینی کلاس‌های خاک در منطقه گیرنده، از مدل جنگل تصادفی که بر اساس روابط بین کلاس‌های خاک و متغیرهای کمکی با قدرت تفکیک ۳۰ متر در منطقه دهنده حاصل شده

4- Overall accuracy  
5- User's accuracy  
6- Producer's accuracy  
7- Kappa coefficient

1- Out-Of- Bag, OOB  
2- Mean Decrease Gini  
3- Random Forest

### نتایج و بحث

در مطالعات نقشه‌برداری رقومی خاک، تعمیم مدل ساخته شده از داده‌های موجود در یک منطقه (منطقه دهنده) به منطقه دیگر (منطقه گیرنده) مستلزم شباهت متغیرهای محیطی نماینده عوامل خاک‌سازی در دو منطقه است. در دو منطقه مطالعاتی سعادت‌شهر و سیدان شباهت متغیرهای محیطی، امکان تئوریک تعمیم مدل ساخته شده در منطقه دهنده به منطقه گیرنده را فراهم آورده است.

بر اساس مطالعه نقاط مشاهداتی از پیش تعیین شده در دشت سعادت‌شهر به‌عنوان منطقه دهنده، خاک‌رخ‌های مطالعه شده در سه رده آلفی‌سولز، اینسپتی‌سولز و انتی‌سولز طبقه‌بندی شدند. در سطح گروه بزرگ چهار خاک هاپلوزرال‌فز، هاپلوزرپتز، کلسی‌زرپتز و زراورتنز شناسایی شد. در سطح زیرگروه به‌ترتیب تکامل، خاک‌ها در کلاس‌های کلسیک هاپلوزرال‌فز، تیبیک کلسی‌زرپتز، کلسیک هاپلوزرپتز، تیبیک هاپلوزرپتز، تیبیک زراورتنز و

فلوونتیک هاپلوزرپتز قرار گرفتند. شواهدی از انتقال رس در خاک‌های آلفی‌سولز و انتقال و تغییر شکل مواد در افق‌های زیرسطحی در رده اینسپتی‌سولز از جمله مهم‌ترین فرایندهای شناخته‌شده در منطقه است. خاک‌های زیرگروه تیبیک زراورتنز و تیبیک هاپلوزرپتز اغلب مربوط به حاشیه‌های دشت بر روی واحد مخروط افکنه‌های آبرفتی و واریزه‌ای هستند که تکامل خاکرخی کمی دارند. در ارتفاعات پایین‌تر و به‌سمت مناطق میانی دشت، خاک‌ها دارای زیرگروه‌های کلسیک هاپلوزرپتز، تیبیک کلسی‌زرپتز، فلوونتیک هاپلوزرپتز و کلسیک هاپلوزرال‌فز بودند (جدول ۲).

در دشت سیدان به‌عنوان منطقه گیرنده در نقاط مطالعاتی از بین کلاس‌های بالا، خاک کلسیک هاپلوزرال‌فز مشاهده نشد، ولی در مناطق دارای شرایط پایدارتر و بر روی فیزیوگرافی مناطق فلات مرتفع، زیرگروه پتروکلسیک کلسی‌زرپتز مشاهده شد که جزء خاک‌های منطقه دهنده نیست (جدول ۲).

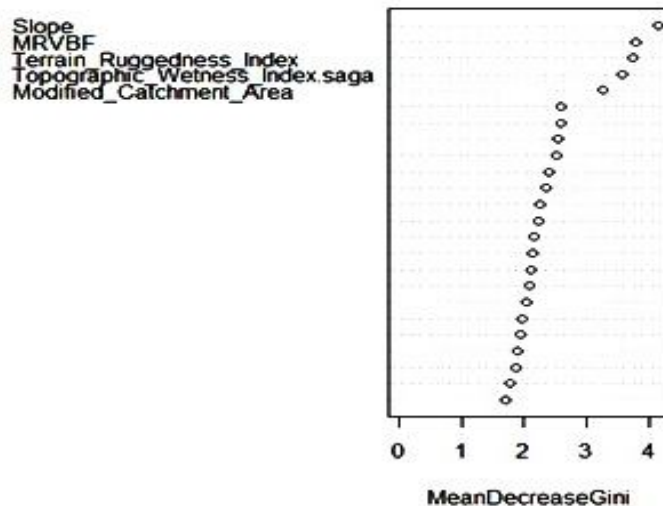
جدول ۲- فراوانی کلاس‌های خاک بر اساس سامانه رده‌بندی جامع آمریکایی (۲۰۱۴) در سطح زیرگروه در مناطق دهنده و گیرنده.

**Table 2. Frequency of soil classes at subgroup level based on the Soil Taxonomy (2014) system in donor and recipient areas.**

درصد فراوانی کلاس‌های خاک		کلاس خاک Soil class
منطقه گیرنده Recipient area	منطقه دهنده Donor area	
26	30	تیبیک زراورتنز Typic Xerorthents
22	10	تیبیک هاپلوزرپتز Typic Haploxerepts
4	6	فلوونتیک هاپلوزرپتز Fluventic Haploxerepts
19	7	کلسیک هاپلوزرپتز Calcic Haploxerepts
22	39	تیبیک کلسی‌زرپتز Typic Calcixerepts
-	8	کلسیک هاپلوزرال‌فز Calcic Haploxerafs
7	-	پتروکلسیک کلسی‌زرپتز Petrocalcic Calcixerepts

توپوگرافی و مساحت حوزه اصلاح شده از بیشترین اهمیت برای پیش‌بینی متغیر هدف برخوردار شدند (شکل ۲).

بر اساس مقادیر شاخص میانگین کاهش جینی در مدل جنگل تصادفی از بین تمام متغیرهای محیطی، شیب، شاخص همواری دره با درجه تفکیک بالا، شاخص ناهمواری‌های توپوگرافی، شاخص خیزی



شکل ۲- اهمیت نسبی پارامترهای کمکی برای پیش‌بینی کلاس‌های خاک در منطقه دهنده بر اساس میانگین کاهش جینی.

Figure 2. Relative importance of covariates to predict soil classes in donor area based on the mean decrease in Gini index.

کاربر برای یک کلاس نشان می‌دهد در کل مشاهدات آن کلاس، مدل موفق به پیش‌بینی خوبی بوده و کم‌تر به خطا به‌جای سایر کلاس‌ها پیش‌بینی شده است. این در حالی است که صحت بالای تولیدکننده نشان از آن دارد که پیش‌بینی خوبی برای کلاس موردنظر انجام شده است.

در پیش‌بینی کلاس‌های خاک در منطقه دهنده اطلاعات، صحت کلی و ضریب کاپا به‌ترتیب ۷۲ درصد و ۰/۵۹ به‌دست آمد. این نتایج بیانگر ارتباط خوب بین داده‌های مشاهده‌ای و پیش‌بینی در منطقه دهنده است. جدول ۳، نتایج صحت کاربر و صحت تولیدکننده را برای کلاس‌های مختلف خاک در منطقه دهنده نشان می‌دهد. مقادیر بالای صحت

جدول ۳- نتایج ارزیابی مدل جنگل تصادفی برای پیش‌بینی کلاس‌های خاک بر اساس سامانه رده‌بندی جامع آمریکایی (۲۰۱۴) در سطح زیرگروه در منطقه دهنده.

**Table 3. Evaluation of RF model to predict soil classes at subgroup level based on the Soil Taxonomy (2014) system in donor area.**

داده‌های آموزشی Training data		کلاس خاک Soil classes
PA	UA	
1	0.78	تیپیک زراورتنز Typic Xerorthents
0.5	0.34	تیپیک هاپلوزرپتز Typic Haploxerepts
0	NaN	فلوونتییک هاپلوزرپتز Fluventic Haploxerepts
0	NaN	کلسیک هاپلوزرپتز Calcic Haploxerepts
0.82	0.75	تیپیک کلسی‌زرپتز Typic Calcixerepts
0.5	1	کلسیک هاپلوزرالفز Calcic Haploxeralfs

UA: User's accuracy, PA: Procedure's accuracy, NaN: No predicted class.

خاک‌های کم‌تر تکامل‌یافته هاپلوزرپتز مقدار و روند نامنظم مواد آلی با عمق خاک است، به‌نظر می‌رسد که متغیرهای کمکی مورد استفاده در منطقه نتوانسته‌اند شرایط خاص حاکم بر تشکیل این خاک را رصد نمایند. بهترین پیش‌بینی مدل مربوط به کلاس‌های تیپیک کلسی‌زرپتز و تیپیک زراورتنز است که بر اساس نقاط مشاهداتی، بیش‌ترین گسترش جغرافیایی در منطقه را دارا هستند. در موارد معدودی، مدل خاک تیپیک هاپلوزرپتز را برای محل گسترش کلاس‌های تیپیک زراورتنز پیش‌بینی نموده که به‌نظر می‌رسد به‌دلیل پراکنش جغرافیایی و روند تکاملی نزدیک به هم این دو خاک است.

در منطقه دهنده، از خاک‌هایی که مدل موفق به پیش‌بینی آن‌ها شده است، به‌جز خاک کلسیک هاپلوزرالفز در سایر موارد با توجه به مقدار بزرگ‌تر صحت تولیدکننده از صحت کاربر، کلاس‌های خاک با بیش برآورد<sup>۱</sup> مواجه بوده‌اند. مدل مورد استفاده نتوانسته است خاک‌های دو زیرگروه کلسیک هاپلوزرپتز و فلوونتییک هاپلوزرپتز را تخمین بزند. به هر حال شرایط تشکیل خاک‌های کلسیک هاپلوزرپتز بسیار نزدیک به خاک‌های غالب تیپیک کلسی‌زرپتز در منطقه بوده و تفاوت آن‌ها در مقدار آهک ثانویه است که این ویژگی به مدل داده نشده است. برای خاک فلوونتییک هاپلوزرپتز که ویژگی متمایزکننده آن با

#### 1- Overestimation

خاک به جای سایر خاک‌ها پیش‌بینی شده است. نقطه قوت به‌کارگیری مدل با کل داده‌ها، علاوه بر تخمین بهتر برای خاک تیبیک زراورتنز، پیش‌بینی خاک زیرگروه فلوننتیک هاپلوزرپنز در منطقه گیرنده است. در عین حال که در منطقه دهنده این نوع خاک به‌درستی پیش‌بینی نشده بود، نزدیکی عوامل محیطی مؤثر بر تشکیل این خاک به یکی یا بیش‌تر خاک‌های این زیرگروه در منطقه دهنده درخور توجه است. با مقایسه ضریب کاپا در نتایج به‌کارگیری دو مجموعه داده‌های مورد نظر و محاسبه سازگاری شانس برای هر دو مورد، مشخص شد پیش‌بینی این کلاس خاک حاصل سازگاری مشاهده‌ای بوده و شانس نقشی در پیش‌بینی این کلاس خاک نداشته است. این موضوع نشان داد می‌توان رابطه مستقیمی بین تعداد نقاط مشاهداتی در مجموعه داده‌های آموزشی در منطقه دهنده با کلاس‌های قابل پیش‌بینی در منطقه گیرنده برقرار نمود. در تمام مواردی که مدل برای پیش‌بینی کلاس‌های کلسیک هاپلوزرپنز، تیبیک هاپلوزرپنز و پتروکلسیک کلسی‌زرپنز با خطا مواجه بوده است، خاک تیبیک کلسی‌زرپنز پیش‌بینی شده است که احتمالاً به‌دلیل فراوانی بالای این کلاس خاک در منطقه بوده است. اختلاف خاک‌های مورد اشاره بیش‌تر مربوط به سطح تغییر شکل یا انتقال کربنات‌ها در عمق خاک‌رخ است.

عدم پیش‌بینی زیرگروه پتروکلسیک کلسی‌زرپنز که خاستگاه آن فیزیوگرافی فلات‌های مرتفع است، به‌دلیل عدم وجود این نوع واحد فیزیوگرافی در منطقه دهنده دور از انتظار نبود. در منطقه گیرنده به‌جز خاک تیبیک کلسی‌زرپنز که در آن صحت تولیدکننده بیش از صحت کاربر است، در سایر موارد خاک‌ها با کم تخمینی<sup>۱</sup> مواجه بودند.

انتظار می‌رود دو منطقه دهنده و گیرنده اطلاعات با توجه به نزدیکی جغرافیایی و شرایط یکسان تأثیر عوامل خاک‌سازی بر تشکیل خاک‌ها از جمله اقلیم مشابه، همواری دشت‌ها، ساختار مشابه زمین‌شناسی ارتفاعات مجاور و الگوی کشت و رفتار کشاورزان، شرایط مشابهی نیز برای تشکیل خاک‌ها دارا باشند (۱۹). مشاهدات انجام‌شده در منطقه گیرنده به‌منظور راستی‌آزمایی مدل انتقال‌یافته از منطقه دهنده نشان داد زیرگروه‌های مشابهی در منطقه گیرنده گسترش دارند ولی دامنه جغرافیایی گسترش هر خاک متفاوت از منطقه دهنده است. نتایج تعمیم مدل نشان داد صحت کلی و ضریب کاپا در منطقه گیرنده با لحاظ نمودن ۷۰ درصد مجموعه داده‌های آموزشی به‌ترتیب ۴۵ درصد و ۰/۲۸ و برای کل مجموعه داده‌های منطقه دهنده، به‌ترتیب ۵۲ درصد و ۰/۳۸ است. به‌طورکلی لحاظ نمودن مدل ساخته شده با کل داده منطقه دهنده نتایج بهتری نسبت به مدل ساخته شده با ۷۰ درصد این داده‌ها را نشان داد. جدول ۴، نتایج ارزیابی مدل در منطقه گیرنده با دو مجموعه داده متفاوت را نشان می‌دهد.

در منطقه گیرنده، بهترین پیش‌بینی با هر دو مجموعه داده‌های آموزشی مربوط به کلاس‌های تیبیک زراورتنز و تیبیک کلسی‌زرپنز است. با توجه به مقادیر صحت کاربر و صحت تولیدکننده برای زیرگروه تیبیک زراورتنز، علاوه بر این که خاک‌های این کلاس به‌خوبی پیش‌بینی شده‌اند، این کلاس به‌جای خاک دیگری نیز پیش‌بینی نشده است. تخصیص مقدار حداکثر صحت تولیدکننده و مقدار کم صحت کاربر برای خاک تیبیک کلسی‌زرپنز، نشان‌دهنده پیش‌بینی خوب کلاس‌های مشاهده شده این خاک و همچنین موارد نسبتاً زیادی است که این

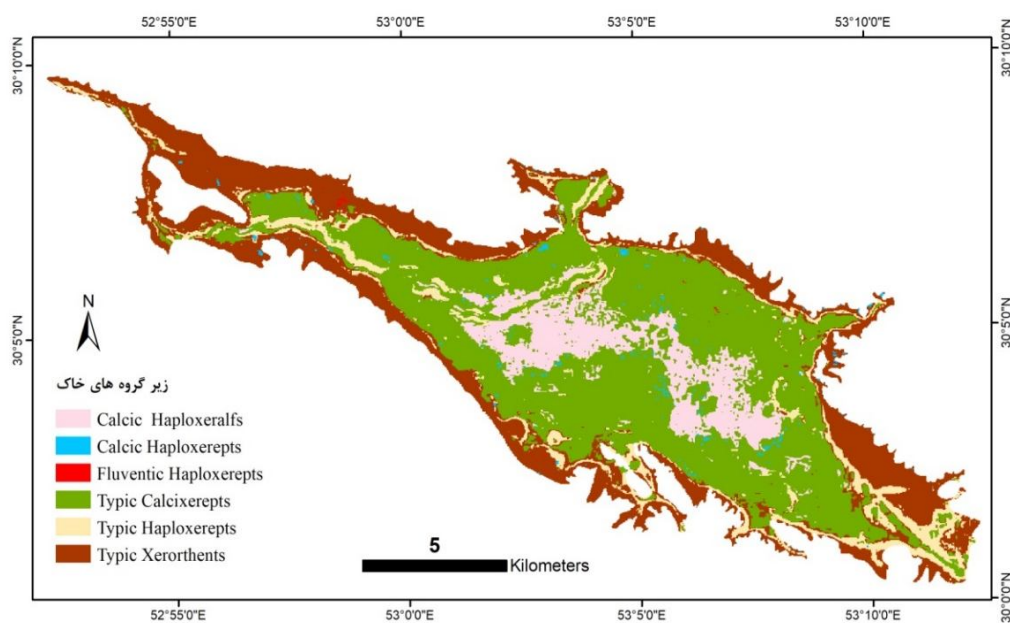
جدول ۴- نتایج ارزیابی مدل جنگل تصادفی برای پیش‌بینی کلاس‌های خاک بر اساس سامانه رده‌بندی جامع آمریکایی (۲۰۱۴) در سطح زیرگروه در منطقه گیرنده با دو مجموعه داده آموزشی.

**Table 4. Evaluation of RF model to predict soil classes at subgroup level based on the Soil Taxonomy (2014) system in recipient area using two training datasets.**

کل داده‌ها Total data		۷۰ درصد داده‌ها 70% of total data		کلاس خاک Soil classes
PA	UA	PA	UA	
0.86	1	0.72	1	تیبیک زراورتنتر Typic Xerorthents
0.17	0.5	0.17	0.25	تیبیک هاپلوزرپتر Typic Haploxerepts
1	1	0	NaN	فلوونتیک هاپلوزرپتر Fluventic Haploxerepts
0	NaN	0	NaN	کلسیک هاپلوزرپتر Calcic Haploxerepts
1	0.34	1	0.34	تیبیک کلسی‌زرپتر Typic Calcixerepts
0	NaN	0	NaN	پتروکلسیک کلسی‌زرپتر Petrocalcic Calcixerepts

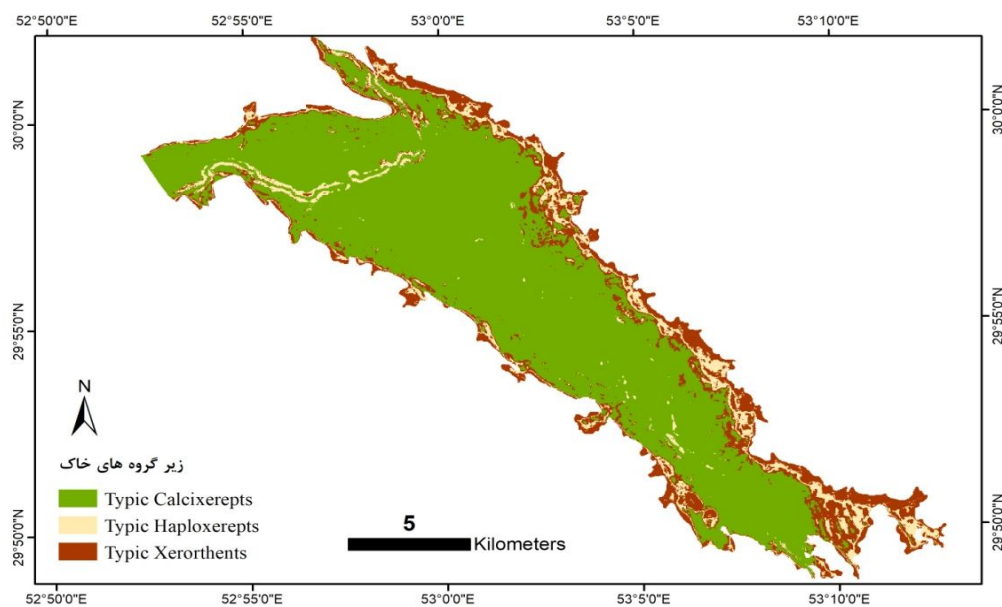
مورفولوژیکی، فیزیکی، شیمیایی و مینرالوژیکی خود را از موقعیت قرارگیری بر پستی و بلندی به دست می‌آورند. بنابراین به نظر می‌رسد در شرایطی که خاک‌ها متأثر از عوامل خاک‌سازی محدودتری بوده و یا یک عامل خاک‌ساز در تشکیل خاک به‌طور مشهودی بر سایر عوامل غلبه بیشتر داشته باشند، تعمیم ویژگی‌ها و یا کلاس‌های خاک به مناطق مشابه بهتر انجام می‌گیرد. به هر حال در شرایطی که عوامل زیادی بر تشکیل خاک‌ها دخالت داشته باشند، مدل کردن تأثیر هر عامل و همچنین اثر متقابل عوامل مختلف خاک‌ساز در تشکیل خاک دشوارتر است. نقشه‌های کلاس‌های پیش‌بینی‌شده در سطح زیرگروه برای دو منطقه گیرنده و تعمیم به ترتیب در شکل‌های ۳ و ۴ نشان داده شده‌اند.

به‌طورکلی در مطالعات نقشه‌برداری رقومی خاک که برای پیش‌بینی کلاس‌های خاک در یک منطقه از داده‌ها و مدل‌های ساخته شده در منطقه مشابه یا غیرمشابه دیگر (خارج از منطقه مورد نظر) استفاده شده است، نتایج به‌دست آمده با محدودیت روبه‌رو بوده است (۱ و ۱۱). در این مطالعه این نتیجه با مقایسه پیش‌بینی کلاس‌های خاک در دو منطقه ساخت مدل و منطقه خارج از آن حاصل گردید. در مناطق خشک و نیمه‌خشک مواد مادری خاک‌ها بیش‌تر تحت تأثیر پستی و بلندی متحول شده و اقلیم و سایر فرایندهای خاک‌ساز در مرتبه‌های بعد قرار گرفته و از اهمیت کم تا ناچیز برخوردار هستند (۲۲). الگوی پوشش گیاهی در دشت‌های کشاورزی متأثر از نوع خاک‌ها است که خصوصیات



شکل ۳- نقشه کلاس‌های خاک پیش‌بینی‌شده در سطح زیرگروه در دشت سعادت شهر.

Figure 3. Map of predicted soil classes at subgroup level in Saadat Shahr plain.



شکل ۴- نقشه کلاس‌های خاک پیش‌بینی‌شده در سطح زیرگروه در دشت سیدان.

Figure 4. Map of predicted soil classes at subgroup level in Seidan plain.

## نتیجه‌گیری کلی

در سطح انتقال و تجمع کربنات‌کلسیم در خاکرخ باعث شده اغلب این خاک‌ها با عنوان تیپیک کلسی‌زرپتز (خاک غالب منطقه) پیش‌بینی شوند. بنابراین با توجه به محدودیت روش نقشه‌برداری رقومی در پیش‌بینی خاک‌هایی که به لحاظ ژنتیکی شرایط تشکیل مشابهی دارند، باید خصیصه‌هایی در قالب متغیر کمکی به مدل داده شود که مدل با درک فرآیندهای تشکیل خاک‌ها، بتواند بر این محدودیت فائق آید. پیش‌بینی خاک فلوونوتیک هاپلوزرپتز در منطقه تعمیم پس از به‌کار بردن داده‌های آموزشی پیش‌تر (کل مشاهده‌ها منطقه دهنده) نشان داد که افزایش داده‌های آموزشی می‌تواند شانس پیش‌بینی کلاس‌های خاک بیش‌تری را در منطقه گیرنده به همراه داشته باشد. به‌طورکلی در مطالعات تعمیم مدل‌های خاک، ضروری است پژوهش‌های بیش‌تری برای دستیابی به مدل‌های کاربردی‌تر و متغیرهای کمکی مؤثرتر در یافتن و انتقال موفقیت‌آمیز مدل‌های مبتنی بر روابط خاک- زمین‌نما انجام گیرد. این مطالعات به‌ویژه برای پیش‌بینی کلاس‌های خاک در سطوح پایین‌تر طبقه‌بندی می‌تواند در اولویت مطالعات نقشه‌برداری رقومی خاک قرار گیرد.

در این پژوهش در قالب روش‌های نقشه‌برداری رقومی خاک، صحت پیش‌بینی کلاس‌های خاک با استفاده از مدل جنگل تصادفی در سطح زیرگروه در یک منطقه دارای داده‌های جدید و همچنین انتقال مدل ساخته شده به‌منظور تهیه نقشه زیرگروه‌های خاک خارج از منطقه ساخت مدل، مورد ارزیابی قرار گرفت. از بین متغیرهای محیطی، شیب، شاخص همواری دره با درجه تفکیک بالا، شاخص ناهمواری‌های توپوگرافی، شاخص خیسی توپوگرافی و مساحت حوزه اصلاح شده نقش مؤثرتری در پیش‌بینی کلاس‌های خاک نشان دادند. مقایسه کارایی مدل به‌کار رفته در هر دو منطقه نشان داد در عین نزدیکی جغرافیایی دو منطقه، دقت مدل برای پیش‌بینی کلاس‌های خاک در محل ساخت مدل (منطقه دهنده) بهتر از نتایج پیش‌بینی در منطقه خارج از ساخت مدل (منطقه گیرنده) بود. در هر دو منطقه بهترین تخمین متعلق به دو زیرگروه تیپیک کلسی‌زرپتز و تیپیک زراورتنز بود. زیرگروه اول، خاک غالب در دو منطقه و زیرگروه دوم به لحاظ موقعیت تشکیل، روند تکاملی و خصوصیات خاک از تفاوت فاحشی نسبت به سایر زیرگروه‌های موجود در منطقه برخوردار بود. شباهت خاک‌های دارای اختلاف

## منابع

1. Abbaszadeh Afshara, F., Ayoubib, S., and Jafari, A. 2018. The extrapolation of soil great groups using multinomial logistic regression at regional scale in arid regions of Iran. *Geoderma*. 315: 1. 367-48.
2. Breiman, L. 2001. Random forests. *Machine Learning*. 45: 1. 5-32.
3. Brungard, C.W., Boettinger, J.L., Duniway, M.C., Wills, S.A., and Edwards, T.C. 2015. Machine learning for predicting soil classes in three semi-arid landscapes. *Geoderma*. 239-240: 1. 68-83.
4. Bui, E.N. 2004. Soil survey as a knowledge system. *Geoderma*. 120: 1-2. 17-26.
5. Carré, F., and Girard, M.C. 2002. Quantitative mapping of soil types based on regression kriging of taxonomic distances with landform and land cover attributes. *Geoderma*. 110: 3-4. 241-263.
6. Coll, C., Galve, J.M., Sanchez, J.M., and Caselles, V. 2010. Validation of Landsat-7/ETM+ thermal-band calibration and atmospheric correction with ground-based measurements. *IEEE Transactions on Geoscience and Remote Sensing*. 48: 1. 547-555.



7. Congalton, R. 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*. 37: 1. 35-46.
8. Debella-Gilo, M., and Etzelmüller, B. 2009. Spatial prediction of soil classes using digital terrain analysis and multinomial logistic regression modeling integrated in GIS. Examples from Vestfold County, Norway. *Catena*. 77: 1. 8-18.
9. Gallant, J.C., and Austin, J.M. 2015. Derivation of terrain covariates for digital soil mapping in Australia. *Soil Research*. 53: 1. 895-90.
10. Grimm, R., Behrens, T., Marker, M., and Elsenbeer, H. 2008. Soil organic carbon concentrations and stocks on Barro Colorado Island-Digital soil mapping using random forests analysis. *Geoderma*. 146: 1-2. 102-113.
11. Grinand, C., Arrouays, D., Laroche, B., and 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.
12. Guo, P.T., Li, M.F., Luo, W., Tang, Q.F., Liu, Z.W., and Lin, Z.M. 2015. Digital mapping of soil organic matter for rubber plantation at regional scale: an application of Random Forest plus residual kriging approach. *Geoderma*. 237-238: 1. 49-59.
13. Heung, B.C., Bulmer, C.E., and Schmitdt, M.G. 2014. Predictive soil parent material mapping at a regional-scale: A random forest approach. *Geoderma*. 214-215: 1. 141-154.
14. Ho, H.C., Knudby, A., Sirovyak, P., Xu, Y., Hodul, M., and Henderson, S.B. 2014. Mapping maximum urban air temperature on hot summer days. *Remote Sensing of Environment*. 154: 1. 38-45.
15. Jenny H. 1941. *Factors of Soil Formation, a System of Quantitative Pedology*. McGraw-Hill, New York, 281p.
16. Lagacherie, P. 2002. Cartographie de la diversité des sols viticoles de versant par imagerie à haute résolution: contribution à la connaissance des terroirs, Montpellier, France.
17. Lagacherie, P., Legros, J.P., and Burrough, P.A. 1995. A soil survey procedure using the knowledge on soil pattern of a previously mapped reference area. *Geoderma*. 65: 3-4. 283-301.
18. Mahler, P.J. 1970. *Manual of Multipurpose Land Classification*. Report no. 212. Soil and Water Research Institute, Tehran. Iran.
19. Mallavan, B.P., Minasny, B., and McBratney, A.B. 2010. Homosoil: a methodology for quantitative extrapolation of soil information across the globe. P 137-149. In: J.L. Boettinger (ed.) *Digital Soil Mapping: Bridging Research, Environmental Application, and Operation*. Springer, London.
20. Malone, B.P., Sanjeev, K.J., Minasny, B., and McBratney, A.B. 2016. Comparing regression-based digital soil mapping and multiple-point geostatistics for the spatial extrapolation of soil data. *Geoderma*. 262: 1. 243-253.
21. McBratney, A.B., Mendonça Santos, M.L., and Minasny, B. 2003. On digital soil mapping. *Geoderma*. 117: 1-2. 3-52.
22. Mehnatkesh, A., Ayoubi, S., Jalalian, A., and Sahrawat, K.L. 2013. Relationships between soil depth and terrain attributes in a semi-arid hilly region in western Iran. *J. Moun. Sci*. 10: 1. 163-172.
23. Minasny, B., and McBratney, A.B. 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computer and Geoscience*. 32: 9. 1378-1388.
24. Minasny, B., and McBratney, A.B. 2007. Spatial prediction of soil properties using EBLUP with the Matern covariance function. *Geoderma*. 140: 1. 324-336.
25. Moore, I.D., Gessler, P., Nielsen, G., and Peterson, G. 1993. Soil attribute prediction using terrain analysis. *Soil Sci. Soc. Amer. J.* 57: 2. 443-452.
26. Pahlavan Rad, M.R., Toomanian, N., Khormali, F., Brungard, C.W., Komaki, C.B., and Bogaert, P. 2014. Updating soil survey maps using random forest and conditioned Latin hypercube sampling in the loess derived soils of northern Iran. *Geoderma*. 232-234: 1. 97-106.

27. RStudio. 2015. RStudio: Integrated Development Environment for R, Boston, MA. Available at <http://www.r-studio.com>. (Visited 20 November 2018).
28. Saga Development Team. 2011. System for Automated Geoscientific Analyses (SAGA). Available at <http://saga-gis.org/en/index.html> (visited 12 August 2012).
29. Schoeneberger, P.J., Wysocki, D.A., Benham, E.C., and Broderson, W.D. 2012. Field book for describing and sampling soils, version 3.0. USDA Natural Resources Conservation Service, National Soil Survey Center, Lincoln, NE.
30. Sim, J., and Wright, C.C. 2005. The kappa statistic in reliability studies: use, interpretation and sample size requirements. *Physical Therapy*. 85: 3. 257-268.
31. Soil and Water Research Institute. 1999. Semi detailed soil survey of Saadat Shahr, Sivand, Seydan and Arsenjan. Soil and Water Research Institute of Iran, Ministry of Agriculture, Tehran, Iran. (In Persian)
32. Soil Survey Staff. 2014. Keys to soil taxonomy, 12<sup>th</sup> edition. USDA Natural Resources Conservation Service.
33. Taghizadeh-Mehrjardi, R., Nabiollahi, K., Minasny, B., and Triantafyllis, J. 2015. Comparing data mining classifiers to predict spatial distribution of USDA-family soil groups in Baneh region, Iran. *Geoderma*. 253-254: 1. 67-77.
34. Thompson, J.A., Pena-Yewtukhiq, E.M., and Grove, J.H. 2006. Soil-landscape modeling across a physiographic region: topographic patterns and model transportability. *Geoderma*. 133: 1-2. 57-70.
35. United State Department of Agriculture. Soil Conservation Service. 1993. Soil survey manual. Soil Survey. Div. Staff. US. Department of Agriculture. Handbook. 18. Washington, DC.
36. Zhu, A.X., Hudson, B., Burt, J., Lubich, K., and Simonson, D. 2001. Soil mapping using GIS, expert knowledge, and fuzzy logic. *Soil Sci. Soc. Amer. J.* 65: 5. 1463-1472.
37. Zhu, A.X., Liu, J., Du, F., Zhang, S.J., Qin, C.Z., Burt, J., Behrens, T., and Scholten, T. 2015. Predictive soil mapping with limited sample data. *Europ. J. Soil Sci.* 66: 1. 535-547.



---

## Evaluating extrapolation of random forest model to predict soil classes at subgroup level

M. Jamshidi<sup>1</sup>, \*M.A. Delavar<sup>2</sup>, R. Taghizadeh-Mehrjerdi<sup>3</sup> and C. Brungard<sup>4</sup>

<sup>1</sup>Ph.D. Student, Dept. of Soil Science, University of Zanjan and Scientific Staff of Soil and Water Research Institute (SWRI), Agricultural Research, Education and Extension Organization (AREEO),  
<sup>2</sup>Associate Prof., Dept. of Soil Science, University of Zanjan, <sup>3</sup>Assistant Prof., Dept. of Soil Science, University of Ardakan, <sup>4</sup>Assistant Prof., Dept. of Plant and Environmental Sciences, New Mexico State University  
Received: 10.24.2018; Accepted: 01.12.2019

---

### Abstract

**Background and Objectives:** Many soil maps that produced in Iran are in medium scale related to the soil survey projects that have been done over the past six decades. In many cases, soil maps have not been updated due to the high cost of soil survey activities in conventional methods. A proposed solution to overcome limitations of the conventional soil survey is digital soil mapping (DSM) that extensively used for producing soil maps in many countries recently. The extrapolation method in which soil pattern rules in reference area is used for soil class prediction in other areas as a cost-effective method have been mentioned by some soil surveyors. To achieve the main advantages of extrapolation in DSM, in this research we evaluated the use of random forest model in a reference area (donor area) for producing soil taxonomic classes at subgroup level in a site out of the reference area (recipient area).

**Materials and Methods:** In this study two neighboring areas in Fars Province in southern Iran were selected: 1) Saadat Shahr plain as donor site and 2) Seidan plain as recipient area. Two agricultural plain have a moderately similar environmental condition such as elevation, geology, physiography and climate and agriculture behavior. In donor area, 82 soil profiles were excavated, described and analyzed. Latin hypercube sampling (LHS) was used as a statistical method in donor area. In recipient area, 27 locations were determined on some parallel transects across the plain. All soils were classified according to USDA Soil Taxonomy (2014). Random forest (RF) in R statistical software was used to predict soil classes in donor area. Then the constructed model in donor area saved and applied to the recipient area. 25 variables related to soil forming factors consist of 1) primary and secondary train attributes and 2) remote sensing indices obtained from Landsat 8 satellite, OLI sensor imagery were used in this study. All auxiliary environmental covariate layers were resampled to a 30 m resolution. Producer's, users and overall accuracy and kappa index calculated according to the agreement of the field surveyed with predicted soil classes.

**Results:** Using RF algorithm from the 25 variables related to soil forming factors, five primary and secondary train attributes consist of slope, multiresolution index of valley bottom flatness (MRVBF), terrain ruggedness index, topographic wetness index and modified catchment area were selected as influential covariates. An overall accuracy of 72% and a Kappa index of 0.59 in the donor area, illustrating the relatively desirable agreement between observed and predicted soil classes. For extrapolating evaluation, the result of RF model with 70% of soil samples in the donor area was compared with the output of the transported RF model using 27 observations of the validation dataset. The overall accuracy of the external validation was 45% and the Kappa index was 0.28. Transferring the RF model constructed by all soil samples of the donor area

---

\* Corresponding Author; Email: amir-delavar@znu.ac.ir

(100%) showed a better result of soil prediction in the recipient area. The overall accuracy and the Kappa index of the external validation was 52% and 0.38, respectively. From the six soil subgroup classes, the best predicted classes were Typic Calcixerepts and Typic Xerorthents. Some classes were too sparse and the model was unable to predict them correctly.

**Conclusion:** The results showed that the model extrapolation in the framework of DSM could be a powerful tool for producing soil map in the area of Iran where soil maps are not available or updating the present soil maps are time consuming and costly. The low-cost and time saving method reported here, encourages soil surveyors to select model extrapolation for their survey activities.

**Keywords:** Digital soil mapping, Latin hypercube sampling (LHS), Soil forming factors