

پیش‌بینی مکانی گروه بزرگ‌های خاک با استفاده از مدل‌های رگرسیونی و درخت تصمیم در منطقه جنوب شرق ایران

فریده عباس زاده افشار^{۱*}، شمس‌اله ایوبی^۲ و اعظم جعفری^۳

- ۱- استادیار گروه علوم خاک، دانشکده کشاورزی، دانشگاه جیرفت، جیرفت، ایران
- ۲- استاد گروه علوم خاک، دانشکده کشاورزی، دانشگاه صنعتی اصفهان، اصفهان، ایران
- ۳- استادیار گروه علوم خاک، دانشکده کشاورزی، دانشگاه باهنر کرمان، کرمان، ایران

تاریخچه مقاله

دریافت: ۱۳۹۵/۱۱/۱۱

پذیرش نهایی: ۱۳۹۶/۱۱/۱۹

کلمات کلیدی:

نقشه برداری رقومی خاک،

گروه بزرگ خاک،

درخت تصمیم،

مدل رگرسیونی

* عهده دار مکاتبات

Email: f.abbaszade@yahoo.com

چکیده

نقشه توزیع مکانی کلاس‌های خاک برای استفاده مناسب از خاک و تصمیم‌گیری‌های مدیریتی مهم است. نقشه برداری رقومی خاک می‌تواند توزیع مکانی کلاس‌های خاک را به صورت کمی پیش‌بینی کند. ماشین یادگیری یک اصطلاح کلی برای مجموعه گسترده‌ای از مدل‌ها برای کشف الگوهای موجود در داده‌ها و پیش‌بینی متغیرهای مورد مطالعه است. این مطالعه با هدف مقایسه سه مدل رگرسیون لجستیک چندجمله‌ای (MLR)، رگرسیون درختی توسعه یافته (BRT)، درخت تصمیم (DT) و کارایی آن‌ها در پیش‌بینی گروه بزرگ‌های خاک در منطقه بوم استان کرمان طراحی شده است. یک الگوی نمونه برداری طبقه‌بندی شده تصادفی در منطقه‌ای به مساحت صد هزار هکتار تعریف شد و در نهایت، ۱۲۶ خاکرخ حفر و بر اساس سیستم رده‌بندی آمریکایی (۲۰۱۰) تشریح و طبقه‌بندی گردید. نتایج حاصل از مدل سازی نشان داد که نقشه سطوح ژئومرفولوژی، یک ابزار مهم در روش‌های نقشه برداری رقومی خاک است که به افزایش دقت پیش‌بینی کمک می‌کند. پس از سطوح ژئومرفیک، اجزای سرزمین و شاخص‌های سنجش از دور به عنوان پارامترهای کمکی مؤثر شناخته شدند. شاخص‌های ارزیابی مدل‌ها از جمله صحت کلی، شاخص کاپا و درجه برابری، به ترتیب برای مدل MLR، ۰/۷۱، ۰/۶۵، ۰/۵۷، مدل BRT، ۰/۸۵، ۰/۸۱، ۰/۴۵ و مدل DT، ۰/۸۶، ۰/۸۴ و ۰/۴۸ محاسبه شد. نتایج مقایسه دقت ارزیابی مدل‌ها نشان داد که بهترین پیش‌بینی مربوط به مدل درخت تصمیم بود. این نتایج نشان می‌دهد که ساختار درختی ایجاد شده بین متغیر هدف و متغیرهای انتخاب شده در مدل، باعث افزایش دقت این مدل نسبت به مدل‌های رگرسیونی شده است. نتایج کلی نشان داد که نقشه برداری رقومی خاک، می‌تواند به عنوان یک روش ارزیابی منابع خاک استفاده شود. علاوه بر این، قابلیت اطمینان نقشه‌های برآورد شده می‌تواند شروع یک بحث جدید بین متخصصان منابع زمین و خاکشناسان باشد. این اطلاعات هم‌چنین می‌تواند برای تکمیل مجموعه داده‌های موجود در کشور نیز، استفاده شوند.

مقدمه

خاک، یک عامل کلیدی در تنظیم و اصلاح چرخه ماده و انرژی در سطح زمین است؛ در نتیجه پیش شرط لازم برای ارزیابی مدیریت محیط زیست و زمین، به دست آوردن اطلاعات کافی در مورد آن خواهد بود (۲۶). روش‌های استاندارد و معمول برای تهیه نقشه ویژگی‌های خاک، بر مشاهده‌های و آنالیز نقطه‌ای خاک بنا نهاده شده‌اند، که این روش‌ها از نظر زمانی و اقتصادی برای زمین‌نماهای^۱ بزرگ، مقرون به صرفه نیستند. هم‌چنین هیچ معیار آماری برای نمونه‌برداری خاک در نقشه‌برداری سنتی^۲ وجود ندارد و ممکن است منجر به نمونه‌برداری اریب در بعضی مناطق شود (۱۸)؛ بنابراین، نقشه‌های استخراج شده از این روش، نمی‌توانند ویژگی‌های دینامیکی و پیوسته زمین‌نما را نشان دهند (۲۹).

در پاسخ به مشکلات سیستم نقشه‌برداری قدیم، به تدریج روش‌های کمی توسعه یافتند. روش‌های کمی قادر به بیان توزیع جغرافیایی خاک‌ها به شکل بهتر و دقیق‌تر می‌باشند. این روش‌ها تحت عنوان پدومتری^۳ در خاک‌شناسی معرفی شده‌اند. در حقیقت، نقشه‌برداری با روش‌های پدومتری، نقشه‌برداری رقومی^۴ نیز نامیده می‌شود (۲۰).

نقشه‌برداری رقومی خاک، در واقع توسعه یک مدل عددی یا آماری از رابطه بین متغیرهای محیطی و ویژگی‌های خاک است که به منظور تولید نقشه رقومی به کار می‌رود (۲۸). نقشه‌برداری رقومی خاک، سامانه‌هایی را برای اطلاعات مکانی خاک، گردآوری و ایجاد می‌کند که می‌تواند کاربران را در تصمیم‌سازی برای رسیدگی به مسائل و مشکلات محیطی و کشاورزی کمک نماید (۱۲). یکی از جنبه‌های اساسی در نقشه‌برداری رقومی خاک، استفاده از مدل‌های مختلف به منظور ساده‌سازی

پیچیدگی‌های موجود در سامانه طبیعی خاک است. بر این اساس، مدل‌های خاک - زمین‌نما، بیانگر شکل ساده شده‌ای از روابط پیچیده موجود بین خاک و زمین‌نما می‌باشند که فرآیندهای تکاملی خاک و الگوی پراکنش آن را نشان می‌دهند (۷).

ینی^۵ (۱۱) خاک را به عنوان تابعی از پنج فاکتور آب و هوا، جانداران، توپوگرافی، مواد مادری و زمان توصیف کرد. او معادله‌هایی برای توصیف عوامل و فاکتورهای مؤثر بر ویژگی‌های خاک تنظیم کرد. مک‌براتنی و همکاران^۶ (۱۵) با مروری بر مدل‌های مختلف تخمین و برآورد مکانی خاک و با توجه به نکات کلیدی مذکور، اقدام به ارائه یک چارچوب تعمیم‌یافته برای تخمین و برآورد مکانی خاک نمودند؛ به عبارت دیگر، مدل پنج فاکتوری ینی را به یک مدل هفت فاکتوری از طریق کمی‌سازی و بهره‌مندی از مدل‌های رقومی ارتفاع، سنجش از دور، سنسورهای سنجش از نزدیک خاک و دیگر تکنیک‌های پهنه‌بندی، بسط و تعمیم دادند (۱۵). پایه و اساس نقشه‌برداری رقومی خاک مبتنی بر معادله اسکورپن^۷ می‌باشد. در این معادله ویژگی‌های یا کلاس خاک با یک مجموعه متغیرهای محیطی (متغیرهای کمکی) ارتباط داده می‌شود. جزء اصلی معادله اسکورپن، مدل ارتباط‌دهنده متغیرهای محیطی و متغیر وابسته (مدل پیش‌بینی کننده)^۸ می‌باشد:

$$S_c = f(s, c, r, o, p, a, n) + e \quad (1)$$

که در معادله اسکورپن S_c : ویژگی یا کلاس‌های خاک، s : مربوط به اطلاعات خاک که از نقشه خاک یا پایگاه داده یا دانش کارشناس به دست می‌آید. c : اقلیم، o : جانداران از جمله فعالیت‌های انسانی و پوشش گیاهی، r : پستی و بلندی یا توپوگرافی، p : مواد مادری، a : زمان و n : موقعیت مکانی است. تابع f یک طبقه‌بندی نظارت شده یا برنامه یادگیری نظارت شده را نشان می‌دهد (۱۵). مدل

5- Jenny

6- McBratney et al.

7- Scorpan

8- Predictive model

1- Landscape

2- Traditional soil mapping

3- Pedometry

4- Digital soil mapping

خاک قابل دسترس در کشور ایران وجود دارد، و بسیاری از داده‌های جمع‌آوری شده منتشر نشده‌اند. این نوع محدودیت داده، یک مسئله مشترک در بسیاری از مناطق جهان است که در آن توزیع مکانی اطلاعات خاک کیمیا و یا ناقص هستند. با توجه به نیاز کشور به مجموعه داده‌های رقومی اطلاعات خاک، این مطالعه با هدف کاربرد روش‌های نقشه‌برداری رقومی خاک (رگرسیون لاجیستیک چندجمله‌ای^۱، رگرسیون درختی توسعه‌یافته^۲ و درخت تصمیم) برای پیش‌بینی گروه بزرگ‌های خاک و ارزیابی کارایی این مدل‌ها در منطقه بوم استان کرمان طراحی گردید.

مواد و روش‌ها

منطقه مطالعاتی

منطقه مورد مطالعه بین عرض‌های جغرافیایی ۲۸ درجه و ۵۲ دقیقه تا ۲۹ درجه و ۹ دقیقه شمالی و طول‌های جغرافیایی ۵۸ درجه و ۴ دقیقه تا ۵۸ درجه و ۲۸ دقیقه شرقی در جنوب شرقی ایران و در شرق استان کرمان قرار گرفته است (شکل ۱). این منطقه، مساحتی حدود یک‌صد هزار هکتار را می‌پوشاند. زمین‌نماهای اصلی این منطقه، شامل مخروط‌افکنه، پلایا، باهادا، تپه و کوه می‌باشند. میانگین بارندگی سالیانه این منطقه ۶۴ میلی‌متر، متوسط درجه حرارت آن ۲۲ درجه سانتی‌گراد، میانگین سالیانه تبخیر با تشکک تبخیر کلاس A در آن، ۳۰۰۰ میلی‌متر در دوره آماری ۲۰ ساله (۱۳۹۲-۱۳۷۱) محاسبه شد. هم‌چنین میانگین ارتفاع منطقه ۱۰۵۰ متر از سطح دریاست (۱۶). این منطقه بر اساس سیستم رده‌بندی آمریکایی با توجه به نقشه رژیم‌های حرارتی و رطوبتی خاک ایران، دارای رژیم رطوبتی اریدیک و رژیم حرارتی هایپرترمیک است (۲). این منطقه توسط کوه‌های آهکی و آتشفشانی از شمال غرب تا جنوب شرق احاطه شده است.

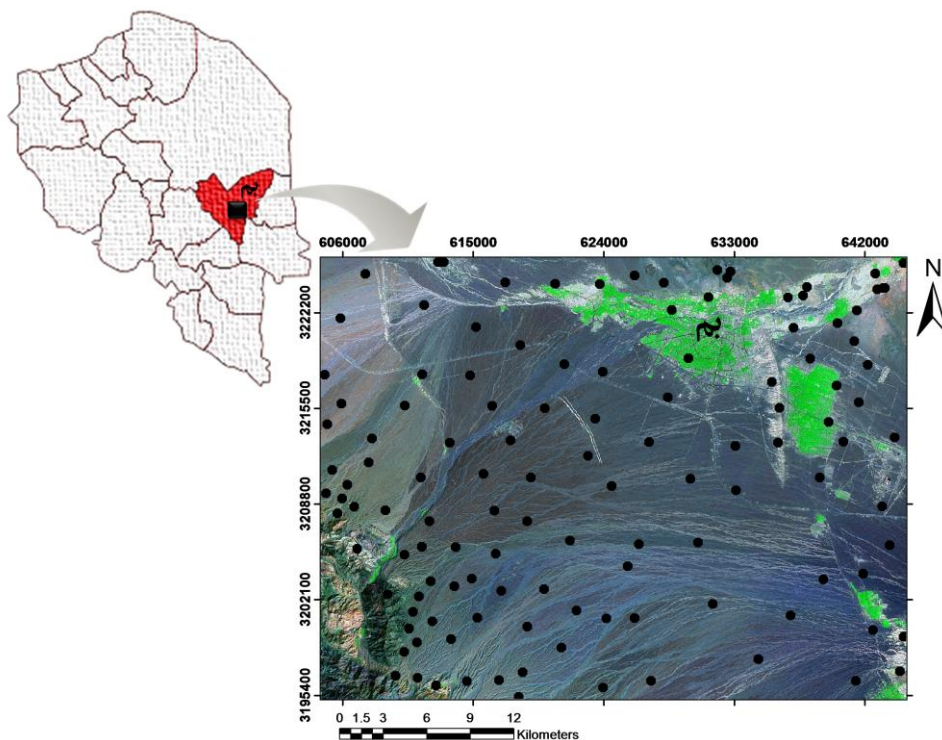
اسکورپن برای پیش‌بینی کمی کلاس‌های خاک و ویژگی‌های پیوسته خاک بر اساس مشاهدات تجربی استفاده می‌شود. علاوه بر این، در مدل‌سازی‌های هیدرولوژی و تناسب اراضی، یکی از فاکتورهای کلیدی خاک و یا همان نقشه خاک است. برای پیش‌بینی ویژگی‌های خاک یا کلاس خاک، مدل‌های مختلفی مورد بررسی قرار گرفته‌اند که از آن جمله می‌توان به رگرسیون چند متغیره^۱، رگرسیون غیرخطی^۲، شبکه عصبی مصنوعی^۳، رگرسیون درختی^۴، سیستم فازی^۵، درخت تصمیم^۶ و الگوریتم ژنتیک^۷ اشاره کرد.

کمپن و همکاران (۱۲) از روش رگرسیون لاجیستیک چندمتغیره برای کمی کردن روابط بین متغیرهای کمکی و گروه بزرگ خاک استفاده کردند. هدف از این تحقیق، به‌روزرسانی نقشه ملی خاک بدون کار صحرایی اضافی از طریق نقشه‌برداری رقومی خاک با استفاده از داده‌های قبلی خاک بود. سان و همکاران (۲۲) از دو روش درخت تصمیم و تفکیک خطی برای نقشه‌برداری رقومی رده‌های خاک در مناطق کوهستانی هنگ‌کنگ استفاده کردند. نتایج این مطالعه اهمیت انتخاب متغیرهای کمکی خاک در مدل‌سازی برای نقشه‌برداری خاک را نشان داد. در ایران، برای پیش‌بینی گروه‌های بزرگ خاک از تکنیک‌های نقشه‌برداری رقومی خاک، از جمله رگرسیون چندجمله‌ای، رگرسیون درختی، درخت تصمیم و شبکه عصبی مصنوعی توسط محققان متعددی مانند اسفندیارپور و همکاران (۶)، جعفری و همکاران (۱۰)، تقی‌زاده و همکاران (۲۳ و ۲۵) و پهلوان‌زاد و همکاران (۱۷) انجام گرفته است.

با توجه به محدودیت‌های روش نقشه‌برداری سنتی که به آن‌ها اشاره شد، اطلاعات محدودی از توزیع مکانی

- 1- Multivariate regression
- 2- Nonlinear regression
- 3- Artificial neural network
- 4- Tree regression
- 5- Fuzzy logic
- 6- Decision tree (DT)
- 7- Genetic algorithms

- 8- Multinomial regression logistic (MLR)
- 9- Boosted regression trees (BRT)



شکل (۱) موقعیت جغرافیایی منطقه مورد مطالعه در شهرستان بام، استان کرمان و نقاط نمونه‌برداری
Figure (1) Location of the study area and sampling points in Bam, Kerman province

هشت گروه بزرگ خاک شناسایی شد که شامل هاپلوسالیدز^۲، هاپلوکلسیدز^۳، هاپلوچیپسیدز^۴، هاپلو کمبیدز^۵، کلسی‌چیپسیدز^۶، کلسی‌آرجیدز^۷، پتروکلسیدز^۸ و توری‌ارتنتز^۹ هستند.

متغیرهای کمکی

از یک مدل رقومی ارتفاع^{۱۰} که از وب سایت مدل رقومی ارتفاع جهانی استر^{۱۱} به دست آمد، برای تعیین اجزاء سرزمین و از تصویر ماهواره‌ای ETM+ برای تعیین شاخص‌های گیاهی، شاخص رس و شاخص شوری استفاده گردید (۲۷). اجزاء سرزمین، حاصل

زمین‌شناسی منطقه بیش‌تر شامل رسوبات قدیمی و اخیر دوران کواترنر است. از مهم‌ترین کاربری اراضی منطقه مورد مطالعه، می‌توان به کشاورزی آبی، دیم و مرتع اشاره کرد.

مطالعات میدانی و نمونه‌برداری

یک الگوی نمونه‌برداری طبقه‌بندی‌شده تصادفی^۱ با استفاده از نقشه‌های رقومی زمین‌شناسی، ژئومورفولوژی (نقشه سطوح ژئومرفیک) و توپوگرافی (نقشه ارتفاع رقومی) پیاده گردید. در این مطالعه، لایه نمونه‌برداری منطبق بر تفاوت در زمین‌ریخت‌ها (نقشه ژئومورفولوژی)، توپوگرافی (مدل رقومی ارتفاع) و زمین‌شناسی تعریف شد. در نهایت ۱۲۶ خاک‌رخ حفر و تشریح گردید. تمامی مقاطع بر اساس معیارهای رده‌بندی آمریکایی (۲۱) تشریح، طبقه‌بندی و مورد نمونه‌برداری قرار گرفتند. با تشریح خاک‌رخ‌های حفر شده بر اساس سیستم رده‌بندی آمریکایی در منطقه مورد مطالعه،

-
- 2- Haplosalids
 - 3- Haplocalcids
 - 4- Haplogypsids
 - 5- Haplocambids
 - 6- Calcigypsids
 - 7- Calscigypsids
 - 8- Petrocalcids
 - 9- Torriorthents
 - 10- Digital elevation model
 - 11- Aster Global Digital Elevation Model

-
- 1- Stratified sampling scheme

مدل درخت تصمیم با استفاده از بسته "C50" در نرم افزار R برآزش داده شد.

رگرسیون درختی توسعه یافته از مجموعه یادگیری ماشین، ترکیبی از دو تکنیک آماری بوستینگ^۲ و رگرسیون درختی است (۵). بوستینگ یک روش مرحله‌ای رو به جلو است که در آن، مدل‌های درختی به صورت تکرارپذیر با زیرمجموعه‌ای از داده‌های آموزشی برآزش داده می‌شوند. در برآزش رگرسیون درختی توسعه یافته، باید دو پارامتر، نرخ یادگیری و پیچیدگی درخت مشخص گردند. نرخ یا مقدار یادگیری، سهم هر درخت متوالی را در مدل نهایی تعیین می‌کند. پیچیدگی درخت، اثرات اصلی یا اثرات متقابل بین متغیرها را نشان می‌دهد.

مدل‌های رگرسیون درختی توسعه یافته با استفاده از بسته "gbm"^۳ در نرم افزار R برآزش داده شدند. برای برآزش مدل، پارامترهای قراردادی کد BRT استفاده شد که نرخ یادگیری ۰/۰۱، پیچیدگی درخت ۱ و اعتبار سنجی متقابل ۱۰۵ مرتبه بود.

رگرسیون لجستیک، جزء مدل‌های خطی تعمیم یافته است که می‌تواند برای توصیف رابطه چندین متغیر مستقل (X) با یک متغیر وابسته دو یا چند حالتی (Y) مورد استفاده قرار گیرد. در این مطالعه، از مدل رگرسیون لجستیک چندجمله‌ای برای پیش‌بینی گروه بزرگ‌های خاک (متغیر هدف) استفاده شد. این مدل در نرم افزار R با استفاده از بسته nnet قابل اجراست. اهمیت ضریب رگرسیون هر متغیر پیش‌بینی کننده برای هر متغیر با استفاده از معیار آماری والد ارزیابی گردید. این معیار در واقع از تقسیم ضریب رگرسیون هر متغیر مستقل به خطای استاندارد آن به دست می‌آید. بنابراین، می‌توان اهمیت هر پیش‌بینی کننده را استنباط نمود؛ به این معنی که هر چه مقدار مطلق این پارامتر بزرگتر باشد، اهمیت متغیر پیش‌بینی کننده بیش تر می‌شود (۹).

ارزیابی دقت مدل‌های پیش‌بینی کننده

برای بررسی صحت مدل مورد استفاده، داده‌ها به طور تصادفی به داده‌های آموزشی و اعتبارسنجی تقسیم گردیدند. داده‌های آموزشی، ۸۰ درصد و داده‌های اعتبارسنجی، ۲۰ درصد کل داده‌ها را شامل شدند. هر مدل با داده‌های آموزشی برآزش داده شد و سپس پیش‌بینی برای داده‌های اعتبارسنجی انجام گردید.

از مدل رقومی ارتفاع و شاخص‌های سنجش از دور در جدول ۱ ارائه شده‌اند. متغیرهای کمکی در محیط سامانه جغرافیایی SAGA محاسبه و استخراج شدند. روش استخراج پارامترهای مزبور بر مبنای منبع هنگل و همکاران^۱ (۸) بود. شایان ذکر است که تصاویر ماهواره‌ای و مدل رقومی ارتفاع استفاده شده دارای ابعاد ۳۰×۳۰ متر بودند. شاخص‌های سنجش از دور و اجزاء سرزمین، نقشه ژئومرفولوژی و نقشه خاک در نرم افزار SAGA به سیستم مرجع جهانی یکسانی (WGS 84 UTM 40) تبدیل شدند.

مدل سازی

درخت تصمیم یک نوع طبقه‌بندی کننده نظارت شده سلسله مراتبی است که جزء روش‌های دانش پایه طبقه‌بندی به شمار می‌آید؛ به عبارت دیگر، پدیده‌های موضوعی در این روش با داشتن اطلاعات در مورد آن‌ها از بالا به پایین با اتخاذ تصمیمات مناسب، کلاس بندی می‌شوند.

مدل درختی، یک روش ناپارامتری الگوریتمی است که قادر به پیش‌بینی متغیرهای کمی یا متغیرهای طبقه بندی شده (کلاس‌های خاک) بر اساس مجموعه‌ای از متغیرهای پیش‌بینی کننده کمی و کیفی است. برای اولین بار این روش توسط بریمن و همکاران (۳) گسترش داده شد. درخت تصمیم، یکی از جدیدترین روش‌های داده کاوی می‌باشد که مشاهده‌ها و اندازه گیری‌های انجام شده در مورد مقادیر پارامترهای تأثیرگذار بر مقدار یک پدیده مورد نظر را به شکل قوانینی در آورده و از آن‌ها به منظور پیش‌بینی استفاده می‌کند. درخت تصمیم با پرسش یک مجموعه سوال و تعیین مقادیر ممکن برای جواب سوال، داده‌های ورودی از ریشه را در مسیرهای موجود حرکت داده و در نهایت به یک برگ می‌رساند. داده‌های قرار گرفته در برگ همگن، ویژگی یکسانی دارند و آن ویژگی برگ است (۳).

در روش درخت تصمیم، مجموعه‌ای از شرط‌های منطقی، به صورت یک الگوریتم با ساختار درختی برای طبقه بندی یا پیش‌بینی کمی یک متغیر به کار می‌رود. ایجاد درخت تصمیم دو مرحله دارد: مرحله اول، مرحله ایجاد و رشد درخت است که این مرحله شامل پیوند و انشعاب است. مرحله دوم، مرحله توقف و هرس است. هدف مرحله اخیر، به حداقل رساندن خطای پیش‌بینی است (۲۳).

2- Boosting

3- Generalized Boosted Regression Models

1- Hengl et al.

عباس زاده افشار و همکاران: پیش‌بینی مکانی گروه بزرگ‌های خاک...

جدول (۱) ویژگی‌های سرزمینی مورد استفاده به عنوان متغیرهای پیش‌بینی کننده در مدل سازی

Table (1) Train attributes used for spatial prediction

نام متغیر (Variable name)	راه‌نما (Legend)	ماهیت فاکتور (Nature factor)	داده‌های محیطی (Auxiliary data)
ارتفاع (Elevation)	EL	توپوگرافی (Topography)	مدل رقومی ارتفاع (Digital elevation model)
شیب (Slope angle)	SL		
جهت شیب (Aspect)	As		
انحنای میانگین (Mean curvature)	MCur		
انحنای نیمرخ (Profile curvature)	PrCur		
انحنای سطح (Plan curvature)	PICur		
شاخص رطوبتی (Topographic wetness index)	TWI		
شاخص همواری دره با درجه تفکیک بالا (Multi-resolution valley bottom flatness index)	MrVBF		
شاخص قدرت جریان (Stream Power Index)	SPI		
شاخص گیاهی نرمال شده (Normalized difference vegetation Index)	NDVI	انعکاس نور (Reflection)	تصویر لندست (ETM+) (Landsat image)
شاخص نسبت گیاهی (Ratio Vegetation Index)	RVI		
شاخص گیاهی عمودی (Perpendicular Vegetation Index)	PVI		
شاخص رس (Clay Index)	CI		
شاخص شوری (Salinity Index)	SI		
شاخص گچ (Gypsum Index)	GI		
شاخص گیاه & خاک (Soil & Vegetation Index)	SAI		
شاخص کربنات (Carbonate Index)	CaI		
۱۷ سطح ژئومورفیک (17 geomorphic surfaces)	GS	ژئومورفولوژی (Geomorphology)	نقشه ژئومورفولوژی (Geomorphology map)
۳۰ واحد زمین‌شناسی (30 geological units)	Geol	زمین‌شناسی (Geology)	نقشه زمین‌شناسی (Geology map)

تولید کننده^۳، شاخص کاپا^۴ و درجه برابری^۵ استفاده گردید.

اعتبارسنجی با مقایسه داده‌های مشاهده شده و پیش‌بینی شده صورت گرفت. یک آرایه خطا برای محاسبه معیارهای دقت از جمله صحت کلی^۱، صحت کاربر^۲، قابلیت اطمینان

- 2- User accuracy (UA)
- 3- Producer accuracy (PA)
- 4- Kappa index
- 5- Brier scores

- 1- Overall accuracy (OA)

نتایج و بحث

مدل سازی گروه بزرگ های خاک

در جدول ۲ و شکل ۲ متغیرهای کمکی مورد استفاده در پیش بینی گروه بزرگ های خاک توسط مدل رگرسیون لاجیستیک چندجمله ای، رگرسیون درختی توسعه یافته و درخت تصمیم ارائه شده است. نتایج حاصل از این مدل ها نشان داد مهم ترین متغیری که تغییرات کلاس های خاک را تشریح می کند، سطح ژئومرفیک است. بعد از سطوح ژئومرفیک، پارامترهای سرزمین و داده های سنجش از دور در پیش بینی گروه بزرگ های خاک شرکت داشتند. این نتایج نشان داد که سطح ژئومورفولوژیک یک پارامتر بسیار مهم است که علت آن را می توان به این که، سطوح ژئومورفولوژی اخیراً تشکیل شده اند و ارتباط خوبی با فرآیندهای خاک سازی منطقه دارند نسبت داد. اثر فرآیندهای ژئومورفولوژی روی الگوی توزیع خاک ها و توسعه آن ها توسط تعدادی از محققان بررسی شده است. به عنوان مثال اسکول و همکاران^۱ (۱۹)، جعفری و همکاران (۱۰) و تقی زاده و همکاران (۲۴) تأکید کردند که سطوح ژئومورفیک مهم ترین متغیر کمکی برای شناسایی خاک ها در مناطق بیابانی و خشک است.

در مناطق با توپوگرافی متغیر، مشخصه های توپوگرافی، معمولاً پیش بینی کننده مناسبی در مدل سازی هستند. اغلب این ویژگی ها به عنوان پیش بینی کننده کلیدی در نقشه برداری رقومی خاک استفاده می شوند (۱۵). در این مطالعه، ویژگی های سرزمین، به دلیل این که توپوگرافی یک فاکتور مهم در خاک سازی است، به عنوان یکی از فاکتورهای اصلی در پیش بینی گروه های بزرگ خاک انتخاب شدند. حضور ویژگی های پستی و بلندی به همراه فرآیندهای ژئومرفیک در پیش بینی گروه های بزرگ خاک، نشان داد که این فاکتورها بر آید نیروهای مؤثر بر تشکیل خاک در منطقه مورد مطالعه هستند.

هم چنین شاخص های سنجش از دور، از جمله شاخص پوشش گیاهی و خاک (SAVI)، شاخص رس، شاخص شوری، شاخص گچ و کربنات هم جزء متغیرهای کمکی، قابلیت مناسبی در پیش بینی کلاس های خاک داشتند. آکسوی و همکاران^۲ (۱) گزارش کردند ترکیب اجزای سرزمین و داده های سنجش از دور، ابزاری قدرتمند در نقشه برداری خاک است.

صحت کلی ارتباط بین همه داده های مورد استفاده و

داده های طبقه بندی شده (ff,tt) را تشریح می کند:

$$\text{Overall accuracy} = \frac{tt + ff}{tt + ft + tf + ff} \quad (2)$$

صحت کاربر، شامل کلاس های حضور صحیح پیش بینی شده (tt) به مجموع کلاس های حضور صحیح پیش بینی شده و کلاس های عدم حضور که به غلط جزء کلاس های حضور مشاهده شده، پیش بینی شدند (tf) می باشد:

$$\text{users' accuracy} = \frac{tt}{tt+tf} \quad (3)$$

قابلیت اطمینان تولید کننده ارتباط بین همه کلاس های صحیح پیش بینی شده و مجموع کلاس های صحیح پیش بینی شده کلاس های حضور مشاهده شده که به غلط جزء کلاس های عدم حضور پیش بینی شدند (ft) می باشد:

$$\text{producers' reliability} = \frac{tt}{tt+ft} \quad (4)$$

شاخص کاپا، یک شاخص قوی است که نسبت احتمال حضور یا عدم حضور کلاس، را که به درستی به وسیله مدل پیش بینی شدند، محاسبه می کند؛ بنابراین شاخص کاپا همیشه کمی کم تر از خلوص نقشه است.

$$\text{kappa index} = \frac{\text{observed accuracy} - \text{chance agreement}}{1 - \text{chance agreement}} \quad (5)$$

که در آن، دقت مشاهده به عنوان مجموع ارزش قطر در ماتریس و شانس خطا شامل مجموع ارزش غیر قطر ماتریس است. شاخص کاپا بالاتر نشان دهنده عملکرد مدل بالاست. دامنه تغییرات آماره کاپا بین صفر تا یک است. شاخص کاپا بالاتر از ۰/۸، ۰/۴-۰/۸ و کم تر از ۰/۴ به ترتیب نشان دهنده توافق قوی، متوسط و ضعیف هستند (۲۴).

درجه برابری تفاوت بین کلاس واقعی و احتمال تخمین زده شده از کلاس واقعی را محاسبه می کند:

$$BS = \frac{1}{n} \sum_{j=1}^r \sum_{i=1}^n (F_{ij} - E_{ij})^2 \quad (6)$$

کلاس داده های تست، F_{ij} برآورد احتمال تعداد مشاهده های n_j متعلق به کلاس i و E_{ij} یک متغیر شاخص، از جمله $E_{ij}=1$ اگر n_j یک گروه بزرگ خاک و ۰ دیگر گروه های بزرگ خاک است (۴). یک نقشه با صحت خوب دارای مقادیر صحت کلی، صحت کاربر، قابلیت اطمینان تولید کننده و شاخص کاپا نزدیک به یک و درجه برابری پایین است.

1- Scull et al.

2- Aksoy et al.

عباس زاده افشار و همکاران: پیش‌بینی مکانی گروه بزرگ‌های خاک...

جدول (۲) متغیرهای مورد استفاده در پیش‌بینی گروه بزرگ خاک توسط مدل MLR و BRT
Table (2) Selected variables in soil great group prediction by MLR and BRT model

متغیرهای مورد استفاده در مدل* (Used Variables in model)	مدل (model)
PVI, EI, NDVI, WI, CaI, GS, MrVBF,	رگرسیون لاجیستیک چندجمله‌ای (Multinomial regression logistic)
متغیرهای مورد استفاده (Used Variables)	گروه‌های بزرگ خاک (Soil great groups)
GS+MrVBF+EL	هاپلوسالیدز (Haplosalids)
GS+SI+WI+SAVI+SP	هاپلوچیسیدز (Haplogypsids)
GS+MrVBF+EL	هاپلوکلیدز (Haplocalids)
GS+SAVI+SI+GI+CI+EL	هاپلوکمیدز (Haplocambids)
GS+EL+CaI+As	کلسی‌چیسیدز (Calcigypsids)
GS+EL+CI+SL+SR+RVI	کلسی‌آرچیدز (Calciargids)
GS+MrVBF+DI+EL	پتروکلیدز (Petrocalcids)
GS+CaI+As+EL+TWI	توری‌ارتنتز (Torriorthents)

* علائم اختصاری ارائه شده به طور کامل در جدول ۱ تعریف شده‌اند.

*The abbreviated symbols are described in Table 1.

خاک، گروه بزرگ هاپلوکلیدز، کم‌ترین دقت پیش‌بینی را در داده‌های اعتبارسنجی با دقت کاربر ۰/۲۵ با مدل رگرسیون لاجیستیک چندجمله‌ای، دارا بود. با این حال، مدل‌های درختی نسبت به مدل‌های رگرسیونی بهتر عمل کردند؛ زیرا شامل یک ساختار درختی هستند و در نتیجه پیچیدگی مدل را افزایش می‌دهد؛ بنابراین به نظر می‌رسد که نه تنها روابط خاک - محیط پیچیده‌اند، بلکه دارای یک ساختار سلسله مراتبی نیز می‌باشند. اگر چه دلیل ساختار متغیر سلسله مراتبی مشخص نیست، اما ممکن است تا حدی به اثرات فرآیندهای خاکسازي در مقیاس‌های مختلف مربوط باشد. دقت کاربر و قابلیت اطمینان تهیه‌کننده، به ما اجازه می‌دهد که سطح بیش‌برآورد^۱ و کم‌برآورد^۲ در پیش‌بینی گروه‌های بزرگ خاک را تخمین بزنیم (۱۴).

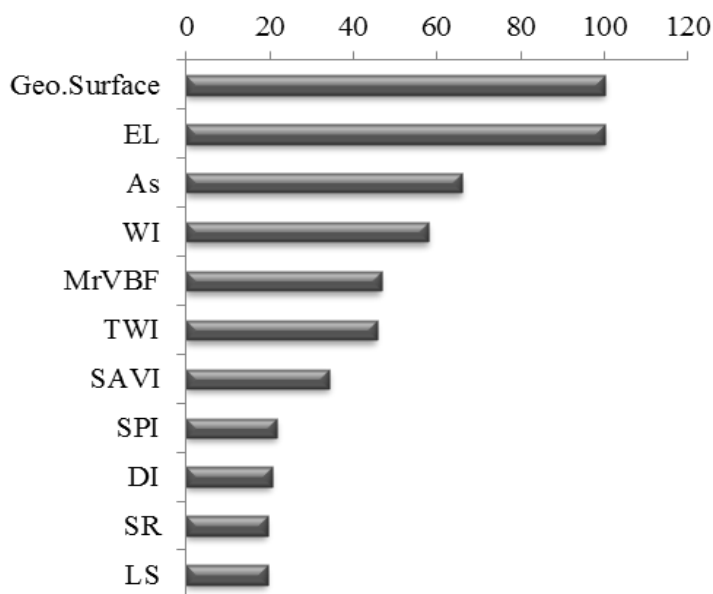
ویژگی‌های مختلف مدل رگومی ارتفاع هم به لحاظ منطقی و ریاضی و هم از نظر تجربی، دارای رابطه نزدیکی با ویژگی‌های محیطی و خاک هستند. استفاده از این ویژگی‌ها در شناسایی خاک می‌تواند تا حد زیادی بر دقت کار بیفزاید و از صرف هزینه و زمان بکاهد.

مقایسه دقت ارزیابی مدل‌های برازش داده شده

شاخص‌های دقت پیش‌بینی گروه بزرگ‌های خاک حاصل از مدل‌های برازش شده در داده‌های اعتبارسنجی، در جدول ۳ ارائه شده است. نتایج نشان می‌دهد که بیش‌ترین دقت برای پیش‌بینی گروه‌های بزرگ خاک مربوط به کلاس‌های هاپلوسالیدز، کلسی‌چیسیدز و پتروکلیدز با دقت کاربر ۱ در داده‌های اعتبارسنجی است. نکته قابل توجه این بود که این گروه‌های بزرگ توسط هر سه مدل برازش داده‌شده با دقت بالا پیش‌بینی شدند، که احتمالاً این موضوع را می‌توان به انتخاب مناسب متغیرهای پیش‌بینی‌کننده توسط این مدل‌ها نسبت داد. در بین کلاس‌های

1- Overestimate

2- Underestimate



شکل (۲) نتایج اهمیت متغیرهای محیطی در پیش‌بینی گروه‌های بزرگ خاک بر حسب درصد توسط مدل درخت تصمیم (DT) علائم اختصاری ارائه شده به طور کامل در جدول ۱ تعریف شده‌اند.

Figure (2) Selected variables in soil great group prediction by decision tree model the abbreviated symbols are described in Table 1.

باشند؛ به گونه‌ای که این مدل‌ها بتوانند با دنیای واقعی انطباق بیشتری داشته باشند.

نتایج این بررسی نشان داد دقت مدل درخت تصمیم در پیش‌بینی کلاس‌های خاک نسبت به مدل رگرسیونی بیشتر است. این نتایج نشان می‌دهد که ساختار درختی ایجاد شده بین متغیر هدف و متغیرهای انتخاب شده در مدل، باعث افزایش دقت این مدل‌ها نسبت به مدل‌های رگرسیونی شده است. البته شایان ذکر است که تفاوت معنی‌داری بین مدل‌هایی که دارای ساختار درختی هستند، در شاخص‌های دقت دیده نشد؛ بنابراین، مدل درخت تصمیم را می‌توان به‌عنوان بهترین مدل برای پیش‌بینی کلاس‌های خاک در سطح گروه بزرگ، در منطقه مورد مطالعه توصیه نمود. برانگراد و همکاران^۱ (۴) بیش‌ترین شاخص کاپا را مقادیر ۰/۵۵ و ۰/۶۹ برای کلاس‌های خاک در دو سایت مورد مطالعه با استفاده از مدل‌های مختلفی به‌دست آوردند. کیم و همکاران^۲ (۱۳) نیز، شاخص کاپا ۰/۷۹ را برای پنج مجموعه خاک در منطقه فلوریدا آمریکا محاسبه کردند. هم‌چنین تقی‌زاده و همکاران (۲۳) با استفاده از مدل‌های مختلف

دقت کاربر برای گروه‌های بزرگ هاپلوچیسیدز و هاپلوکمیدز در مدل MLR، هاپلوکمیدز و پتروکلسیدز در مدل BRT و توری‌ارتنتز در مدل DT بیش‌تر از قابلیت اطمینان تهیه‌کننده اندازه‌گیری شد؛ این مسئله نشان می‌دهد مدل‌ها این گروه‌های بزرگ را، در منطقه مورد مطالعه کم برآورد کرده‌اند. گروه بزرگ توری‌ارتنتز در مدل MLR، هاپلوکلسیدز در مدل BRT و هاپلوچیسیدز در مدل DT، با قابلیت اطمینان بیش‌تر از دقت کاربر واضح است که پیش‌بینی آن دارای بیش‌برآورد بوده است. نتایج مقایسه دقت ارزیابی مدل‌های برآورد داده شده در جدول ۴ ارائه شده است. نتایج دقت ارزیابی نشان داد که بهترین پیش‌بینی مربوط به مدل درخت تصمیم است. دقت کلی و مقادیر زیاد شاخص کاپا و درجه برآیر کم، تأییدکننده این مطلب است. همه روش‌های ذکر شده در این مطالعه بر فرض دقیق بودن متغیرهای مورد مطالعه و مشاهده‌های مربوط به آن‌ها استوار است و روابط بین متغیرها نیز، دقیق فرض می‌شود. حال آن‌که در سیستم‌های طبیعی مانند خاک، عموماً مشاهده‌ها یا روابط ناشناخته بین متغیرها مطرح است. در چنین شرایطی، از مدل‌هایی باید بهره گرفت که قادر به ارائه الگوهای مناسب‌تر

1- Brungard et al.

2- Kim et al.

عباس زاده افشار و همکاران: پیش‌بینی مکانی گروه بزرگ‌های خاک...

منطقه و تفسیر عکس‌های هوایی، جهت حرکت آبراهه‌ها و شیب منطقه به‌طرف زمین‌ریخت پلایا است و باعث تجمع رسوبات ریزدانه در این سطح شده است. در منطقه مورد مطالعه، بیش‌تر باغ‌های خرما در سطح ژئومرفیک پلایا قرار دارند. آبیاری غرقابی باغ‌های خرما، باعث شست‌وشو و خارج شدن نمک و گچ از این خاک‌ها شده و افق‌های کمبیک و کلسیک تشکیل گردیده‌اند. بر اساس این مشاهده‌ها، خاک‌های این سطح در گروه‌های بزرگ هاپلوکلسیدز و هاپلوکمبیدز طبقه‌بندی شدند.

همان‌طور که در شکل یاد شده مشاهده می‌شود، بیش‌ترین سطح مربوط به گروه بزرگ توری‌ارتنتز و هاپلوچیسیدز است. گروه بزرگ توری‌ارتنتز در کوه‌ها، رسوبات لایه‌بندی و مخروط‌افکنه و باهادای جدید و فعال دیده می‌شود. هم‌چنین گروه بزرگ هاپلوچیسیدز در مخروط‌افکنه بریده‌بریده و مناطقی از باهاداهایی که پایدار شده‌اند و سطح آن‌ها پوشیده از سنگ‌فرش بیابانی است، قابل مشاهده بود. این دو کلاس خاک، تعداد نقاط بیش‌تری نیز نسبت به سایر کلاس‌های خاک داشتند.

یادگیری ماشین دامنه شاخص کاپای ۰/۷۸-۰/۵۲ در گروه‌های بزرگ و دامنه ۰/۷۴-۰/۸۴ را برای رده خاک به‌دست آوردند. تقی‌زاده و همکاران (۲۵) از دو روش شبکه عصبی و درخت تصمیم برای پیش‌بینی کلاس‌های خاک استفاده کردند. نتایج آن‌ها نشان داد که مدل درخت تصمیم در جداسازی خاک‌ها کارایی بالاتری دارد. به‌طور کلی نتایج نشان داد تکنیک‌های خاک‌آماری، به‌خصوص درخت تصمیم، می‌توانند فرآیند نقشه‌برداری خاک‌ها را در گستره‌ای وسیع و متشکل از هر گونه عوارض طبیعی انجام دهد.

نقشه رقومی گروه بزرگ‌های خاک

با توجه به این که مدل درخت تصمیم بهترین صحت در پیش‌بینی گروه‌های بزرگ خاک در منطقه مورد مطالعه را دارا بود، در شکل ۳ نقشه خروجی مدل درخت تصمیم در منطقه مورد مطالعه ارائه شده است. در نقشه پیش‌بینی شده، هشت گروه بزرگ موجود در منطقه مورد مطالعه تشخیص داده شد. آنچه که از شکل ۳ منتج می‌شود این است که، گروه بزرگ هاپلوسالیدز در پدیمت و گروه‌های بزرگ کلسی‌چیسیدز، کلسی‌آرجیدز و پتروکلسیدز را در باهادای بریده بریده و باهادا به‌خوبی پیش‌بینی شدند. هم‌چنین با اطلاعات تصویر ماهواره‌ای

جدول (۳) مقایسه دقت پیش‌بینی گروه‌های بزرگ خاک توسط مدل‌های برازش شده در داده‌های اعتبارسنجی

Table (3) Comparison of predictive quality of soil great groups prediction in calibration data set

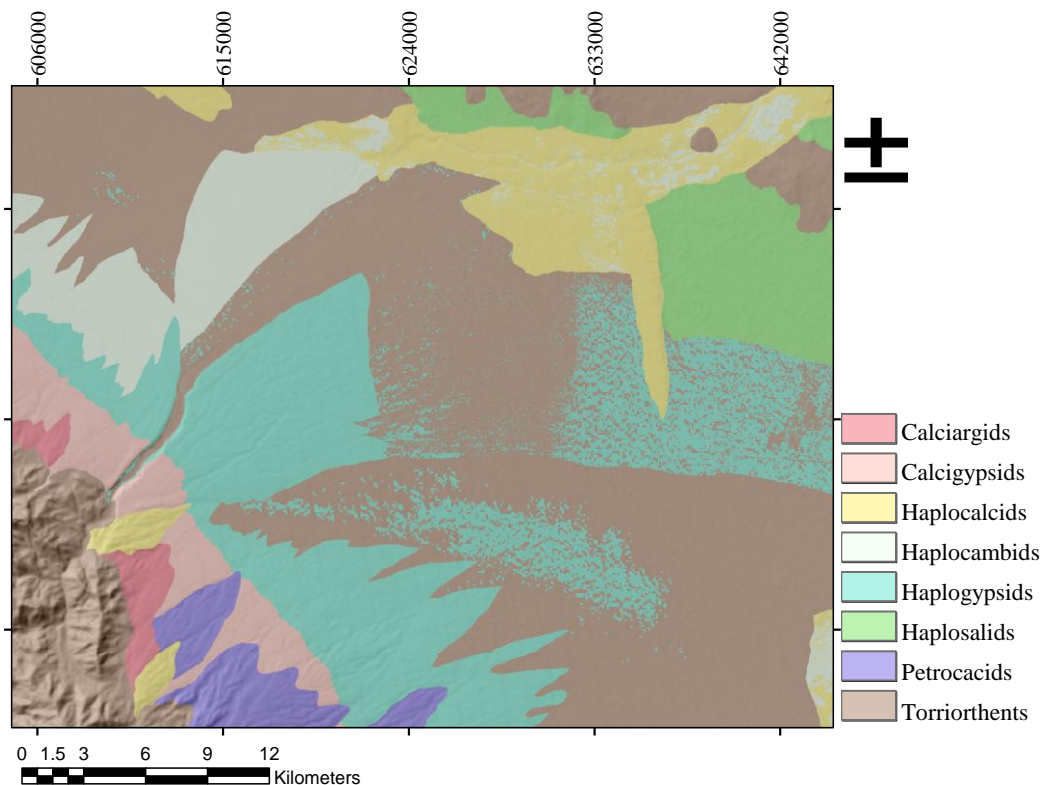
MLR		BRT		DT		گروه‌های بزرگ خاک (Soil great groups)
UA	PR	UA	PR	UA	PR	
1	1	1	1	1	1	هاپلوسالیدز (Haplosalids)
0.83	0.50	1	1	0.80	0.88	هاپلوچیسیدز (Haplogypsids)
0.25	0.25	0.40	0.67	0.67	0.67	هاپلوکلسیدز (Haplocalsids)
0.50	0.40	0.67	0.40	0.75	0.75	هاپلوکمبیدز (Haplocambids)
1	1	1	1	1	1	کلسی‌چیسیدز (Calcigypsids)
1	1	0.50	1	1	1	کلسی‌آرجیدز (Calciargids)
1	1	1	0.50	1	1	پتروکلسیدز (Petrocalcids)
0.61	1	1	1	0.88	0.80	توری‌ارتنتز (Torriorthents)

MLR: رگرسیون لاجستیک چندجمله‌ای، BRT: رگرسیون درختی چندکلاسه، DT: درخت تصمیم طبقه‌بندی، UA: دقت کاربر، PR: قابلیت اطمینان

تولیدکننده

جدول (۴) مقایسه دقت ارزیابی مدل‌های برازش‌شده در منطقه مورد مطالعه
 Table (4) Comparison of prediction quality of models in the study area

درجه برای (Brier score)	شاخص کاپا (Kappa index)	خلوص نقشه (Map Purity)	مدل‌های یادگیری ماشین (Machine learning models)
0.57	0.65	0.71	رگرسیون لاجیستیک چندجمله‌ای (Multinomial regression logistic)
0.48	0.81	0.85	رگرسیون درختی توسعه‌یافته (Boosted regression trees)
0.47	0.84	0.86	درخت تصمیم طبقه‌بندی (Decision tree)



شکل (۳) توزیع مکانی گروه‌های بزرگ خاک با استفاده از مدل درخت تصمیم

Figure (3) Spatial distribution of the soil great groups derived from decision tree model

از دور در پیش‌بینی کلاس‌های خاک نشان داد که این فاکتورها برآیند و نتیجه نیروهای مؤثر خاکسازي در منطقه مورد مطالعه هستند. در صورتی که همه فاکتورهای خاکسازي به‌طور هم‌زمان در مدل‌سازی استفاده شوند، بهترین پیش‌بینی‌ها حاصل می‌شود. هم‌چنین وجود ارتباط قوی بین داده خاک و پارامترهای محیطی نیز، یکی از فاکتورهای تاثیرگذار بر روی

نتیجه‌گیری

نتایج حاصل از تکنیک‌های نقشه‌برداری رقومی خاک در منطقه مورد مطالعه نشان‌دهنده این بود که سطح ژئومرفیک تأثیر معنی‌دار و مهم در پیش‌بینی گروه‌های بزرگ خاک را دارد و به افزایش دقت پیش‌بینی مدل‌ها کمک می‌کند. فرآیندهای ژئومرفیک به‌همراه پارامترهای سرزمین و شاخص‌های سنجش

عباس زاده افشار و همکاران: پیش‌بینی مکانی گروه بزرگ‌های خاک...

خاکشناسان باشد. این اطلاعات هم‌چنین می‌تواند برای تکمیل مجموعه داده‌های موجود در کشور نیز، استفاده شوند. هم‌چنین این مطالعه مقایسه‌ای از تکنیک‌های یادگیری ماشین برای اهداف طبقه‌بندی در علوم خاک را فراهم می‌نماید و ممکن است به انتخاب مدل مناسب برای نقشه‌برداری رقومی خاک و پژوهش‌های مدل‌سازی ژئومورفولوژی در آینده کمک شایانی کند.

دقت مدل خواهد بود. به‌طورکلی نتایج نشان داد که تکنیک‌های یادگیری ماشین می‌توانند فرآیند نقشه‌برداری خاک‌ها را در سطح وسیع انجام داده، روش سنتی را ارتقاء بخشند، و قابلیت استفاده آن‌ها را برای قشر وسیعی از شاخه‌های علمی امکان‌پذیر نماید. هم‌چنین نقشه‌برداری رقومی خاک، می‌تواند به عنوان یک روش ارزیابی منابع خاک استفاده شود. علاوه بر این، قابلیت اطمینان نقشه‌های برآوردشده می‌تواند شروع یک بحث جدید بین متخصصان منابع زمین و

منابع

1. Aksoy, E., Ozsoy, G., and Sabri Dirim, M. 2009. Soil mapping approach in GIS using Landsat satellite imagery and DEM data. *African Journal Agriculture Research*, 4(11): 1295-1302.
2. Banaie, M.H. 1998. Iranian soil moisture and temperature regime map. Agricultural Research, Education and Extension Organization, Soil and Water Research Institute of Iran.
3. Breiman L., Friedman, H.J., Olshen, A.R., and Stone, J.C. 1984. Classification and regression trees. Wadsworth Publishing Company. California, U.S.A.
4. Brungard, C.W., Boettinger, J.L., Duniway, M.C., Wills, S.A., and Edwards Jr, T.C. 2015. Machine learning for predicting soil classes in three semi-arid landscapes. *Geoderma*, 239–240: 68–83.
5. Elith, J., Leathwick, J.R., and Hastie, T. 2008. A working guide to boosted regression trees. *Journal Animal Ecology*, 77: 802–813.
6. Esfandiarpour Borujeni, I., Mohammadi, J., Salehi, M.H., Toomanian, N. and Poch, R. 2010. Assessing geopedological soil mapping approach by statistical and geostatistical methods: A case study in the Borujen region, Central Iran. *Catena*, 82: 1-14.
7. Grunwald, S. 2005. Environmental Soil-Landscape Modeling, Geomorphic Information Technologies and Pedometrics. Taylor and Francis.
8. Hengl, T., Heuvelink, G., and Stein, A. 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma*, 120(1–2): 75–93.
9. Hengl, T., Toomanian, N., Reuter, H.I., and Malakouti, M.J. 2007. Methods to interpolate soil categorical variables from profile observations: lessons from Iran. *Geoderma*, 140: 417–427.
10. Jafari. A., Finke P.A, Van deWauw, J., Ayoubi, S., and Khademi, H. 2012. Spatial prediction of USDA- great soil groups in the arid Zarand region, Iran: comparing logistic regression approaches to predict diagnostic horizons and soil types. *European Journal Soil Science*, 63: 284–298.

11. Jenny, H. 1941. Factors of soil formation. A System of Quantitative Pedology. McGraw-Hill, New York.
12. Kempen, B., Brus, D.J., Heuvelink, G.B.M., and Stoorvogel, J.J. 2009. Updating the 1:50,000 Dutch soil map using legacy soil data: A multinomial logistic regression approach. *Geoderma*, 151: 311–326.
13. Kim, J., Grunwald, S., Rivero, R.G., and Robbins, R. 2012. Multi-scale modeling of soil series using remote sensing in a wetland ecosystem. *Soil Science Society American Journal*, 76: 2327–2341.
14. Lacoste, M., Lemercier, B., and Walter, C. 2011. Regional mapping of soil parent material by machine learning based on point data. *Geomorphology*, 133: 90–99.
15. McBratney, A.B., Santos, M.L.M., and Minasny, B. 2003. On digital soil mapping. *Geoderma*, 117(1-2): 3-52.
16. Meteorological Organization of Iran. 2011. Meteorological statistics of Bam city, Kerman province. <http://www.irimo.ir/>.
17. 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: 97–106.
18. Schoeneberger, P. J., Wysocki, D.A., Benham, E.C., and Broderick, W.D. 2003. Field book for describing and sampling soils, Version 2.0. Natural Resources Conservation Service. National Soil Survey Center, Lincoln.
19. Scull, P., Franklin, J., and Chadwick, O.A. 2005. The application of classification tree analysis to soil type prediction in a desert landscape. *Ecological Modelling*, 181: 1–15.
20. Smith, C. A. S., Daneshfar, B., and Frank, G. 2012. Use of weights of evidence statistics to define inference rules to disaggregate soil survey maps. In Minasny, B., Malone, B.P., and McBratney, A. (Eds.), *Digital Soil Assessments and Beyond: Proceedings of the 5th Global Workshop on Digital Soil Mapping*. CRC Press, Sydney. pp: 215–220.
21. Soil Survey Staff. 2010. *Keys to Soil Taxonomy*, 11th, NRCS, USDA.
22. Sun, X.L., Yu-Guo, Z., Gan-Lin, Z., Sheng-Chun, W., Yu-Bon M. and Ming-Hung, W. 2011. Application of a digital soil mapping method in producing soil orders on mountain areas of hong kong based on legacy soil data. *pedosphere*, 21: 339–350.
23. Taghizadeh-Mehrjardi, R., Minasny, B., McBratney, A.B., Triantafyllidis, J., Sarmadian, F., and Toomanian, N. 2012. Digital soil mapping of soil classes using decision trees in central Iran. In Minasny, B., and McBratney M. (Eds.), *Digital Soil Assessments*. Taylor and Francis Group, London.

24. Taghizadeh-Mehrjardi, R., Nabiollahi, K., Minasny, B., and Triantafilis, J. 2015. Comparing data mining classifiers to predict spatial distribution of USDA-family soil groups in Baneh region, Iran. *Geoderma*, 253–254: 67–77.
25. Taghizadeh-Mehrjardi, R., Sarmadian, F., Omid, M., Savabeghi, Gh., Toomanian, N., Rosta, M.J., and Rahimian, M.H. 2013. comparison of artificial neural network and decision tree methods for mapping soil units in ardakan region. *Iranain Journal of Soil and Water Research*, 44(2): 173-182.
26. Tavares Wahren, F., Julich, S., Nunes, J.P., Gonzalez-Pelayo, O., Hawtree, D., Feger, K.H., and Jacob Keizer, J. 2016. Combining digital soil mapping and hydrological modeling in a data scarce watershed in north-central Portugal. *Geoderma*, 264: 350–362.
27. The Ministry of Economy, Trade and Industry of Japan (METI) and the National Aeronautics and Space Administration (NASA). 2012. Aster Global Digital Elevation Model (Aster GDEM). NASA Official. URL <http://www.gdem.aster.ersdac.or.jp>.
28. Triantifilis, J., Earl, N.Y., and Gibbs, I.D. 2012. Digital soil-classmapping across the Edgeroi district using numerical clustering and gamma-ray spectrometry data. In Minasny, B., Malone, B.P., and McBratney, A. (Eds.), *Digital Soil Assessments and Beyond: Proceedings of the 5th Global Workshop on Digital Soil Mapping*. CRC Press, Sydney. pp: 187–191.
29. Zhu, A., Hudson, B., Burt, J., Lubich, K., and Simonson, D. 2001. Soil mapping using GIS, expert knowledge, and fuzzy logic. *Soil Science Society American Journal*, 65: 1463-1472.