



مدل‌سازی پراکنش مکانی کلاس‌های خاک با استفاده از الگوریتم‌های یادگیری ماشین

در بخشی از اراضی استان زنجان

مستانه رحیمی مشکله، محمد امیر دلاور*، محمد جمشیدی و امین شریفی فر

دانشجوی دکتری گروه علوم خاک دانشکده کشاورزی، دانشگاه زنجان، ایران؛ mastanehrahimi@znu.ac.ir

دانشیار گروه علوم خاک دانشکده کشاورزی، دانشگاه زنجان، ایران؛ amir-delavar@znu.ac.ir

استادیار موسسه تحقیقات خاک و آب، سازمان تحقیقات، آموزش و ترویج کشاورزی، کرج، ایران؛ mohammadjamshidi@yahoo.com

پژوهشگر گروه علوم خاک پردیس کشاورزی، دانشگاه تهران، ایران؛ sharififar@ut.ac.ir

« مقاله پژوهشی »

دریافت: ۱۴۰۱/۱۲/۱۵ و پذیرش: ۱۴۰۲/۴/۱۴

چکیده

نقشه‌برداری رقومی از روش‌های نوینی است که از الگوریتم‌های یادگیری ماشین و متغیرهای محیطی پیروی کرده و به دلیل صرفه‌جویی در زمان و هزینه انجام مطالعه برای پهنه‌بندی خصوصیات و کلاس‌های خاک در سطوح مختلف طبقه‌بندی کاربرد گسترده‌ای در سطح جهانی دارد. این پژوهش باهدف مقایسه کارایی مدل‌های یادگیری ماشین در پیش‌بینی مکانی زیرگروه‌های خاک در بخشی از اراضی استان زنجان انجام شد. برای این منظور، بر اساس الگوی طبقه‌بندی تصادفی با میانگین فاصله ۵۰۰ متر، ۱۴۸ خاک‌رخ حفر و تشریح گردید و مطابق با سامانه جامع رده‌بندی خاک به روش آمریکایی طبقه‌بندی شد. خاک‌ها در سطح زیرگروه در پنج کلاس تیبیک کلسی‌زریپتز، تیبیک هاپلوزریپتز، جیپسیک هاپلوزریپتز، تیبیک زراورتننز و لیتیک زراورتننز طبقه‌بندی شد. متغیرهای محیطی شامل نقشه‌های ژئومورفولوژی، زمین‌شناسی، توپوگرافی و داده‌های حاصل از تصاویر سنجش‌ازدور بود. ۵۷ متغیر محیطی به‌عنوان نمایندگان عوامل خاک‌سازی استخراج گردید و با استفاده از روش تحلیل مؤلفه اصلی و نظر کارشناسان، مؤثرترین متغیرهای محیطی انتخاب شد. مدل‌سازی خاک - زمین‌نما با استفاده از الگوریتم‌های جنگل تصادفی، درخت تصمیم توسعه‌یافته و رگرسیون لجستیک چندجمله‌ای در محیط نرم‌افزار Rstudio انجام شد. صحت کلی و ضریب کاپا برای ارزیابی کلاس‌های خاک در سطح زیرگروه به ترتیب در مدل رگرسیون لجستیک چندجمله‌ای ۶۵٪ و ۴۱٪، در مدل جنگل تصادفی ۶۵٪ و ۳۲٪ و در مدل درخت تصمیم توسعه‌یافته ۶۰٪ و ۳۵٪ به دست آمد. صحت کاربر و صحت تولیدکننده نشان داد که مدل رگرسیون لجستیک چندجمله‌ای برآورد قابل قبولی در پیش‌بینی مکانی زیرگروه‌های خاک ارائه می‌دهد. متغیرهای عمق دره، فاصله تا شبکه آبراهه، شاخص همواری بالای پشته با درجه تفکیک بالا و شاخص طول در جهت شیب دارای بیش‌ترین اهمیت در مدل رگرسیون لجستیک چندجمله‌ای بود. کلاس‌های خاک با فراوانی بیشتر صحت بالاتری داشت. که نشان می‌دهد رابطه مستقیمی بین فراوانی کلاس‌های خاک در داده‌های آموزشی با صحت نتایج پیش‌بینی کلاس‌های خاک برقرار است.

واژه‌های کلیدی: رگرسیون لجستیک چندجمله‌ای، ضریب کاپا، نقشه برداری رقومی خاک، ویژگی‌های توپوگرافی

* - آدرس ایمیل نویسنده مسئول: amir-delavar@znu.ac.ir



در این رابطه S_a : ویژگی‌های قابل پیش‌بینی خاک، S_c : کلاس خاک موردنظر، C : اقلیم، O : پوشش گیاهی یا فعالیت‌های جانوری و انسانی، T : توپوگرافی، P : مواد مادری خاک، a : سن و یا زمان سپری‌شده، n : موقعیت مکانی نمونه‌ها و پیش‌بینی‌ها و e : مقدار خطای پیش‌بینی است که در مدل اسکورپن، تابع مختصات و موقعیت مکانی آن‌ها است. در این رابطه، ویژگی‌ها یا کلاس‌های خاک با متغیرهای محیطی^۳ ارتباط داده می‌شوند. به این صورت که در هر منطقه، متغیرهای محیطی وجود دارند که می‌توانند با ویژگی‌ها یا کلاس‌های خاک در محل همبستگی داشته باشند. یافتن این ارتباط و پیش‌بینی ویژگی‌ها یا کلاس‌های خاک در نقاط دیگر، سبب تهیه نقشه رقومی خاک در تمام منطقه مورد مطالعه می‌گردد (تقی‌زاده مهرجردی و همکاران، ۲۰۱۴)؛ به واسطه نیاز تهیه نقشه‌های خاک و ارائه اطلاعات به‌روز و قابل‌اعتمادتر درباره خاک‌ها همراه با ارتقای تفسیر نتایج، به‌نحوی که افراد غیرمتخصص نیز بتوانند از این اطلاعات استفاده کنند، استفاده از روش‌های نوین در تهیه نقشه خاک ضروری است. در این راستا روش نقشه‌برداری رقومی خاک^۴ با هدف تهیه نقشه‌های دقیق بر مبنای مدل‌سازی کمی روابط بین کلاس‌ها یا ویژگی‌های خاک و داده‌های محیطی نماینده عوامل خاک‌سازی به‌عنوان یک روش جایگزین برای تهیه نقشه‌های خاک به روش سنتی معرفی شده است (هارتمینک و همکاران، ۲۰۰۸؛ کمپن و همکاران، ۲۰۰۹).

کلید اصلی نقشه‌برداری رقومی خاک، استفاده مؤثر از متغیرهای کمکی برای درک روابط خاک - زمین - نما است (فن و همکاران، ۲۰۲۲). انتخاب مجموعه‌ای از متغیرهای کمکی که با تجزیه و تحلیل ویژگی‌های محیطی یک منطقه روابط خاک - زمین نما را مشخص کنند، اهمیت بسیاری دارد (ژانگ و همکاران، ۲۰۱۷، لیانگ و همکاران، ۲۰۲۰). محققان مختلف از مدل‌های رقومی

خاک به‌عنوان یک محیط چندمنظوره برای رشد گیاهان یک جزء کلیدی برای دستیابی به اهداف توسعه پایدار شناخته می‌شود (ایون و همکاران، ۲۰۲۱). امروزه استفاده بهینه از اراضی به‌عنوان یک چالش بزرگ مدیریتی در دنیا مطرح است (گارگ و همکاران، ۲۰۲۰)، بنابراین تصمیم‌گیرندگان به اطلاعات مکانی دقیق خاک نیاز دارند تا اطمینان حاصل کنند خاک و اراضی با روشی کارآمد و پایدار مورد استفاده، مدیریت و حفاظت قرار می‌گیرند (هلفنستاین و همکاران، ۲۰۲۲). نقشه‌های خاک از مهم‌ترین منابع اطلاعات شناسایی خاک در کشور هستند و اغلب با روش‌های سنتی (مرسوم) تهیه شده و یکی از مهم‌ترین محدودیت‌های استفاده از روش‌های نقشه‌برداری مرسوم خاک تأکید بر قابلیت‌های فردی و درک خاک‌شناس از روابط خاک و زمین‌نما است که معمولاً با خطا همراه است (النگر، ۲۰۰۷). محدودیت‌های دیگر شامل بررسی میدانی طولانی است که فرآیند نقشه‌برداری را وقت‌گیر و غیر اقتصادی می‌کند (لاگاجری و مک‌براتنی، ۲۰۰۶) و نقشه‌های تهیه شده اغلب توسط کاربران نهایی، غیرقابل دسترس و بدون استفاده باقی می‌مانند (سانترا و همکاران، ۲۰۲۱). به‌منظور رفع محدودیت‌های ذکر شده، مک‌براتنی و همکاران (۲۰۰۳) با مرور مدل‌های مختلف تخمین و برآورد مکانی خاک و توجه به علوم ریاضی و آمار برای انتخاب متغیرهای محیطی مناسب در تهیه نقشه خاک‌ها به ارائه یک چارچوب تعمیم‌یافته تحت عنوان مدل اسکورپن^۱ (رابطه ۱) اقدام کردند؛ آن‌ها با دخیل نمودن موقعیت مکانی و دیگر مشخصه‌های خاک، اقدام به بسط و توسعه مدل چندعاملی ینی^۲، به‌منظور برآورد و تخمین یک ویژگی یا کلاس خاک نمودند و انقلابی گسترده در نقشه‌برداری خاک در سراسر جهان در دو دهه اخیر به‌وجود آوردند.

$$S_a \text{ or } S_c = f(s, c, o, r, p, a, n) + e \quad (1)$$

3. Environmental covariates
4. Digital soil mapping, DSM

1. Scorpan
2. Jeny

بزرگ حاصل از داده‌های ماهواره‌ای یا پارامترهای مشتق شده از مدل رقومی ارتفاع است (پادریان و همکاران، ۲۰۲۰). این داده‌ها به عنوان کنترل‌کننده‌ها یا پیش‌بینی‌کننده‌های محیطی که نماینده عوامل خاک‌سازی هستند، نقش دارند (هنگل و همکاران، ۲۰۱۸). الگوریتم‌های مختلفی در نقشه‌برداری رقومی خاک استفاده می‌شود که از جمله این موارد می‌توان به جنگل تصادفی^۲، درخت تصمیم توسعه‌یافته^۳ و رگرسیون لجستیک چندجمله‌ای^۴ اشاره کرد. پژوهش‌های متعددی توسط محققان در سال‌های اخیر بر روی مدل‌سازی و پیش‌بینی مکانی کلاس‌های خاک انجام گرفته است (کمپن و همکاران، ۲۰۰۹؛ سای‌لین و همکاران ۲۰۱۱؛ تقی‌زاده مهرجردی و همکاران، ۲۰۱۵؛ پلگرینو و همکاران، ۲۰۱۶؛ مصلح و همکاران، ۱۳۹۶).

با وجود گذشت بیش از شش دهه از انجام مطالعات خاک‌شناسی در ایران هنوز نقشه‌های خاک در کشور به‌طور کامل تهیه نشده و از این رو تهیه نقشه‌های خاک باکیفیت و دقت قابل قبول با استفاده از روش‌های نوین نقشه‌برداری می‌تواند برای اراضی فاقد نقشه موردتوجه قرار گیرد. از طرفی مطالعات نقشه‌برداری رقومی خاک در استفاده از متغیرهای محیطی بیشتر معطوف به استفاده از فاکتورهای رقومی ارتفاع و سنجش‌ازدور بوده و استفاده از اطلاعات ژئومورفولوژی و زمین‌شناسی به‌عنوان متغیرهای محیطی که به‌طور مستقیم می‌توانند بر پراکنش انواع خاک‌ها اثرگذار باشند کمتر استفاده شده است؛ بنابراین هدف از انجام این مطالعه تلفیق اطلاعات نقشه‌های ژئومورفولوژی و زمین‌شناسی همراه با سایر متغیرهای محیطی برای تهیه نقشه‌های رقومی خاک با استفاده از مدل‌های یادگیری ماشین جنگل تصادفی، درخت تصمیم توسعه‌یافته و رگرسیون لجستیک چندجمله‌ای است.

مواد و روش‌ها

ارتفاع، داده‌های طیفی سنجش‌ازدور و نقشه‌های حاصل از مطالعات موروثی به‌عنوان متغیرهای محیطی استفاده کرده‌اند. تقی‌زاده مهرجردی و همکاران (۲۰۱۴) برای نقشه‌برداری رقومی کلاس‌های خاک در منطقه اردکان استان یزد از ویژگی‌های توپوگرافی و نقشه‌های ژئومورفولوژی، زمین‌شناسی و کاربری اراضی، موسوی و همکاران (۱۳۹۸) در مدل‌سازی و پیش‌بینی مکانی کلاس خاک در بخشی از اراضی دشت قزوین با اطلاعات مدل رقومی ارتفاع و شاخص‌های سنجش‌ازدور، مقصودی و همکاران (۱۳۹۹) برای پیش‌بینی مکانی واحدهای خاک در اراضی استان ایلام تنها از داده‌های حاصل از مدل رقومی ارتفاع به‌عنوان متغیرهای محیطی و عباس‌زاده افشار و همکاران (۱۳۹۷) از ویژگی‌های پستی و بلندی همراه با فرآیندهای ژئومورفیک در پیش‌بینی گروه بزرگ‌های خاک در بخشی از اراضی جنوب شرق ایران استفاده کردند. ملاح و همکاران (۲۰۲۲) در پیش‌بینی کلاس‌های بافتی خاک در مزارع شمال ایران از تصاویر سنجش‌از راه دور، مدل رقومی ارتفاع و داده‌های زیست اقلیم به‌عنوان متغیر محیطی استفاده کردند.

کاربرد روش‌های یادگیری ماشین در زمینه‌های مختلف علوم به‌ویژه در دهه گذشته به‌سرعت افزایش یافته است (میناسنی و مک‌براتی، ۲۰۱۶). دسترسی روزافزون به داده‌های خاک که می‌توانند به‌طور مؤثری از روش‌های دورسنجی و مطالعات سنجش‌از نزدیک^۱ به دست آیند و الگوریتم‌هایی که آزادانه در دسترس هستند، منجر به پذیرش سریع روش‌های یادگیری ماشین برای تجزیه و تحلیل داده‌های خاک شده است (پادریان و همکاران، ۲۰۲۰). تحقیقات علوم خاک، به‌ویژه، در زمینه پدومتریک، از مدل‌های آماری برای یادگیری یا درک داده‌ها استفاده کرده و چگونگی پراکنش خاک در مکان و زمان را بیان می‌کنند (مک‌براتی و همکاران، ۲۰۱۹). یکی از کاربردهای مهم روش‌های یادگیری ماشین عمدتاً شناسایی و پیش‌بینی الگوهای موجود در چندین مجموعه داده

2. Random Forest, RF
3. Boosted Decision Tree, BRT
4. Multinomial Logistic Regression, MNLR

1. Proximal Sensing

خشک شدن از الگ ۲ میلی‌متری عبور داده شد و تجزیه‌های فیزیکی و شیمیایی شامل بافت خاک (بویوکاس، ۱۹۶۲)، واکنش خاک (پری و لاوتن‌شلاگر، ۱۹۸۴)، کربنات کلسیم معادل خاک (لانیون و هیلد، ۱۹۸۳)، ظرفیت تبادل کاتیونی (سامنر و میلر، ۱۹۹۶)، قابلیت هدایت الکتریکی (ریچاردز، ۱۹۵۴)، کربن آلی (والکی و بلک، ۱۹۳۴) بر روی تمام نمونه‌ها و در صورت مشاهده صحرایی بر روی برخی نمونه‌های دارای گچ (آرتیدا و همکاران، ۲۰۰۶) انجام شد. خاک‌رخ‌ها بر اساس نتایج تشریح و اطلاعات تجزیه‌های فیزیکی و شیمیایی نمونه‌های خاک در سیستم جامع رده‌بندی خاک به روش آمریکایی (هیئت شناسایی خاک^۱، ۲۰۱۴) تا سطح فامیل طبقه‌بندی شدند.

متغیرهای محیطی

متغیرهای محیطی مورد استفاده شامل اطلاعات نقشه‌های ژئومورفولوژی، نقشه زمین‌شناسی، داده‌های سنجش‌ازدور و توپوگرافی هستند. برای این منظور نقشه زمین‌شناسی با مقیاس ۱:۲۵۰۰۰۰ منطقه تهیه‌شده توسط سازمان زمین‌شناسی کشور در محیط Arc-GIS نسخه ۱۰٫۷ زمین مرجع و رقومی شد (شکل ۲، الف). از مدل رقومی ارتفاع با قدرت تفکیک مکانی ۳۰×۳۰ متر سنجنده استر، ۱۸ شاخص پستی‌وبلندی در محیط نرم‌افزار SAGA GIS (نسخه ۷٫۹) استخراج شد. ۳۶ شاخص سنجش‌ازدور از تصاویر سنجنده (OLI/TIRS) ماهواره لندست ۸ با قدرت تفکیک مکانی ۳۰×۳۰ متر (USGS 2014) پس از اعمال تصحیحات رادیومتریک و اتمسفری در محیط نرم‌افزار ENVI (نسخه 5.3) تهیه و استخراج شد. نقشه ژئومورفولوژی منطقه بر اساس تلفیق لایه‌های اطلاعاتی شامل واحدهای لندفرم و مواد مادری به همراه تفسیر تصاویر ماهواره‌ای بر اساس رویکرد سلسله‌مراتبی ارائه شده توسط زینک (زینک، ۲۰۱۵) تهیه گردید (جدول ۱، شکل ۲-ب). از میان متغیرهای محیطی

منطقه مورد مطالعه به مساحت ۱۳۸۲۳ هکتار از توابع شهرستان ایچرود استان زنجان است. این منطقه در جنوب غربی استان زنجان با مختصات جغرافیایی ۴۷ درجه و ۹۱ دقیقه تا ۴۸ درجه و ۱۱ دقیقه طول شرقی و ۳۶ درجه و ۳۷ دقیقه تا ۳۶ درجه و ۳۱ دقیقه عرض شمالی واقع شده است (شکل ۱). متوسط بارندگی منطقه ۳۴۰ میلی‌متر، متوسط دمای سالانه ۱۴ درجه سلسیوس و میانگین ارتفاع منطقه ۱۴۸۲ متر از سطح دریا است (سالنامه آماری زنجان، ۱۳۹۸). خاک‌های منطقه دارای رژیم حرارتی مزیک^۱ و رژیم رطوبتی زیریک^۲ هستند (موسسه تحقیقات خاک و آب، ۱۳۸۹). سازندهای زمین-شناسی عمده منطقه شامل لایه‌های کربناته، سنگ آهک، کنگلومرا و مواد آتشفشانی و فیزیوگرافی منطقه شامل دو واحد اراضی تپه‌ماهور^۳ و دشت‌های دامنه‌ای^۴ است. اراضی تپه‌ماهوری در سطح پستی‌وبلندی به تپه‌های با ارتفاع متوسط^۵ و تپه‌های با ارتفاع کم^۶ تفکیک شده‌اند. با توجه به تقسیم‌بندی زمین‌شناسی، تپه‌های با ارتفاع متوسط غالباً رسوبات واریزه‌ای - بادرفتی آهکی همراه با مارن^۷ و پستی‌وبلندی تپه‌های با ارتفاع کم غالباً شامل مخلوط رسوبات واریزه‌ای - بادرفتی و مارن^۸ هستند.

مطالعات صحرایی و آزمایشگاهی

پس از بررسی نقشه‌های توپوگرافی و استفاده از اطلاعات موروثی در محیط سامانه‌های اطلاعات جغرافیایی، تصاویر برگرفته از گوگل ارث و انجام بازدیدهای صحرایی، ۱۴۸ خاک‌رخ بر اساس الگوی طبقه‌بندی تصادفی با میانگین فاصله ۵۰۰ متر مطابق روش‌های استاندارد مطالعه (راهنمای مطالعات خاک^۹، ۲۰۱۷) و نمونه‌برداری از افق‌های مشخصه سطحی و زیرسطحی هر خاک‌رخ انجام شد. نمونه‌ها پس از هوا

1. Mesic
2. Xeric
3. Hill lands
4. Piedmont plains
5. Medium hill
6. Low hill
7. Marl+lime+alluvio-colluvium
8. Marl+Alluvio-colluvium
9. Soil survey manual

تهیه شده، تعدادی متغیر محیطی بر اساس رویکرد تحلیل مؤلفه اصلی^۱ در نرم‌افزار SPSS نسخه ۲۶,۰,۰ و رتبه-بندی اهمیت نسبی مدل یادگیری ماشین به همراه نظارت کارشناس برای ورود به مدل انتخاب گردید (کوهن و جانسون، ۲۰۱۳).

مدل‌سازی مکانی

بعد از آماده‌سازی داده‌ها و متغیرهای محیطی مؤثر در مدل‌سازی به‌عنوان متغیرهای وابسته و داده‌های مربوط به کلاس‌های خاک به پیش‌بینی کلاس‌های خاک با استفاده از الگوریتم جنگل تصادفی، درخت تصمیم توسعه یافته و رگرسیون لجستیک چندجمله‌ای در محیط نرم‌افزار RStudio (نسخه ۳,۴۹۲) اقدام شد.

جنگل تصادفی

مدل جنگل تصادفی یک تکنیک یادگیرنده فعال و توسعه یافته از مدل طبقه‌بندی و رگرسیون درختی است. در این روش داده‌ها به‌طور تکراری برای به دست آوردن ارتباط بین متغیر پاسخ و متغیرهای مستقل و انجام تخمین جداسازی می‌شوند. در روش جنگل تصادفی برخلاف سایر روش‌های درختی که تعداد محدودی درخت ترسیم می‌کنند، صدها یا هزاران درخت طبقه‌بندی تولید می‌شود (بريمن و کالتر، ۲۰۰۴). این روش یک روش یادگیری گروهی است و برای طبقه‌بندی با ساختن تعداد درختان زیاد عمل می‌نماید (بريمن، ۲۰۰۱). کلیه مراحل مدل-سازی با استفاده از روش یادگیری جنگل تصادفی با استفاده از بسته Random Forest به همراه کدنویسی در محیط نرم‌افزار RStudio انجام شد.

درخت تصمیم توسعه یافته

رگرسیون درختی توسعه یافته به‌عنوان یکی از الگوریتم‌های یادگیری ماشین ترکیبی از دو تکنیک آماری

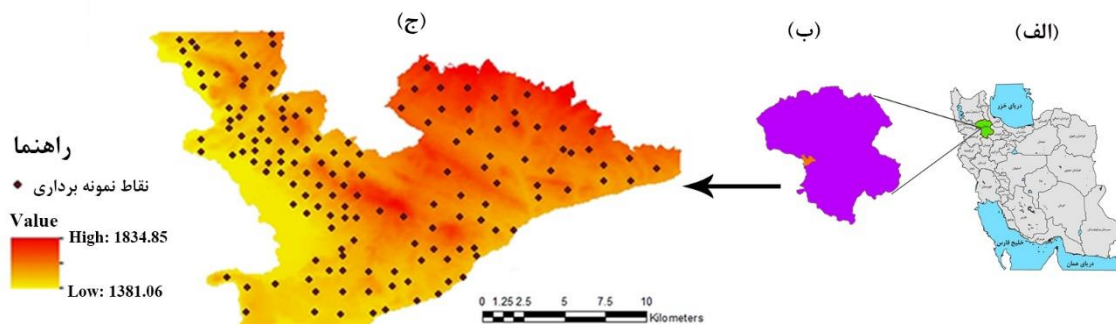
بوستینگ^۲ و رگرسیون درختی است (آبر، ۲۰۰۹). بوستینگ یک روش مرحله‌ای روبه‌جلو است که در آن مدل‌های درختی به‌صورت تکرارپذیر با زیرمجموعه‌ای از داده‌های آموزشی برآزش داده می‌شوند. در برآزش رگرسیون درختی توسعه یافته باید دو پارامتر نرخ یادگیری^۳ و پیچیدگی درخت^۴ مشخص گردند. نرخ یا مقدار یادگیری سهم هر درخت متوالی را در مدل نهایی تعیین می‌کند. پیچیدگی درخت اثرات اصلی یا اثرات متقابل بین متغیرها را نشان می‌دهد (الیت و همکاران، ۲۰۰۸). مدل‌سازی رگرسیون درختی توسعه یافته در نرم‌افزار Rstudio و با بسته C5.0 انجام شد.

رگرسیون لجستیک چندجمله‌ای

مدل رگرسیون لجستیک یک نوع مدل خطی تعمیم یافته است و برای مجموعه داده‌هایی مناسب است که متغیر وابسته به‌صورت کیفی است. این مدل‌ها قادر به توصیف روابط بین مجموعه‌ای از متغیرهای پیش‌بینی کننده و یک متغیر وابسته دویخی است که دارای مقادیر صفر یا چهار است (کلاین باوم و همکاران، ۲۰۰۸). در هر دو حالت مدل رگرسیون لجستیک چندجمله‌ای برای هر کلاس خاک در منطقه مورد مطالعه توسعه و روابط توپوگرافی و واحدهای طبقه‌بندی خاک از داده‌ها تعیین شدند. مدل‌سازی رگرسیون لجستیک در نرم‌افزار Rstudio بر اساس بسته "Caret" انجام شد.

2. Boosting
3. Learning rate
4. Tree complexity

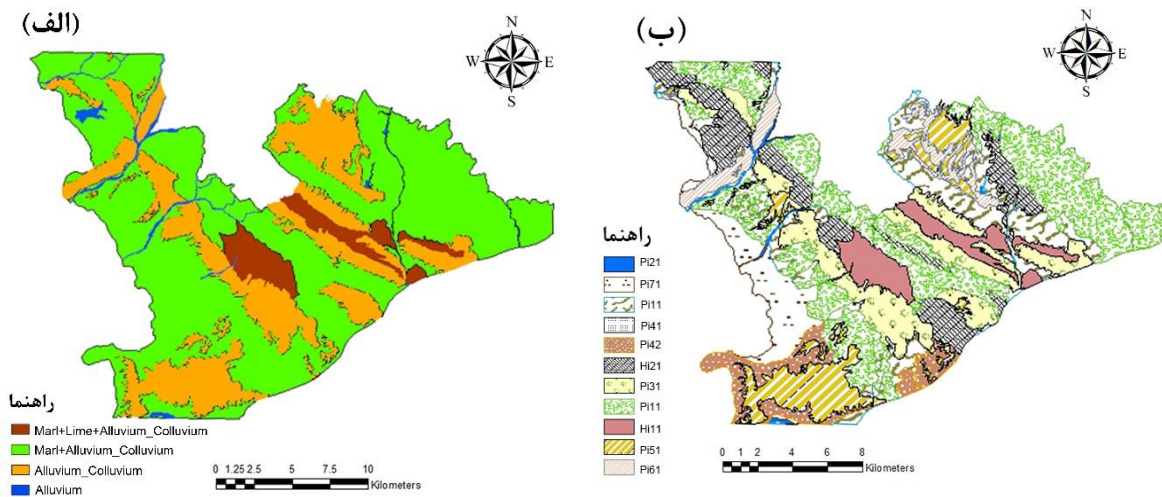
1. Principal Component Analysis, PCA



شکل ۱- موقعیت منطقه و نقاط نمونه‌برداری (الف) ایران، (ب) زنجان و (ج) مدل رقومی ارتفاع منطقه

جدول ۱_ واحدهای تفکیک‌شده در سطح لندفرم بر اساس اطلاعات ژئومورفولوژی در منطقه مورد مطالعه

| واحد نقشه | اجزای لندفرم | شکل زمین | سنگ‌شناسی/منشأ | پستی‌وبلندی/قالب | زمین‌نما |
|--------------|--------------|---------------------|------------------------------------|------------------------------------|--------------|
| Consociation | Hi111 | Slope facet complex | Marl + Lime + Alluvium - Colluvium | Medium hill | تپه‌ماهور |
| | Hi211 | Summit | | | |
| | Hi212 | Shoulder | | | |
| Complex | Hi213 | Backslope | Marl + Alluvium - Colluvium | Low hill | |
| | Hi214 | Footslope | | | |
| | Hi215 | Toeslope | | | |
| Association | Pi111 | High glacial | Alluvium - Colluvium | Glacial, Dissected | |
| Complex | Pi211 | Middle glacial | Marl + Alluvium - Colluvium | Glacial, Moderately dissected | |
| Association | Pi311 | Low glacial | Marl + Alluvium - Colluvium | Glacial, Low dissected | |
| Association | Pi411 | Side slope | Alluvium | Glacial terrace, Dissected | دشت دامنه‌ای |
| Consociation | Pi421 | Tread | Alluvium - Colluvium | | |
| Association | Pi511 | Side slope | Marl + Alluvium - Colluvium | Glacial terrace, Slightly eroded | |
| Association | Pi611 | Side slope | Alluvium - Colluvium | Coalescing fan | |
| Association | Pi711 | Upper part | Marl + Alluvium - Colluvium | Channeled recent alluvial deposits | |
| | Pi712 | Middle part | | | |
| | Pi713 | Lower part | | | |



شکل ۲_الف) نقشه زمین‌شناسی (ب) نقشه ژئومورفولوژی منطقه مورد مطالعه

شاخص دارای مقداری بین صفر و یک است. اگر کاپا برابر با صفر باشد نشان‌دهنده طبقه‌بندی کاملاً تصادفی و مقدار منفی نشان‌دهنده خطا در طبقه‌بندی و اگر این مقدار برابر با یک باشد نشان‌دهنده طبقه‌بندی کاملاً صحیح است.

$$Kappa = N \sum_{i=1}^k X_{ii} - \sum_{i=1}^k (X_{i+} \times X_{+i}) / N^2 - \sum_{i=1}^k (X_{i+} \times X_{+i}) \quad (3)$$

در این رابطه X_{i+} و X_{+j} به ترتیب مجموع حاشیه‌ای برای ردیف i و ستون j هستند. مقادیر کاپا بیشتر از $0/8$ نشان‌دهنده توافق یا دقت قوی بین نقشه طبقه‌بندی و اطلاعات مرجع زمینی است. مقادیر بین $0/4$ و $0/8$ نشان‌دهنده توافق متوسط و مقادیر کمتر از $0/4$ نشان‌دهنده توافق ضعیف است (کنگلتون، ۱۹۹۱).

برای ارزیابی هر طبقه خاک، دو شاخص صحت تولیدکننده و صحت کاربر نیز محاسبه شد. دقت تولیدکننده از تقسیم تعداد کل پیکسل‌های صحیح یک کلاس بر تعداد کل پیکسل‌های آن کلاس از داده‌های مرجع زمین (کل ستون) محاسبه می‌شود و از رابطه ۴ به دست می‌آید.

$$PA = \frac{X_{ij}}{X_{+j}} \quad (4)$$

در این رابطه X_{ij} و X_{+j} به ترتیب مجموع حاشیه‌ای برای ردیف i و ستون j هستند.

ارزیابی عملکرد مدل‌ها

اعتبار یک مدل به‌طور ساده بیان درصدی از پیش‌بینی‌های انجام‌شده توسط آن مدل است که با واقعیت موجود هماهنگی دارد. به‌منظور آموزش مدل‌ها مجموعه خاک رخ‌ها (متغیرهای محیطی و کلاس‌های خاک) به صورت تصادفی به دو مجموعه با نسبت چهار به یک تقسیم شدند. ۸۰ درصد داده‌ها برای آموزش مدل و ۲۰ درصد برای اعتبارسنجی استفاده شد. ارزیابی مدل‌ها با استفاده از شاخص‌های صحت کلی نقشه، صحت تولیدکننده، صحت کاربر^۳ و ضریب کاپا^۴ انجام شد (جنسن، ۱۹۹۶). صحت کلی نقشه از تقسیم تعداد کل پیکسل‌های که به درستی پیش‌بینی شده (مجموع قطر در ماتریس خطا) بر تعداد کل پیکسل‌های ماتریس خطا (N) به دست می‌آید (رابطه ۲).

$$OA = \sum_{i=1}^k X_{ii} / N \quad (2)$$

در این رابطه X_{ii} تعداد مشاهده‌ها در ردیف i و ستون i است، k تعداد سطرها (کلاس‌های خاک) در ماتریس خطا است.

شاخص کاپا معیاری برای مقایسه طبقه‌بندی مدل خودکار با طبقه‌بندی تصادفی است (رابطه ۳). این

1. Overall Accuracy, OA
2. Producer Accuracy, PA
3. Users Accuracy, UA
4. Kappa Index

جدول (۳) ارائه شده است. نتایج مطالعه خاک‌رخ‌ها نشان می‌دهد خاک‌رخ شماره (۱) خاکی است با ضخامت زیاد و بافت سبک (شنی لومی) و زهکشی مناسب، فاقد گچ بوده و بیشترین میزان میزان آهک در افق‌های این خاک‌رخ مشاهده می‌گردد که از سطح به عمق با روندی نامنظم کاهش می‌یابد. خاک‌رخ شماره (۲) خاکی عمیق و دارای بافت شنی لومی است این خاک‌ها بر روی مواد مادری مارن گچی تشکیل شده و دارای نفوذپذیری متوسط می‌باشند. میزان گچ در این خاک‌رخ از سطح به عمق به‌طور نامنظم افزایش پیدا کرده است و رنگ آن در افق‌های سطحی روشن‌تر از سایر خاک‌رخ‌هاست. در این خاک رخ بیشترین میزان قابلیت هدایت الکتریکی به چشم می‌خورد و به نظر می‌رسد رابطه مستقیمی بین میزان گچ و شوری وجود دارد. خاک‌رخ شماره (۳) تنها خاک‌رخی است که ضخامت آن کم است و این موضوع به شیب و توپوگرافی زیاد منطقه مربوط است. بیشترین مقدار سیلت، رس و آهک در این خاک‌رخ قابل مشاهده است. بافت آن لومی رسی است که دارای زهکشی و نفوذپذیری متوسط است. خاک‌رخ شماره (۴) خاکی است عمیق که بیشترین مقدار شن در آن مشاهده می‌شود و بافت آن در تمامی افق‌ها شنی لومی است. بیشترین مقدار کربن آلی در این خاک رخ مشاهده می‌شود که از سطح به عمق روند کاهشی دارد و به‌طور کلی مقدار ماده آلی با عمق کاهش می‌یابد. این خاک‌رخ دارای زهکشی و نفوذپذیری مناسب است. در خاک‌رخ شماره (۵) بیشترین میزان سیلت و رس مشاهده می‌شود و بافت آن سیلت رسی است. این خاک-رخ نیز مانند سایر خاک‌رخ‌ها ضخامت زیاد دارد و نفوذپذیری آن متوسط است.

مطابق با نتایج حاصل شده از جدول (۳)، خاک-رخ‌های شاهد شماره (۱) و (۲) در رده اینسپتی‌سول در مراحل آغازین تکامل و بر روی زمین‌نمای تپه‌ماهور مشاهده شدند. این خاک‌ها بر روی مواد مادری آهکی و گچی تشکیل شده و دارای نفوذپذیری مناسبی هستند. وجود افق مشخصه زیرسطحی کلسیک در این دو خاک-

صحت کاربرد از تقسیم تعداد کل پیکسل‌های صحیح یک کلاس بر تعداد کل پیکسل‌هایی که واقعاً در آن دسته طبقه‌بندی شده‌اند (کل ردیف) محاسبه می‌شود و از رابطه ۵ به دست می‌آید.

$$UA = \frac{X_{ii}}{X_{+i}} \quad (5)$$

در این رابطه X_{+i} و X_{ii} به ترتیب مجموع حاشیه‌ای برای ردیف i و ستون j هستند. دامنه تغییرات صحت تولیدکننده و صحت کاربرد بین صفر و یک است و مقادیر بالاتر نشان‌دهنده عملکرد مناسب‌تر مدل است.

نتایج و بحث

تشریح کلی خاک‌های منطقه

نتایج طبقه‌بندی نشان داد که خاک‌های منطقه در دو رده انتی‌سولز^۱ و اینسپتی‌سولز^۲ قرار می‌گیرند. این خاک‌ها در سطح زیرگروه در پنج کلاس تیبیک کلسی-زریپتز^۳، تیبیک هاپلوزریپتز^۴، جیپسیک هاپلوزریپتز^۵، تیبیک زراورتنتر^۶ و لیتیک زراورتنتر^۷ طبقه‌بندی شدند. نتایج محاسبه مساحت گسترش و درصد مساحت هر کدام از کلاس‌ها با توجه به مدل‌های جنگل تصادفی، درخت تصمیم توسعه‌یافته و رگرسیون لجستیک چندجمله‌ای در جدول (۲) ارائه شده است. کلاس خاک تیبیک کلسی-زریپتز ۶۳/۹۳ درصد با مدل جنگل تصادفی، ۷۶/۷۵ درصد با مدل درخت تصمیم توسعه‌یافته و ۶۴/۱۲ درصد با مدل رگرسیون لجستیک چندجمله‌ای و کلاس جیپسیک هاپلوزریپتز، ۰/۸۰ درصد با مدل جنگل تصادفی، ۰/۹۴ درصد با مدل درخت تصمیم توسعه‌یافته و ۰/۷۳ درصد با مدل رگرسیون لجستیک چندجمله‌ای به ترتیب بیش‌ترین و کمترین فراوانی و مساحت را در بین هر سه مدل به خود اختصاص دادند. نتایج ویژگی‌های فیزیکی و شیمیایی و طبقه‌بندی خاک‌رخ‌ها در منطقه مورد مطالعه در

1. Entisols
2. Inceptisols
3. Typic Calcixerepts
4. Typic Haploxerepts
5. Gypsic Haploxerepts
6. Typic Xerorthents
7. Lithic Xerorthents

شانه شیب دارای عمق کم و زهکشی متوسط است؛ و با توجه به موقعیت مکانی آن امکان نفوذ عمقی آب باران کم و در نتیجه افزایش آبدوی و فرسایش ذرات ریز در خاک سطحی می‌شود. خاک‌رخ شاهد شماره (۴) در رده انتی‌سول قرار داشته و به دلیل واقع شدن در زمین‌نمای دشت دامنه‌ای و تأثیر رسوبات دوره‌ای فاقد تکامل پروفیلی است. در این زمین‌نما شیب زمین تا حد زیادی کاهش یافته و تقریباً به صورت مسطح درآمد است و ضخامت لایه‌های رسوب قابل توجه است.

رخ حاصل فرآیند تجمع آهک ثانویه در افق‌های زیرین است. تشکیل افق کلسیک به دلیل افزایش میزان کربنات کلسیم از سطح به عمق تحت تأثیر فرآیند آبشویی از لایه‌های سطحی و انتقال و انباشت آن‌ها در لایه‌های زیرسطحی یا تجمع در محل آهک ثانویه در افق خاک است (تان و همکاران، ۲۰۱۴؛ شیتزل و تامپسون، ۲۰۱۵). خاک‌رخ شاهد شماره (۳) خاکی جوان، بدون تکامل پروفیلی و در رده انتی‌سول طبقه‌بندی شده است. این خاک‌رخ با قرار گرفتن در زمین‌نمای تپه‌ماهور و موقعیت

جدول ۲- فراوانی کلاس‌های خاک در سطح زیرگروه با استفاده از الگوریتم‌های یادگیری ماشین

| کلاس‌های خاک | زمین‌نما | مدل‌های یادگیری ماشین | | | | | |
|-------------------|--------------|-----------------------|---------|------------------------|----------|---------------------------|-------|
| | | جنگل تصادفی | | درخت تصمیم توسعه یافته | | رگرسیون لجستیک چندجمله‌ای | |
| | | درصد | هکتار | درصد | هکتار | درصد | هکتار |
| تیبیک کلسی زیرتپه | تپه‌ماهور | ۶۳/۹۳ | ۸۸۳۷/۸۹ | ۷۶/۷۵ | ۱۰۶۱۰/۱۳ | ۸۸۶۳/۶۳ | ۶۴/۱۲ |
| جیسیک هاپلوزریتز | | ۰/۸۰ | ۱۰۹/۶۱ | ۰/۹۴ | ۱۳۱/۱۳ | ۱۰۲/۱۶ | ۰/۷۳ |
| لیتیک زراورتنز | دشت دامنه‌ای | ۶/۶۶ | ۹۲۱/۲۷ | ۰/۸۹ | ۱۲۳/۸۷ | ۹۰۹/۱۲ | ۶/۵۷ |
| تیبیک زراورتنز | | ۱۹/۴۸ | ۲۶۹۲/۸۳ | ۱۰/۷۳ | ۱۴۸۳/۶۴ | ۲۷۰۷/۱۲ | ۱۹/۵۸ |
| تیبیک هاپلوزریتز | | ۹/۱۳ | ۱۲۶۱/۷۲ | ۱۰/۶۶ | ۱۴۷۴/۵۵ | ۱۲۴۱/۲۷ | ۸/۹۷ |

به دلیل متأثر بودن از مواد مادری (مارن گچی) و اثر شیب از توسعه قابل توجهی در منطقه برخوردار نیستند. خاک‌رخ‌های ۴ و ۵ به دلیل وقوع فرآیندهای فرسایش و رسوب‌گذاری از تکامل پروفیلی کم و از طرفی گستره کمتر زمین‌نمای دشت دامنه‌ای نسبت به تپه‌ماهور برخوردار هستند.

متغیرهای محیطی منتخب

بر اساس اهداف تحقیق و روش‌های انتخاب متغیر محیطی از میان ۵۷ متغیر محیطی تولیدشده، ۱۰ متغیر محیطی شامل اطلاعات نقشه‌های ژئومورفولوژی، اطلاعات زمین‌شناسی و ویژگی‌های مستخرج از مدل رقومی ارتفاع شامل تجزیه و تحلیل سایه‌اندازی تپه‌ها، طلوع خورشید^۲، عمق دره^۳، شاخص طول در جهت

فرزام‌نیا و همکاران (۱۳۹۴) در مطالعه خود گزارش نمودند که خاک‌های تشکیل شده در زمین‌نمای دشت دامنه‌ای به علت فرآیندهای متوالی فرسایش و رسوب‌گذاری دارای ضخامت بیشتر و تحول پروفیلی کمتر هستند. خاک‌رخ شاهد شماره (۵) خاکی است از رده اینسپتی‌سول، نسبتاً عمیق با زهکشی مناسب که در انتهای زمین‌نمای دشت دامنه‌ای قرار گرفته است. وجود افق مشخصه کمبیک در این خاک‌رخ نشان‌دهنده توسعه کم و مراحل اولیه تکامل است.

نتایج نشان داد که خاک‌رخ شاهد ۱ بیش‌ترین فراوانی را در مقایسه با سایر خاک‌رخ‌های شاهد به خود اختصاص داده است. این خاک‌رخ عمدتاً در مناطقی از زمین‌نما قرار دارند که امکان تجمع املاح و زمان کافی برای وقوع فرآیندهای خاک‌سازی در آن فراهم است و دارای شرایط پایداری بوده و گستره قابل توجهی از منطقه را به خود اختصاص داده است. خاک‌رخ‌های شاهد ۲ و ۳

1. Analytical Hillshading

2. Sunset

3. Valley Depth

لجستیک چندجمله‌ای به ترتیب برابر ۶۵، ۶۰ و ۶۶ درصد پیش‌بینی‌شده و مدل رگرسیون لجستیک چندجمله‌ای بهترین عملکرد برای پیش‌بینی زیرگروه‌های خاک را نشان می‌دهد. در پژوهشی مشابه توسط فاتحی و همکاران (۱۳۹۴) در بخشی از اراضی استان کرمانشاه نشان داده شد که صحت کلی نقشه پیش‌بینی زیرگروه‌های خاک با روش‌های رگرسیون لجستیک چندجمله‌ای و درختان طبقه‌بندی به ترتیب ۵۰ و ۴۷ درصد است که مؤید آن است که مدل رگرسیون لجستیک توانایی بیشتری برای پیش‌بینی زیرگروه‌های خاک در منطقه مورد مطالعه داشته است. جین و همکاران (۲۰۱۸) در مطالعه نقشه‌برداری رقومی کلاس‌های خاک در غرب هائیتی با مدل رگرسیون لجستیک چندجمله‌ای و مدل درختان تصادفی به این نتیجه رسیدند که رگرسیون لجستیک چندجمله‌ای با مقدار شاخص کاپا ۰/۴۵ کارایی بهتری نسبت به مدل درخت تصادفی داشته است.

نتایج دو شاخص صحت کاربر و صحت تولیدکننده برای کلاس‌های خاک در سطح زیرگروه در جدول ۶ ارائه شده است. بر اساس نتایج به‌دست‌آمده، زیرگروه تیبیک کلسی‌زپتیز که ۶۴/۱۲ درصد از نقاط مشاهداتی را به خود اختصاص داده است در مدل رگرسیون لجستیک چندجمله‌ای به ترتیب دارای صحت کاربر ۶۴ درصد و صحت تولیدکننده ۹۴ درصد است که نشان می‌دهد این مدل دارای توانایی بالاتری برای پیش‌بینی این کلاس نسبت به دو مدل جنگل تصادفی (صحت کاربر ۶۱ درصد و صحت تولیدکننده ۸۵ درصد) و درخت تصمیم توسعه‌یافته (صحت کاربر ۶۲ درصد و صحت تولیدکننده ۸۰ درصد) است. از طرفی زیرگروه جیپسیک هاپلوزپتیز با کمترین فراوانی نقاط مشاهداتی (۰/۷۳ درصد) کمترین صحت کاربر و صحت تولیدکننده در مدل رگرسیون لجستیک چندجمله‌ای (صحت کاربر ۲۰ درصد و صحت تولیدکننده ۳۴ درصد) را دارا است. جین و همکاران (۲۰۱۸) بالاترین مقدار صحت کاربر و صحت تولیدکننده را ۷۹ و ۶۹ درصد و کمترین مقدار را برای

شیب^۱، فاصله تا شبکه آبراهه^۲، شاخص رطوبتی توپوگرافی^۳ و شاخص همواری بالای پشته با درجه تفکیک بالا^۴ به‌عنوان مؤثرترین متغیرهای محیطی برای پیش‌بینی کلاس‌های خاک شناخته‌شده و به‌عنوان ورودی مدل انتخاب گردید (جدول ۴). اثر فرآیندهای ژئومورفولوژی در توسعه خاک توسط تعدادی از محققین مانند اسکول و همکاران (۲۰۰۵)، جعفری و همکاران (۲۰۱۲)، تقی‌زاده و همکاران (۲۰۱۵) بررسی شده است. آن‌ها تأکید کردند اطلاعات ژئومورفولوژی مهم‌ترین متغیر کمکی برای تخمین نقشه کلاس‌های خاک در مناطق خشک و نیمه‌خشک است. ادھیکاری و همکاران (۲۰۱۴) در مطالعه خود بیان کردند نقشه زمین‌شناسی در کنار مقدار رس خاک سطحی و ارتفاع، مهم‌ترین پارامترهای محیطی برای بیان تغییرات خاک با استفاده از سامانه رده‌بندی جهانی هستند. جعفری و همکاران (۲۰۱۳) و ویس و لاگچری (۲۰۱۵) در مطالعه خود گزارش کردند که ویژگی‌های توپوگرافی و فرآیندهای ژئومورفیک در پیش‌بینی کلاس‌های خاک از عوامل مؤثر بر تشکیل خاک در مناطق مورد مطالعه هستند.

کارایی مدل‌های پیش‌بینی کننده

نتایج مقادیر صحت‌سنجی پیش‌بینی مکانی هر یک از کلاس‌های خاک بر اساس چهار شاخص صحت کلی، شاخص کاپا، صحت کاربر و صحت تولیدکننده در جداول ۵ و ۶ ارائه شده است. مطابق با نتایج ارائه‌شده، شاخص کاپا در مدل رگرسیون لجستیک چندجمله‌ای ۰/۴۱، در مدل جنگل تصادفی ۰/۳۲ و در مدل درخت تصمیم توسعه‌یافته ۰/۳۵ است که مؤید آن است که این شاخص در مدل رگرسیون لجستیک چندجمله‌ای دارای توافق متوسط و در دو مدل دیگر دارای توافق ضعیف است. مقادیر شاخص صحت کلی برای مدل جنگل تصادفی، مدل درخت تصمیم توسعه‌یافته و رگرسیون

1. LS_Factor
2. Channel Network Distance
3. Topographic Wetness Index, TWI
4. Multi-Resolution Ridge Top Flatness Index, MRRTF

نتایج برازش مدل‌های یادگیری ماشین بر روی کلاس‌های خاک نشان داد که مدل رگرسیون لجستیک چندجمله‌ای بالاترین میزان صحت را برای پیش‌بینی زیرگروه‌های خاک ارائه داده است. شکل ۳ نتایج اهمیت نسبی متغیرهای محیطی و شکل ۴ پراکنش مکانی چهار مورد از مهم‌ترین متغیرهای محیطی در مدل رگرسیون لجستیک چندجمله‌ای را نشان می‌دهد. بر اساس نتایج از میان متغیرهای محیطی منتخب چهار متغیر وابسته به توپوگرافی شامل عمق دره، فاصله تا شبکه آبراهه، شاخص همواری بالای پشته با درجه تفکیک بالا و شاخص طول در جهت شیب دارای بیش‌ترین اهمیت هستند که نشان می‌دهد توپوگرافی مهم‌ترین عامل خاک‌سازی در منطقه مورد مطالعه است. در همین راستا موسوی و همکاران (۱۳۹۸) پارامترهای توپوگرافی را به‌عنوان مهم‌ترین پیش‌ران‌های محیطی برای مدل‌سازی کلاس‌های خاک گزارش نمودند. عباس‌زاده افشار و همکاران (۱۳۹۷) بیان کردند متغیرهای محیطی مستخرج از مدل رقومی ارتفاع به دلیل داشتن همبستگی بالا با کلاس‌های خاک موجب ارتقاء صحت نقشه‌های رقومی خاک تهیه شده است. در مناطق با شدت پستی و بلندی بالا، مشخصه‌های توپوگرافی معمولاً پیش‌بینی‌کننده مناسبی در مدل‌سازی کلاس‌های خاک هستند و اغلب این ویژگی‌ها به‌عنوان پیش‌بینی‌شده کلیدی در نقشه‌برداری رقومی خاک استفاده می‌شوند (مک‌برانتی و همکاران، ۲۰۰۳). کلاس‌ها و ویژگی‌های خاک در مناطق خشک و نیمه‌خشک غالباً به‌وسیله مواد مادری و توپوگرافی تعیین و کنترل می‌شوند (مارکوس و مرکلی، ۲۰۰۷). لاگچری و همکاران (۲۰۰۷) و دبلا و اتزلمور (۲۰۰۹) تأثیر معنی‌دار خصوصیات توپوگرافی را به ترتیب در پیش‌بینی کلاس‌ها و ویژگی‌های خاک و پراکنش گروه‌های خاک با استفاده از مدل رگرسیون لجستیک چندجمله‌ای نشان دادند.

پراکنش مکانی کلاس‌های خاک

برخی از کلاس‌های خاک حداقل یا نزدیک به صفر گزارش نمودند. بر اساس نتایج به دست آمده به نظر می‌رسد کلاس‌هایی که فراوانی بیش‌تری در مشاهده‌ها دارند صحت بالاتری در پیش‌بینی مدل نشان می‌دهند. برای افزایش صحت کلاس‌هایی که دارای کمترین صحت و صحت نامشخص هستند، می‌توان از تعداد نقاط مشاهداتی بیشتری استفاده کرد. دلیل دیگر این صحت نامشخص و کم، داده‌های محیطی به‌کاررفته در پهنه بزرگی از منطقه است که باعث می‌شود تغییرات خاک با صحت کمتری پیش‌بینی شود. صحت پیش‌بینی نسبتاً ضعیفی برای طبقات طبقه‌بندی خاک با حجم کم نمونه در ایران مرکزی گزارش شده است (جعفری و همکاران ۲۰۱۲). در مدل‌سازی گروه‌های خاک مرجع ایران، مشخص شد که گروه‌های خاک مرجع با تعداد نمونه بیشتر، با الگوریتم‌های یادگیری ماشین با دقت بیشتری نسبت به کلاس‌هایی با تعداد نمونه کم مدل‌سازی می‌شوند (تقی‌زاده و همکاران، ۲۰۲۰). هنگل و همکاران (۲۰۰۷) علت این صحت کم در پیش‌بینی کلاس‌های خاک را تعداد کم خاک‌رخ‌های مشاهداتی و همچنین دخیل نبودن سایر متغیرهای محیطی که نماینده سایر عوامل خاک‌سازی در تبیین ارتباط خاک و زمین‌نما باشند، می‌دانند. یانگ و همکاران (۲۰۱۱) در مطالعه خود نتیجه گرفتند تعداد نمونه‌ها، سطح منطقه مورد مطالعه و مقیاس نقشه از جمله خصوصیات هستند که نتایج و عملکرد پیش‌بینی را تحت تأثیر قرار می‌دهند. مارچیتی و همکاران (۲۰۱۱) گزارش کردند مدل رگرسیون لجستیک چندجمله‌ای یک برآورد سریع و قابل‌اعتماد برای متغیرهای طبقه‌ای نظیر کلاس خاک دارد. جعفری و همکاران (۲۰۱۲) بیان کردند در صورت شناخت کامل منطقه مورد بررسی و انتخاب متغیر محیطی مناسب مدل رگرسیون لجستیک چندجمله‌ای پتانسیل بالایی در پیش‌بینی کلاس‌های خاک دارند.

اهمیت نسبی متغیرهای محیطی در مدل رگرسیون

لجستیک چندجمله‌ای

افق زیرسطحی کلسیک و جیپسیک در آن‌ها فراهم شده که در نهایت منجر به تکامل پروفیلی این دو خاک شده است. در مورد تشکیل خاک تیپیک زراورتنز و تیپیک هاپلوزرپتز که بر روی واحد فیزیوگرافی دشت دامنه‌ای واقع شده است می‌توان بیان کرد که دلیل تکامل عدم پروفیلی این خاک‌ها، مربوط به فرآیندهای فرسایش یا رسوب‌گذاری دوره‌ای است که شرایط پایداری زمین‌نما^۱ در این بخش از فیزیوگرافی را به صورت محلی از بین برده است.

با توجه به اینکه مدل رگرسیون لجستیک چندجمله‌ای نسبت به سایر مدل‌های یادگیری ماشین مورد استفاده در این تحقیق از صحت بالاتری برخوردار بود نقشه نهایی پیش‌بینی کلاس‌های خاک بر اساس این مدل در شکل ۵ ارائه شد. زیرگروه تیپیک کلسی‌زرپتز بیشترین و زیرگروه جیپسیک هاپلوزرپتز کمترین پراکنش را در خاک‌های منطقه مورد مطالعه به خود اختصاص داده است. گستره مکانی غالب، کلاس تیپیک کلسی‌زرپتز مربوط به بخش اعظم منطقه به جز مناطق محدودی از نواحی غربی، شمال شرقی و مرکزی است. الگوی پراکنش مکانی متغیر محیطی عمق دره رابطه معکوسی با سه متغیر فاصله تا شبکه آبراهه، شاخص همواری بالای پشته با درجه تفکیک بالا و شاخص طول در جهت شیب دارد (شکل ۳) و به نظر می‌رسد در مناطقی که حداکثر عمق دره مشاهده می‌شود عمدتاً خاک‌های تیپیک زراورتنز، لیتیک زراورتنز و تا حدودی تیپیک هاپلوزرپتز که به ترتیب خاک‌هایی جوان و یا با تکامل پروفیلی حداقلی هستند، گسترش دارند. از دیدگاه پدولوژی چنین می‌توان تفسیر نمود که با افزایش مقدار فاکتور عمق دره در منطقه مطالعاتی شرایط و زمان کافی برای تشکیل و تکامل خاک-هایی که در افق زیرسطحی آن‌ها امکان تجمع رس، آهک و گچ باشد وجود ندارد و در نقطه مقابل در بخش‌های از منطقه مطالعاتی که مقدار عمق دره کاهش یافته تشکیل خاک‌های تیپیک کلسی‌زرپتز و جیپسیک هاپلوزرپتز که خاک‌هایی متکامل‌تری هستند، فراهم شده است (شکل ۴). ارتباط بین پراکنش مکانی کلاس‌های خاک و واحدهای فیزیوگرافی شناسایی شده (جدول ۲) خاک لیتیک زراورتنز خاکی کم‌عمق در شیب زیاد (۱۲ تا ۲۵ درصد) بر روی واحد فیزیوگرافی تپه‌ماهور مشاهده شد که به دلیل واقع شدن بر روی بخش شانه شیب و فرسایش آبی ناپایدار است. دو خاک تیپیک کلسی‌زرپتز و جیپسیک هاپلوزرپتز نیز بر روی واحدهای فیزیوگرافی تپه‌ماهور مشاهده شدند اما به دلیل اینکه عمدتاً بر روی قسمت پایدار این واحد فیزیوگرافی واقع شده‌اند امکان تشکیل دو

جدول ۳- ویژگی‌های فیزیکی، شیمیایی و طبقه‌بندی خاک‌رخ‌ها در منطقه مورد مطالعه

| افق | عمق (cm) | رنگ خاک | در صد نسبی ذرات (%) | | | واکنش خاک ^۴ | قابلیت هدایت الکتریکی ^۵ (dSm ⁻¹) | ظرفیت تبادل کاتیونی ^۶ (Cmol+ kg ⁻¹) | کربن آلی ^۷ (%) | آهک ^۸ (%) | کج ^۹ |
|--|----------|----------|---------------------|-------------------|-----------------|------------------------|---|--|---------------------------|----------------------|-----------------|
| | | | شن ^۱ | سیلت ^۲ | رس ^۳ | | | | | | |
| | | | | | | | | | | | |
| A | ۰-۱۵ | 10YR4/4 | ۵۶ | ۲۸ | ۱۶ | ۷/۹۸ | -/۴۹ | ۱۲/۹۰ | ۰/۴۶ | ۱۶/۷ | - |
| Bk1 | ۱۵-۴۵ | 10YR5/4 | ۵۰ | ۲۲ | ۲۸ | ۸/۳۰ | -/۲۹ | ۱۸/۹۰ | ۰/۲۶ | ۲۶/۱ | - |
| Bk2 | ۴۵-۶۵ | 10YR5/4 | ۶۴ | ۱۶ | ۲۰ | ۸/۱۳ | -/۲۹ | ۱۶/۳۰ | ۰/۰۶ | ۲۷/۹ | - |
| C | ۶۵-۱۵۰ | 10YR5/4 | ۷۴ | ۱۲ | ۱۴ | ۸/۱۹ | -/۴۴ | ۶/۴۰ | ۰/۰۷ | ۲۵/۸ | - |
| خاک‌رخ شاهد شماره ۲ جیسیک هاپلوزریپتر (زمین‌نما: تپه‌ماهور، ارتفاع: ۱۴۴۹، شیب: ۵ تا ۸ درصد) | | | | | | | | | | | |
| A | ۰-۱۵ | 10YR7/2 | ۵۴ | ۳۶ | ۱۰ | ۷/۶۵ | ۲/۸ | ۱۲/۵۰ | ۱/۱۵ | ۲۱/۴ | ۷/۰ |
| Bky1 | ۱۵-۳۰ | 10YR8/2 | ۵۴ | ۳۶ | ۱۰ | ۷/۸۶ | ۲/۷ | ۹/۶۰ | ۰/۶۱ | ۱۶/۹ | ۲۳/۰ |
| Bky2 | ۳۰-۴۷ | 10YR5/4 | ۴۶ | ۳۶ | ۱۸ | ۷/۸۱ | ۲/۹ | ۱۵/۶۰ | ۰/۳۹ | ۱۸/۹ | ۱۰/۰ |
| Cky | ۴۷-۷۵ | 7.5YR5/4 | ۵۲ | ۳۰ | ۱۸ | ۷/۶۲ | ۹/۲۶ | ۱۰/۶۰ | ۰/۳۵ | ۱۸/۳ | ۲۰/۰ |
| Cy | ۷۵-۱۵۰ | 7.5YR5/4 | ۵۸ | ۳۰ | ۱۲ | ۷/۶۲ | ۸/۱۰ | ۱۰/۳۰ | ۰/۱۹ | ۱۳/۱ | ۳۹ |
| خاک‌رخ شاهد شماره ۳ لیتیک زراورتنتر (زمین‌نما: تپه‌ماهور، ارتفاع: ۱۶۶۹۹، شیب: ۱۲ تا ۲۵ درصد) | | | | | | | | | | | |
| A | ۰-۱۰ | 10YR5/6 | ۱۶ | ۴۸ | ۳۶ | ۷/۹۴ | -/۴۳ | ۲۵/۶۰ | ۰/۱۱ | ۱۹/۳ | - |
| C | ۱۰-۴۰ | 10YR5/6 | ۱۲ | ۴۸ | ۴۰ | ۸/۷۷ | ۱/۰۴ | ۲۳/۹۰ | ۰/۱۳ | ۲۳/۳ | - |
| Cr | ۴۰-۸۰ | | ۱۶ | ۵۰ | ۳۴ | ۸/۲۶ | ۲/۵۳ | ۱۹/۲ | ۰/۰۶ | ۲۵/۵ | - |
| خاک‌رخ شاهد شماره ۴ تیپیک زراورتنتر (زمین‌نما: دشت دامنه‌ای، ارتفاع: ۱۴۱۷، شیب: ۲ تا ۵ درصد) | | | | | | | | | | | |
| A | ۰-۲۰ | 10YR6/3 | ۶۸ | ۲۲ | ۱۰ | ۷/۸۱ | -/۶۰ | ۹/۱۰ | ۰/۲۸ | ۱۳/۱ | - |
| C1 | ۲۰-۴۵ | 10YR5/3 | ۶۸ | ۲۰ | ۱۲ | ۷/۷۹ | ۱/۰۱ | ۱۰/۶۰ | ۰/۲۳ | ۱۴/۳ | - |
| C2 | ۴۵-۸۰ | 10YR5/3 | ۶۸ | ۲۰ | ۱۲ | ۷/۹۳ | ۱/۰۳ | ۹/۹۰ | ۰/۲۰ | ۱۱/۵ | - |
| C3 | ۸۰-۱۵۰ | 10YR5/3 | ۸۰ | ۱۴ | ۶ | ۸/۱۵ | -/۶۵ | - | ۰/۱۰ | ۱۲/۶ | - |
| خاک‌رخ شاهد شماره ۵ تیپیک هاپلوزریپتر (زمین‌نما: دشت دامنه‌ای، ارتفاع: ۱۴۷۹، شیب: ۲ تا ۵ درصد) | | | | | | | | | | | |
| A | ۰-۲۵ | 10YR4/4 | ۲۱ | ۵۱ | ۲۸ | ۸/۰۲ | -/۷۶ | ۱۸/۰۰ | ۰/۲۵ | ۲۱/۶ | - |
| Bw1 | ۲۵-۸۰ | 10YR4/3 | ۱۲ | ۴۲ | ۴۶ | ۸/۶۳ | ۱/۵۳ | ۲۸/۵۰ | ۰/۱۶ | ۲۲/۶ | - |
| Bw2 | ۸۰-۱۵۰ | 10YR4/3 | ۸ | ۴۲ | ۵۰ | ۸/۶۶ | ۳/۷۷ | - | ۰/۱۷ | ۱۸/۷ | - |

۱- شن (sand)، ۲- سیلت (silt)، ۳- رس (clay)، ۴- واکنش خاک (pH)، ۵- قابلیت هدایت الکتریکی (EC)، ۶- ظرفیت تبادل کاتیونی (CEC)، ۷- کربن آلی (OC)، ۸- آهک (Calcium carbonate)

جدول ۴- متغیرهای محیطی منتخب برای مدل سازی جهت پیش بینی کلاس های خاک

| منبع | نام متغیر | مقیاس | ماهیت متغیر | متغیر محیطی مورد استفاده |
|------------------------|--|----------|--------------|--------------------------|
| (زینک و همکاران، ۲۰۱۶) | نقشه ژئومورفولوژی (Geomorphology) | ۱:۵۰۰۰۰ | ژئومورفولوژی | نقشه ژئومورفولوژی |
| سازمان زمین شناسی کشور | نقشه زمین شناسی (Geology) | ۱:۲۵۰۰۰۰ | زمین شناسی | نقشه زمین شناسی |
| (ماهواره آوس، ۲۰۱۱) | مدل رقومی ارتفاع (DEM) | ۳۰ متر | | |
| (اولایا، ۲۰۰۴) | تجزیه و تحلیل سایه اندازی تپه ها (Analytical Hillshading) | ۳۰ متر | | |
| (اولایا، ۲۰۰۴) | طلوع خورشید (Sunrise) | ۳۰ متر | | |
| (اولایا، ۲۰۰۴) | عمق دره (Valley Depth) | ۳۰ متر | | |
| (اولایا، ۲۰۰۴) | شاخص طول در جهت شیب (LS_Factor) | ۳۰ متر | توپوگرافی | مدل رقومی ارتفاع |
| (اولایا، ۲۰۰۴) | فاصله تا شبکه آبراهه (Channel Network Distance) | ۳۰ متر | | |
| (اولایا، ۲۰۰۴) | شاخص رطوبتی (TWI) | ۳۰ متر | | |
| (اولایا، ۲۰۰۴) | شاخص همواری بالای پشته با درجه تفکیک بالا (MRRTF) | ۳۰ متر | | |

جدول ۵- صحت پیش بینی سطح طبقه بندی زیر گروه توسط الگوریتم های یادگیرنده

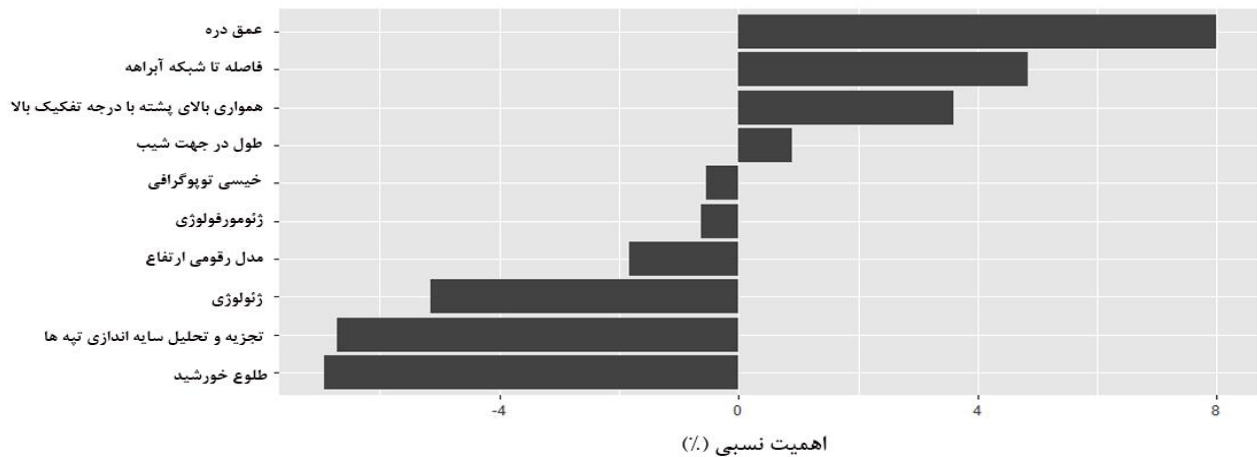
| مدل یادگیری ماشین | شاخص های صحت سنجی | |
|---------------------------|-------------------|-------------|
| | ضریب | صحت کلی (%) |
| جنگل تصادفی | ۰/۳۲ | ۶۵ |
| درخت تصمیم توسعه یافته | ۰/۳۵ | ۶۰ |
| رگرسیون لجستیک چندجمله ای | ۰/۴۱ | ۶۶ |

جدول ۶- صحت تولید کننده و کاربر برای کلاس های خاک در سطح زیر گروه بر اساس مدل های برازش داده شده

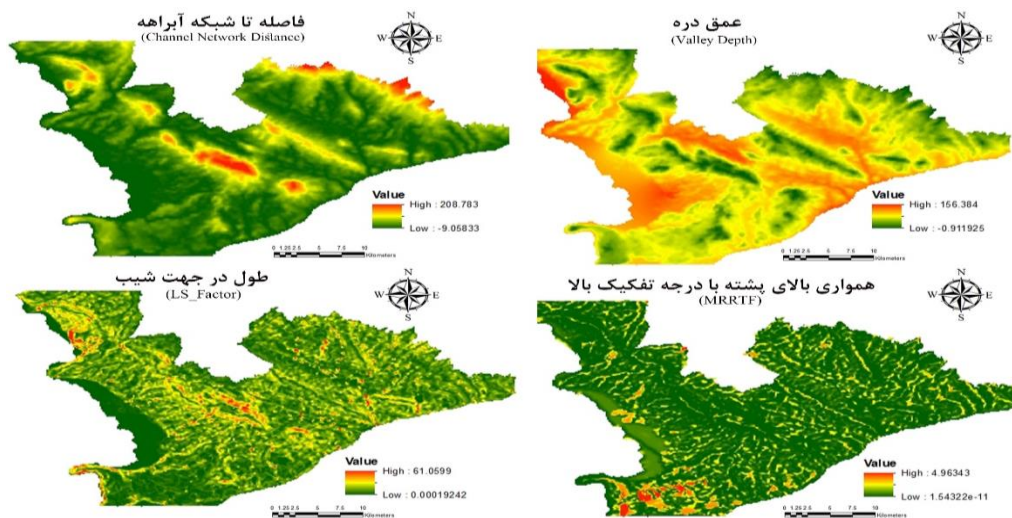
قابلیت اطمینان

| کلاس خاک در سطح زیر گروه | صحت تولید کننده (%) | | | صحت کاربر (%) | | | |
|--------------------------|---------------------|-------------|------------------------|---------------------------|-------------|------------------------|---------------------------|
| | مدل یادگیری ماشین | جنگل تصادفی | درخت تصمیم توسعه یافته | رگرسیون لجستیک چندجمله ای | جنگل تصادفی | درخت تصمیم توسعه یافته | رگرسیون لجستیک چندجمله ای |
| تیبیک کلسی زرپتزر | | ۸۵ | ۸۰ | ۹۴ | ۶۱ | ۶۲ | ۶۴ |
| تیبیک هاپلوزرپتزر | | ۵۰ | ۵۰ | ۶۷ | ۱۰۰ | ۶۷ | ۱۰۰ |
| جیپسیک هاپلوزرپتزر | | ۰ | ۰ | ۳۴ | NaN | NaN | ۲۰ |
| تیبیک زراوتنتزر | | ۳۴ | ۴۰ | ۴۰ | ۶۵ | ۴۰ | ۶۷ |
| لیتیک زراوتنتزر | | ۰ | ۰ | ۱۰۰ | NaN | NaN | ۱۰۰ |

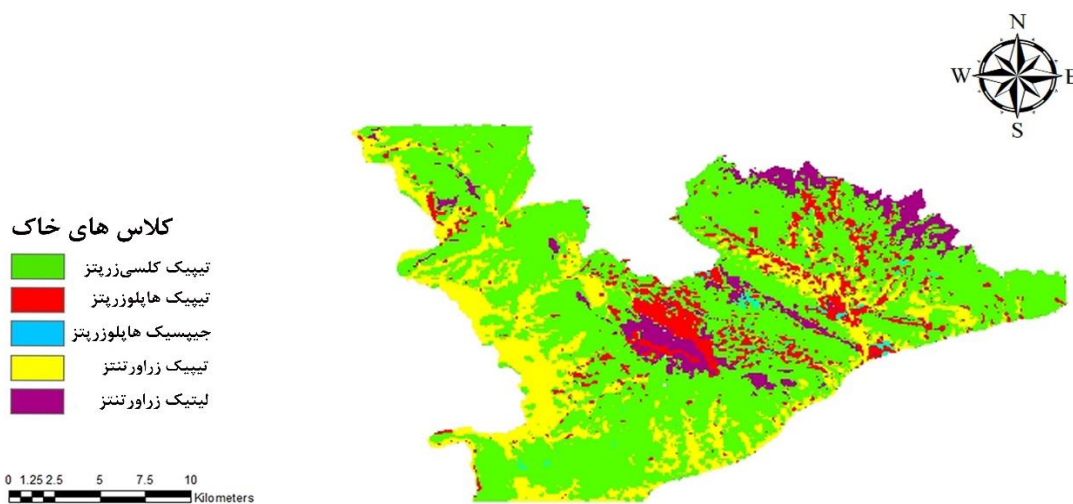
*NaN: عدد نیست، هیچ پیش بینی برای این کلاس انجام نشده است



شکل ۳- اهمیت نسبی متغیرهای محیطی پیش‌بینی کننده کلاس‌های خاک در سطح زیرگروه بر اساس مدل رگرسیون لجستیک چندجمله‌ای



شکل ۴- نقشه مهم‌ترین متغیرهای محیطی پیش‌بینی کننده کلاس خاک در مدل رگرسیون لجستیک چندجمله‌ای



شکل ۵- پراکنش مکانی زیرگروه‌های خاک با استفاده از مدل رگرسیون لجستیک چندجمله‌ای

نتیجه‌گیری کلی

این مطالعه با هدف استفاده از الگوریتم‌های جنگل تصادفی، درخت تصمیم توسعه‌یافته و رگرسیون لجستیک چندجمله‌ای به منظور پیش‌بینی توزیع مکانی کلاس‌های خاک در سطح زیرگروه انجام شد. نتایج نشان داد که روش‌های یادگیری ماشینی یکی از مؤثرترین الگوریتم‌های داده‌کاوی در نقشه‌برداری رقومی خاک هستند. نتایج صحت‌سنجی نشان داد که روش رگرسیون لجستیک چندجمله‌ای با ایجاد ارتباط قوی بین داده‌های مشاهده‌ای صحرایی و متغیرهای محیطی قادر به ارائه نقشه پیش‌بینی مکانی کلاس‌های خاک در سطح زیرگروه با دقت قابل قبول و مورد اطمینان است. استفاده از متغیرهای ژئومورفولوژی و زمین‌شناسی در کنار متغیرهای سنجش‌ازدور و مدل رقومی

ارتفاع می‌تواند در برقراری ارتباط بین خاک و زمین‌نما مؤثر باشد اما علی‌رغم استفاده از این متغیرها، متغیرهای توپوگرافی بیشترین اهمیت را در پیش‌بینی و مدل‌سازی رقومی کلاس‌های خاک داشتند. با وجود اینکه مدل رگرسیون لجستیک چندجمله‌ای توانست نسبت به سایر مدل‌ها دقت قابل قبولی ارائه دهد اما مقادیر پایین شاخص‌های دقت کاربر و تولیدکننده برای کلاس‌هایی که دارای فراوانی حداقل بودند می‌تواند به عدم تعادل داده‌های خاک در منطقه مطالعاتی مربوط باشد و پیشنهاد می‌گردد در مطالعات آتی، تأثیر متعادل‌سازی کلاس‌های خاک در مطالعات نقشه‌برداری رقومی در مناطق شیب‌دار مورد ارزیابی و نتایج صحت‌سنجی حاصله با نتایج تحقیق حاضر در این منطقه مقایسه گردد.

فهرست منابع

۱. سالنامه آماری استان زنجان ۱۳۹۸، «۱- سرزمین و آب‌وهوا»، سازمان آمار کشور.
۲. عباس زاده افشار، ف.، و ایوبی، ش.، و جعفری، ا. ۱۳۹۷. پیش‌بینی مکانی گروه بزرگ‌های خاک با استفاده از مدل‌های رگرسیونی و درخت تصمیم در منطقه جنوب شرق ایران. مهندسی زراعی (مجله علمی کشاورزی)، ۴۱(۲): ۱۳۳-۱۴۶.
۳. فاتحی، ش.، محمدی، ج.، صالحی، م.، مؤمنی، ع.، تومانیان، ت.، جعفری، ا. ۱۳۹۴. انبوه‌شدائی مکانی نقشه‌ی سنتی خاک با استفاده از رگرسیون لاجیستیک چند کلاسه و درختان طبقه‌بندی) مطالعه موردی: زیر حوضه آبخیز مرک در استان کرمانشاه. چهاردهمین کنگره علوم خاک ایران، ۲۸-۳۰ شهریور، رفسنجان، ایران، ۲۰۸-۲۱۳.
۴. فرزاد نیا، پ.، منافی، ش.، ممتاز، ح. ر. ۱۳۹۴. تشکیل و تحول خاک‌های متشکله بر روی رسوبات کواترنر در بخشی از دشت ارومیه. مجله مدیریت خاک و تولید پایدار، ۵(۲): ۹۳-۱۱۱.
۵. مصلح، ز.، صالحی، م. ح.، و جعفری، ع. ۱۳۹۶. نقشه‌برداری رقومی کلاس‌های خاک در سطوح مختلف رده‌بندی آمریکایی با استفاده از مدل رگرسیون درختی توسعه‌یافته در دشت شهرکرد. چهاردهمین کنگره علوم خاک ایران، ۲۸-۳۰ شهریور، رفسنجان، ایران، ۳۴۷-۳۴۳.
۶. مقصودی ز، رستمی‌نیا م، فرامرزی م، کشاورزی ع، رحمانی ا، موسوی س. ا. نقشه‌برداری رقومی کلاس فامیل خاک با استفاده از رویکرد یادگیری ماشین (مطالعه موردی: اراضی نیمه‌خشک غرب ایران). ۱۳۹۹. مجله علوم آب‌و خاک. ۲۴ (۲): ۱۶۳-۱۵۳.

۷. موسوی، س.، و سرمیدیان، ف.، و رحمانی، ا. ۱۳۹۸. مدل‌سازی و پیش‌بینی مکانی کلاس خاک با استفاده از الگوریتم یادگیری رگرسیون درختی توسعه‌یافته و جنگل‌های تصادفی در بخشی از اراضی دشت قزوین. تحقیقات آب‌و خاک ایران (علوم کشاورزی ایران)، ۵۰(۱۰): ۲۵۲۵-۲۵۳۸.
۸. موسسه تحقیقات خاک و آب. ۱۳۸۹. مطالعات پژوهشی مکان‌یابی، خاک‌شناسی و ارزیابی اراضی برای احداث باغات در استان زنجان. نشریه شماره ۱۵۴۷، ۳۶۴ صفحه. کرج. ایران.
9. Abeare, S. 2009. Comparisons of boosted regression tree, GLM and GAM performance in the standardization of yellowfin tuna catch-rate data from the Gulf of Mexico lonline [sic] fishery.
 10. Adhikari, K., Hartemink, A.E., Minasny, B., Bou Kheir, R., Greve, M.B., and Greve M.H. 2014. Digital mapping of soil organic carbon contents and stocks in Denmark. PLoS ONE, 9:1-13.
 11. Artieda, O., Herrero, J., and Drohan, P. J. 2006. Refinement of the differential water loss method for gypsum determination in soils. Soil Science Society of America Journal, 70(6): 1932-1935.
 12. Behrens, T., Schmidt, K., MacMillan, R. A., and Viscarra Rossel, R. A. 2018. Multi-scale digital soil mapping with deep learning. Scientific reports, 8(1): 1-9.
 13. Bouyoucos, G. J. 1962. Hydrometer method improved for making particle size analyses of soils 1. Agronomy journal, 54(5): 464-465.
 14. Breiman, L. 2001. Random forests. Machine learning, 45(1): 5-32.
 15. Breiman, L., and Cutler, A. 2004. Random Forests. Department of Statistics, University of Berkeley.
 16. 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: 68-83.
 17. Congalton, R. G. 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote sensing of environment, 37(1): 35-46.
 18. 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.
 19. Elith, J., Leathwick, J. R., and Hastie, T. 2008. A working guide to boosted regression trees. Journal of animal ecology, 77(4): 802-813.
 20. Elnaggar, A. A. 2007. Development of predictive mapping techniques for soil survey and salinity mapping. Oregon State University.
 21. Evans, D.L., Janes-Bassett, V., Borrelli, P., Chenu, C., Ferreira, C.S., Griffiths, R.I., Kalantari, Z., Keesstra, S., Lal, R., Panagos, P. and Robinson, D.A. 2022. Sustainable futures over the next decade are rooted in soil science. European Journal of Soil Science, 73(1): e13145.
 22. Fan, N.Q., Zhao, F.H., Zhu, L.J., Qin, C.Z. and Zhu, A.X. 2022. Digital soil mapping with adaptive consideration of the applicability of environmental covariates over large areas. International Journal of Applied Earth Observation and Geoinformation, 113:102986.
 23. Garg, K.K., Anantha, K.H., Nune, R., Akuraju, V.R., Singh, P., Gumma, M.K., Dixit, S. and Ragab, R. 2020. Impact of land use changes and management practices on groundwater resources in Kolar district, Southern India. Journal of Hydrology: Regional Studies, 31: 100732.
 24. Hartemink, A. E., A. McBratney and M. d. L. Mendonca-Santos. 2008. Digital soil mapping with limited data. Springer.
 25. Helfenstein, A., Mulder, V.L., Heuvelink, G.B. and Okx, J.P. 2022. Tier 4 maps of soil pH at 25 m resolution for the Netherlands. Geoderma, 410:115659.

26. Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B., and Gräler, B. 2018. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, 6, e5518.
27. Hengl, T., Toomanian, N., Reuter, H., and Malakouti, M.J. 2007. Methods to interpolate soil categorical variables from profile observations: Lessons from Iran. *Geoderma*, 140: 417–427.
28. 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(2): 284–298.
29. Jafari, A., Ayoubi, S., Khademi, H., Finke, P. A., and Toomanian, N. 2013. Selection of a taxonomic level for soil mapping using diversity and map purity indices: a case study from an Iranian arid region. *Geomorphology*, 201:86-97.
30. Jenny, H. 1994. Factors of soil formation: a system of quantitative pedology. Courier Corporation.
31. Jensen, J. R. 1996. Introductory digital image processing: a remote sensing perspective (No. Ed. 2). Prentice-Hall Inc.
32. Jeune, W., Francelino, M. R., Souza, E. D., Fernandes Filho, E. I., & Rocha, G. C. 2018. Multinomial logistic regression and random forest classifiers in digital mapping of soil classes in western Haiti. *Revista Brasileira de Ciência do Solo*, 42.
33. Kempen, B., Brus, D. J., Heuvelink, G. B., and Stoorvogel, J. J. 2009. Updating the 1: 50,000 Dutch soil map using legacy soil data: A multinomial logistic regression approach. *Geoderma*, 151(3-4): 311-326.
34. Kirkham, M. B. 2014. Principles of soil and plant water relations. Academic Press.
35. Kleinbaum, A. M. 2018. Reorganization and tie decay choices. *Management Science*, 64(5): 2219-2237.
36. Kuhn, M., and Johnson, K. 2013. Applied predictive modeling (Vol. 26, p. 13). New York: Springer.
37. Lagacherie, P., McBratny, A.B. and Volts, M. 2007. Digital soil mapping: An introductory perspective. *Developments in soil science* 31(Elsevier, Amsterdam).
38. Lanyon, L. E., and Heald, W. R. 1983. Magnesium, calcium, strontium, and barium. *Methods of Soil Analysis: Part 2 Chemical and Microbiological Properties*, 9: 247-262.
39. Liang, P., Qin, C.-Z., Zhu, A.-X., Hou, Z.-W., Fan, N.-Q., Wang, Y.-J. 2020. A case-based method of selecting covariates for digital soil mapping. *Journal of Integrative Agriculture* 19: 2–11
40. Ma, Y., Minasny, B., Malone, B. P., and Mcbratney, A. B. 2019. Pedology and digital soil mapping (DSM). *European Journal of Soil Science*, 70(2): 216-235.
41. Mallah, S., Delsouz Khaki, B., Davatgar, N., Scholten, T., Amirian-Chakan, A., Emadi, M., Kerry, R., Mosavi, A.H. and Taghizadeh-Mehrjardi, R., 2022. Predicting Soil Textural Classes Using Random Forest Models: Learning from Imbalanced Dataset. *Agronomy*, 12(11): 2613.
42. Marchetti, A., Piccini, C., Santucci, S., Chiuchiarelli, I., and Francaviglia, R. 2011. Simulation of soil types in Teramo province (Central Italy) with terrain parameters and remote sensing data. *Catena*, 85(3): 267-273.
43. Markus, E., and Merkli, C. 2007. Weathering, mineralogical evolution and soil organic matter along aHolocene soil toposequence developed on carbonate-rich materials. *Geomorphology* 97: 675-696.
44. McBratney, A. B., Santos, M. M., and Minasny, B. 2003. On digital soil mapping. *Geoderma*, 117(1-2): 3-52.
45. McBratney, A., de Gruijter, J., and Bryce, A. 2019. Pedometrics timeline. *Geoderma*, 338: 568-575.
46. Olaya, V. 2004. A gentle introduction to SAGA GIS. The SAGA user group.
47. Padarian, J., Minasny, B., and McBratney, A. B. 2020. Machine learning and soil sciences: A review aided by machine learning tools. *Soil*, 6(1): 35-52.

48. Pelegrino, M. H. P., Silva, S. H. G., Menezes, M. D. D., Silva, E. D., Owens, P. R., and Curi, N. 2016. Mapping soils in two watersheds using legacy data and extrapolation for similar surrounding areas. *Ciência e Agrotecnologia*, 40:534-546.
49. Perry Jr, C. R., & Lautenschlager, L. F. 1984. Functional equivalence of spectral vegetation indices. *Remote sensing of environment*, 14(1-3): 169-182.
50. Richards, L.A. 1954. Diagnosis and improvement of saline and alkali soils (Vol. 78, No. 2, p. 154). LWW.
51. Schaetzl, R. J., and Thompson, M. L. 2015. *Soils*. Cambridge university press.
52. 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.
53. Soil science division staff. 2017. "Soil survey manual". USDA Handbook 18.120-131
54. Staff, S. S. 2014. Keys to soil taxonomy. United States Department of Agriculture: Washington, DC, USA.
55. Sumner, M. E., and Miller, W. P. 1996. Cation exchange capacity and exchange coefficients. *Methods of soil analysis: Part 3 Chemical methods*, 5:1201-1229.
56. Santra, P., Meena, H.M. and Yadav, O.P. 2021. Spatial and temporal variation of photosynthetic photon flux density within agrivoltaic system in hot arid region of India. *Biosystems Engineering*, 209:74-93.
57. Taghizadeh-Mehrjardi, R., Minasny, B., Sarmadian, F., and Malone, B. P. 2014. Digital mapping of soil salinity in Ardakan region, central Iran. *Geoderma*, 213:15-28.
58. 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: 67-77.
59. Taghizadeh-Mehrjardi, R., Mahdianpari, M., Mohammadimanesh, F., Behrens, T., Toomanian, N., Scholten, T. and Schmidt, K., 2020. Multi-task convolutional neural networks outperformed random forest for mapping soil particle size fractions in central Iran. *Geoderma*, 376, p.114552.
60. Tan, W. F., Zhang, R., Cao, H., Huang, C. Q., Yang, Q. K., Wang, M. K., and Koopal, L. K. 2014. Soil inorganic carbon stock under different soil types and land uses on the Loess Plateau region of China. *Catena*, 121: 22-30.
61. Vaysse, K., and Lagacherie, P. 2015. Evaluating digital soil mapping approaches for mapping Global Soil Map soil properties from legacy data in Languedoc-Roussillon (France). *Geoderma Regional*, 4: 20-30.
62. Walkley, A., and Black, I. A. 1934. An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil science*, 37(1): 29-38.
63. Wiesmeier, M., Urbanski, L., Hobbey, E., Lang, B., von Lützw, M., Marin-Spiotta, E., van Wesemael, B., Rabot, E., Ließ, M., Garcia-Franco, N. and Wollschläger, U. 2019. Soil organic carbon storage as a key function of soils-A review of drivers and indicators at various scales. *Geoderma*, 333: 149-162.
64. Xiao-Lin, S. U. N., Yu-Guo, Z. H. A. O., ZHANG, G. L., Sheng-Chun, W. U., Yu-Bon, M. A. N., and Ming-Hung, W. O. N. G. 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(3): 339-350.
65. Yang, I., Jiao, Y., Fahmy, S., Zhu, A-X., Hann, S., Burt, J. E., and Qi, F. 2011. Updating conventional soil maps through digital soil mapping. *Soil Science Society of America Journal AbstractPedology*, 75(3): 1044-1053
66. Zhang, G.L., Liu, F., Song, X.D. 2017. Recent progress and future prospect of digital soil mapping: A review. *Journal of Integrative Agriculture* 16 (12): 2871–2885.
67. Zinck, J.A., Metternicht, G., Bocco, G. and Del Valle, H.F. eds. 2015. *Geopedology: An integration of geomorphology and pedology for soil and landscape studies*. Springer.

Modeling Spatial Distribution of Soil Classes Using Machine Learning Algorithms in Some Parts of Zanjan Provice

M. Rahimi Mashkaleh, M. A. Delavar*, M. Jamshidi and A. Sharififar

Ph.D. Student of Department of Soil Science, Faculty of Agriculture, University of Zanjan, Zanjan, Iran;
mastanehrahimi@znu.ac.ir

Associate Prof., Department of Soil Science, Faculty of Agriculture, University of Zanjan, Zanjan, Iran;
amir-delavar@znu.ac.ir

Assistant Prof., Soil and Water Research Institute, Agricultural Research, Education and Extension Organization, Karaj, Iran; mohammadjamshidi@yahoo.com

Researcher, Department of Soil Science, Faculty of Agriculture, University of Tehran, Iran;
sharififar@ut.ac.ir

Received March 6, 2023 and Accepted: July 5, 2023

Abstract

Digital soil mapping (DSM), which uses machine learning (ML) algorithms and environmental covariates, is used worldwide for predicting soil properties and classes, due to being time-effective and cost-saving. This research was conducted to compare ML models and their efficiency in predicting the spatial distribution of soil subgroups in part of Zanjan Province. For this purpose, sampling was carried out through a regular pattern with 500 meters intervals, and 148 soil profile samples were randomly collected and classified. The soils of the region at the subgroup level were categorized in five classes, including Typic Calcixerepts, Typic Haploxerepts, Gypsic Haploxerepts, Typic Xerorthents, and Lithic Xerorthents. Environmental covariates included geomorphological and geological maps, digital elevation model (DEM), and remote sensing (RS), selected by principal component analysis (PCA) and expert knowledge approaches. Fifty-seven environmental variables were extracted as representatives of soil forming factors. Modeling of the soil-landscape relationship was performed using three algorithms, namely, multinomial logistic regression (MNLr), random forest (RF), and boosted regression tree (BRT) in Rstudio software. The results of evaluation metrics such as overall accuracy and Kappa index were 65% and 0.32 for the RF algorithm, 60% and 0.35 for the BRT, and 65% and 0.41 for the MNLr. Referring to the importance of environmental variables, results showed that the four factors of valley depth, LS-Factor, channel network distance, and multiple resolution ridge top flatness index (MRRTF) were the most important variables in MNLr algorithm. Also, the results of two statistics of user's accuracy (UA) and producer's accuracy (PA) showed that the MNLr model performed better in the spatial prediction of soil at subgroup level. Soil classes with higher frequency had higher accuracy. The results of the prediction accuracy of soil classes showed that the model prediction is more accurate for the more frequent types in the observations.

Keywords: Multinomial logistic regression, Kappa Index, Digital soil mapping, Terrain attributes

* - Corresponding author's email: amir-delavar@znu.ac.ir