



طبقه بندی سرزمین جهت تهیه نقشه رقومی خاک با استفاده از تصاویر ماهواره ای و داده

های مدل رقومی ارتفاع در بخشی از اراضی آبیگ استان قزوین

صادق فرج اللهی^۱ ثمرین^۱، فریدون سرمدیان^{۲*}، ساحل دلجوان^۱، متین مولود پور^۱، حکیمه بهاری^۱

۱- دانشجوی کارشناسی ارشد علوم و مهندسی خاک، دانشکده‌گان کشاورزی و منابع طبیعی، دانشگاه تهران، کرج، ایران.

۲- استاد گروه علوم و مهندسی خاک، دانشکده‌گان کشاورزی و منابع طبیعی، دانشگاه تهران. * F.sarmad@ut.ac.ir

چکیده

امروزه رشد سریع جمعیت موجب بهره‌برداری نادرست از منابع حیاتی به ویژه خاک شده است. تحولات اخیر در سنجش از دور و فناوری اطلاعات، امکان گردآوری و به‌روزرسانی داده‌های خاک را فراهم کرده است. در این راستا، نقشه‌برداری رقومی خاک امیدبخش است زیرا می‌تواند کمبود داده‌های خاک در جهان را جبران کند. هدف این پژوهش، تهیه نقشه رقومی خاک در پنج سطح تاکسونومیک با استفاده از مدل درخت تصمیم و الگوریتم C5.0 تقویت‌شده با بوستینگ با کمک داده‌های مورد استفاده شامل تصاویر ماهواره‌ای لندست ۸، مدل رقومی ارتفاع آلوس و نقشه‌های زمین‌شناسی در محدوده‌ای به وسعت ۴۱۰۰۰ هکتار از اراضی شهرستان آبیگ می‌باشد. برای این منظور ۱۲۸ پروفیل خاک با روش شبکه‌بندی تصادفی انتخاب، نمونه‌برداری و طبقه‌بندی شدند. سپس با تحلیل مؤلفه اصلی، ۲۰ متغیر محیطی به‌عنوان نماینده عوامل خاک‌سازی انتخاب گردید. شاخص همواری دره مهم‌ترین متغیر شناخته شد. نتایج نشان داد دقت کلی مدل برای رده ۸۹٪، زیررده ۸۵٪، گروه بزرگ ۵۸٪، زیرگروه ۵۸٪ و فامیلی ۳۵٪ است. همچنین، بوستینگ موجب افزایش دقت و ضریب کاپا در تمامی سطوح گردید. بنابراین، مدل C5.0 تقویت‌شده رویکردی کارآمد برای نقشه‌برداری رقومی خاک به‌شمار می‌آید.

واژگان کلیدی: بوستینگ، سنجش از دور، ضریب کاپا، مدل درخت تصمیم، نقشه‌برداری رقومی خاک

مقدمه

خاک به عنوان بستر حیات و یکی از ارزشمندترین منابع طبیعی، نقش کلیدی در پایداری اکوسیستم‌ها و توسعه کشاورزی دارد. با این حال، طی دهه‌های اخیر، فشار ناشی از رشد جمعیت، تغییر کاربری اراضی و بهره‌برداری نادرست، این منبع حیاتی را با تهدیدهای جدی مواجه کرده است. از این رو، شناسایی دقیق ویژگی‌ها و الگوهای پراکنش مکانی خاک‌ها ضرورتی انکارناپذیر برای مدیریت پایدار منابع طبیعی و برنامه‌ریزی کاربری اراضی به‌شمار می‌رود. در سال‌های اخیر، استفاده از روش‌های پدومتری نشان داده است به علت در نظر گرفتن پیوستگی مکانی داده‌ها، دقت بیشتری در برآورد توزیع مکانی ویژگی‌های خاک نسبت به روش‌های آماری سنتی دارند (McBratney et al., 2000). در این میان، نقشه‌برداری رقومی خاک به عنوان یکی از شاخه‌های اصلی پدومتری مطرح شده و با بهره‌گیری از داده‌های کمکی یا محیطی توانسته است بسیاری از محدودیت‌های خاک‌شناسی مرسوم را برطرف کند. همین ویژگی موجب شده است که طی سال‌های اخیر پروژه‌های متعددی در سراسر جهان با هدف تهیه نقشه‌های رقومی خاک اجرا شوند و پیشرفت چشمگیری در این زمینه حاصل گرد (Minasny and McBratney, 2016). از میان مدل‌های مورد استفاده، مدل‌های درخت تصمیم‌گیری (Decision Tree Model, DTM) به ویژه الگوریتم C5.0 که از توسعه C4.0 و ID3 حاصل شده است (Quillen, 1993). همراه با روش بوستینگ توانسته‌اند نقشه‌های خاک را در سطوح مختلف تاکسونومیک در شهرستان آبیگ پرداخته است.

یادگیری مبتنی بر درخت تصمیم‌گیری منعطف است زیرا قادر به پردازش پیش‌بینی‌کننده‌های عددی، ترتیبی یا گسسته بوده و نیازی به فرضیات معمولی ندارد (Hastie et al., 2009). همچنین یادگیری جنگل تصادفی مفهومی مشابه یادگیری مبتنی بر درخت تصمیم‌گیری است و از مزایای مشابهی برخوردار است (Au et al., 2018).

مواد و روش‌ها

منطقه مورد مطالعه در شهرستان آبیگ استان قزوین قرار دارد و طول جغرافیایی آن بین ۴۴۹۰۱۷ تا ۴۵۲۰۴۷ و عرض جغرافیایی بین ۳۹۹۴۴۶۸ تا ۳۹۸۲۰۲۶ متر برحسب دستگاه مختصات (Universal Transverse Mercator coordinate system, UTM) می‌باشد. این منطقه شامل اراضی کوهستانی و دشت بوده، با میانگین بارندگی سالانه ۲۷۸ میلی‌متر در ایستگاه هواشناسی باغ کوثر و ۲۲۲ میلی‌متر در بوئین زهرا، میانگین دما نیز به ترتیب ۱۳/۸۸ و ۱۶/۸۴ درجه سانتی‌گراد است. اقلیم منطقه معتدل و نیمه خشک و در جنوب شهرستان خشک‌تر است. رژیم حرارتی خاک، ترمیک و رژیم رطوبتی زیریک بوده و در جنوب گاهی آریدیک ضعیف و در برخی نقاط آکوئیک مشاهده می‌شود (بنایی، ۱۳۷۷).

برای تحلیل داده‌ها از نرم‌افزارهای Google Earth، ArcGIS 10.4، SAGA GIS، SAS 9.4، ERDAS IMAGINE 2014، محیط نرم افزاری R به کار گرفته شد. منابع مورد استفاده شامل داده‌های حاصل از ۱۲۸ پروفیل خاک (موقعیت آن‌ها توسط مدل نمونه‌برداری شبکه نامنظم تعیین شد)، نقشه زمین‌شناسی، مدل ارتفاعی آلوس با دقت ۱۲/۵ متر، تصاویر لندست ۸ با قدرت تفکیک ۳۰ متر (مربوط به اکتبر ۲۰۱۷) و داده‌های اقلیمی بود.

در این پژوهش، از مدل رقومی ارتفاع آلوس برای استخراج پارامترهای توپوگرافی نظیر شیب و سایر شاخص‌های مرتبط استفاده شد و پردازش داده‌ها با نرم‌افزار SAGA GIS انجام گرفت، روش کار نیز براساس مدل Hengl و Rossiter (2003) تدوین گردید. تصاویر ماهواره‌ای لندست ۸ مربوط به محدوده آبیگ نیز برای شناسایی پوشش گیاهی، رطوبت خاک، کانی‌ها و دمای سطح زمین به کار گرفته شدند (Travis Nauman, 2009). پردازش و استخراج شاخص‌های مرتبط با پوشش گیاهی و خاک نیز در محیط نرم‌افزار ERDAS IMAGING صورت گرفت.

نقشه ژئوform با استفاده از نقشه زمین‌شناسی و مدل ارتفاعی رقومی تهیه شد؛ این نقشه، که به نقشه ریخت‌زمین نیز معروف است، اشکال سطح زمین و فرآیندهای مرتبط را نشان می‌دهد. برای ترکیب و بهینه‌سازی مرزها از ابزارهای Intersect و Eliminate در نرم‌افزار ArcGIS استفاده شد. نقشه لندفرم نیز با مدل ارتفاعی ماهواره آلوس و نرم‌افزار SAGA بر پایه شاخص (Topographic Position Index, TPI) طبقه‌بندی شد (Hengl and Rossiter, 2003). شاخص TPI موقعیت نسبی هر نقطه از سطح زمین را نسبت به محیط اطرافش نشان می‌دهد. در تهیه نقشه لیتولوژی نیز از نقشه‌های زمین‌شناسی کشور با مقیاس ۱/۱۰۰۰۰۰ و نرم‌افزار ArcGIS 10.4 استفاده و واحدهای استخراج شده به عنوان نقشه مبنا در نظر گرفته شد.

برای آماده‌سازی داده‌ها از روش تحلیل مؤلفه‌های اصلی (Principal Component Analysis, PCA) در SAS 9.4 استفاده شد؛ این روش با حذف متغیرهای همبسته Ramsey و Schafer (2002) مؤلفه‌های اصلی را استخراج می‌کند، این مؤلفه‌ها بیانگر ۹۵٪ واریانس داده‌ها می‌باشند. داده‌های نهایی در سه گروه دسته بندی شدند: ۱. هیدرولوژی سطحی (رطوبت، تجمع جریان، طول و...)، ۲. گروه ارتفاعی (شیب، ارتفاع، عمق دره و...)، ۳. گروه لایه بازتابی (پوشش گیاهی، روشنایی) برای مدل‌سازی، کلاس‌های خاک تا سطح زیرگروه بزرگ به عنوان متغیر وابسته و متغیرهای محیطی و واحدهای ژئوform (کیفی) نیز به مدل اضافه شده و در نهایت تمامی داده‌ها به فرمت رستری با اندازه پیکسل ۱۲/۵ متر و بر پایه مدل ارتفاعی رقومی تهیه شدند.

الگوریتم درخت تصمیم با آغاز از گره‌ریشه، داده‌ها را براساس ویژگی‌های موجود الگوریتم به‌طور سلسله مراتب تا رسیدن به گره‌های برگ دسته‌بندی می‌کند. در این پژوهش پیش‌بینی پنج سطح طبقه‌بندی خاک (رده، زیررده، گروه بزرگ، زیرگروه بزرگ و فامیلی) با الگوریتم C5.0 همراه با روش بوستینگ انجام شد سپس داده‌های خاک (۱۲۸ نمونه) در

محیط R پردازش و خروجی به فرمت (American Standard Code for Information Interchange, ASCII) در GIS استفاده شد. به منظور ارزیابی دقت و اعتبار مدل، از نرم‌افزار Ithir با نسبت ۸۰٪ برای آموزش و ۲۰٪ برای آزمون بهره گرفته شد و اعتبارسنجی نیز با ماتریس خطا انجام شد به طوری که عناصر قطری نشانه طبقه‌بندی درست و عناصر غیرقطری بیانگر خطا بودند.

شاخص‌های اصلی برای سنجش دقت عبارت‌اند از:

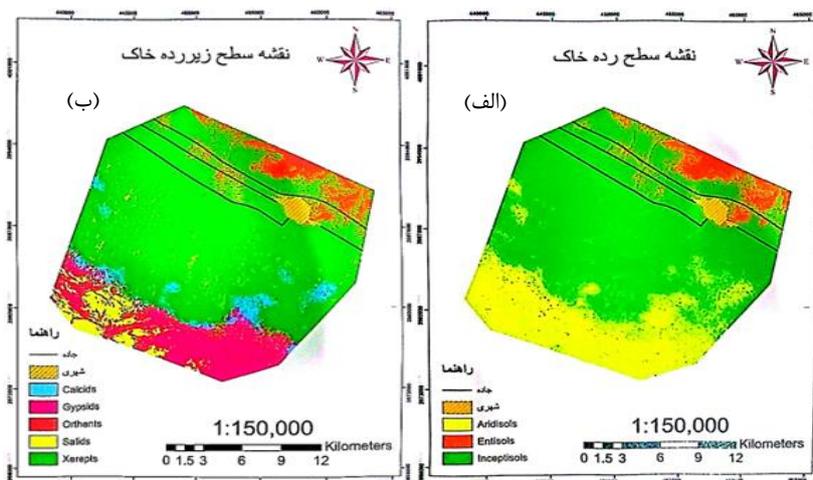
۱. دقت کلی (O.A): نسبت کل پیکسل‌های درست طبقه‌بندی شده به کل پیکسل‌های معلوم می باشد.
۲. دقت تولیدکننده (P.A): احتمال اینکه طبقه‌بندی کننده پیکسلی را به یک کلاس خاص نسبت داده باشد.
۳. دقت کاربر (U.A): بیانگر احتمال تطابق طبقه‌بندی یک کلاس با همان کلاس در نقشه واقعیت زمینی می باشد.
۴. ضریب کاپا (K): میزان انطباق با حالت تصادفی را نشان می‌دهد؛ مقدار ۱۰۰٪ نشان‌دهنده تطابق کامل و مقادیر بین ۰ تا ۱۰۰ درصد نشانگر سطوح مختلف دقت است. مقادیر منفی نیز نتایج ضعیف را نشان می‌دهند.

$$k = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} * x_{+i})}{N^2 - \sum_{i=1}^r (x_{i+} * x_{+i})} \quad (1)$$

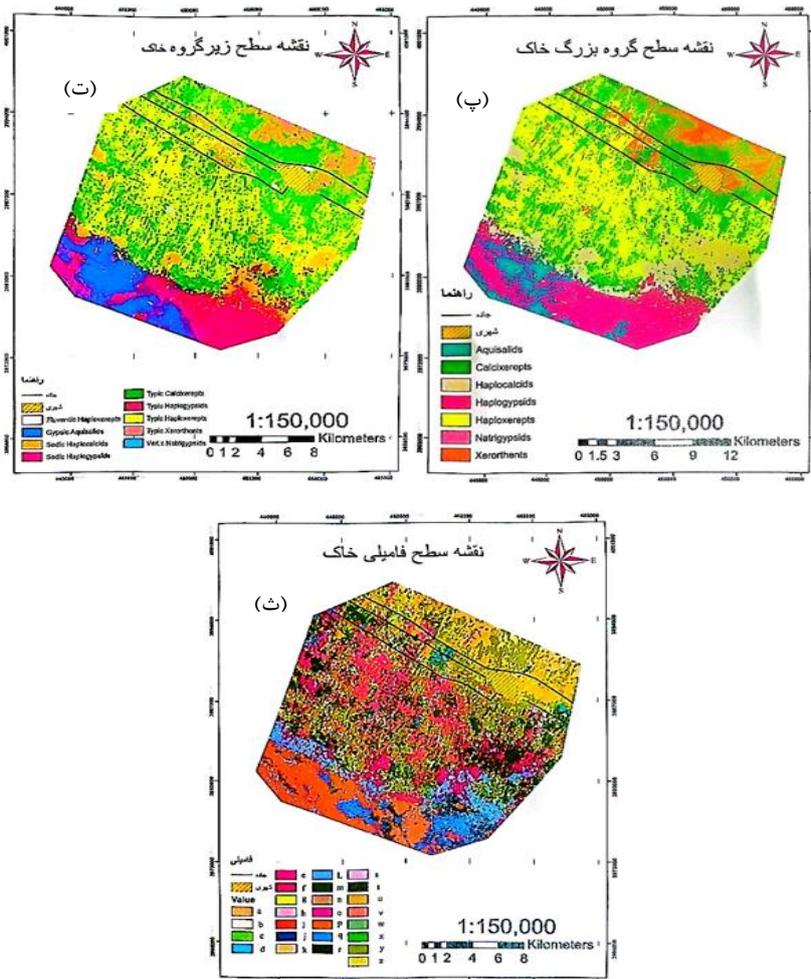
N تعداد کل پیکسل، x_{i+} مجموع عناصر سطر i و x_{+i} مجموع عناصر ستون i می باشد.

نتایج و بحث

در منطقه مورد مطالعه، براساس نمونه‌برداری از ۱۲۸ پروفیل خاک و طبقه‌بندی تاکسونومیک، ۳ نوع خاک در سطح رده شکل ۱ (الف)، ۵ نوع در سطح زیررده شکل ۱ (ب)، ۷ نوع در سطح گروه بزرگ شکل ۱ (پ)، ۹ نوع در سطح زیرگروه شکل ۱ (ت) و ۲۶ نوع در سطح فامیلی شکل ۱ (ث) شناسایی شدند. این منطقه دارای خاک‌هایی با میزان بالای آهک در بخش‌های میانی و شمالی و همچنین مقادیر قابل توجه گچ و نمک در نواحی جنوبی و جنوب شرق است که همراه با مقدار پایین ماده آلی، نشان‌دهنده شرایط خشک و نیمه‌خشک منطقه می باشد. بافت خاک با کاهش ارتفاع تغییر می‌کند؛ به طوری که با حرکت از مناطق مرتفع شمالی به سمت مناطق پایین دست جنوبی، خاک‌ها ریزبافت تر و سنگین تر می‌شوند که ناشی از تجمع رسوبات ریزبافت و فرسایش آبی است. شاخص شوری خاک نیز به دلیل بالا بودن آب زیرزمینی و ویژگی‌های لیتولوژی، در مناطق مرتفع کمتر از ۲ دسی‌زیمنس بر متر و در نواحی پایین دست (Low Land) تا ۱۰۰ دسی‌زیمنس بر متر افزایش می‌یابد.



شکل ۱. نقش رقومی خاک منطقه مورد: (الف) در سطح رده؛ (ب) در سطح زیررده؛ (پ) در سطح گروه بزرگ (ت) در سطح زیرگروه؛ (ث) در سطح فامیلی



ادامه ی شکل ۱. نقش رقومی خاک منطقه مورد : (الف) در سطح رده ؛ (ب) در سطح زیررده؛ (پ) در سطح گروه بزرگ (ت) در سطح زیرگروه؛ (ث) در سطح فامیلی

از نظر جغرافیایی و ویژگی های محیطی، خاک های زراورتنز بدون افق زیرسطحی عمدتاً در شمال یا شمال غربی، خاک های کلسی زریتنز و هاپلوزریتنز در مناطق میانی و خاک های اریدی سول شور و گچی در نواحی جنوبی منطقه پراکنده اند. برای نقشه برداری رقومی خاک، از داده های محیطی متنوعی استفاده شد که شامل نقشه زمین شناسی، تصاویر ماهواره ای لندست ۸، مدل ارتفاعی رقومی با تفکیک مکانی ۱۲/۵ متر از ماهواره آلوس و همچنین شاخص های پوشش گیاهی و شاخص های نوین مانند درخشندگی، شوری، رس و کربنات بودند (تقی زاده و همکاران، ۱۳۹۳). در میان این داده ها، شاخص همواری دره با تفکیک بالا، شاخص رطوبتی توپوگرافی (Topographic Wetness Index, TWI)، شاخص پوشش گیاهی و برخی باندهای بازتابی ماهواره ای بیشترین تأثیر را در مدل سازی طبقه بندی خاک داشتند. همچنین در شناسایی مناطق مستعد رسوب گذاری و شوری خاک، مثل نواحی جنوبی، نقش کلیدی ایفا کردند.

همچنین با استفاده از آنالیز مؤلفه های اصلی، متغیرهای غیرهمبسته و مؤثر برای مدل انتخاب شدند که در نهایت منجر به شناسایی ۲۰ متغیر کلیدی شامل پارامترهای هیدرولوژیکی، توپوگرافی و شاخص های ماهواره ای گردید.

برای پیش بینی کلاس های خاک، الگوریتم درخت تصمیم C5.0 در پنج سطح تاکسونومیک (رده، زیررده، گروه بزرگ، زیرگروه و فامیلی) به کار گرفته شد؛ در سطوح رده و زیررده، مدل با استفاده از بوستینگ به دقت ۱۰۰٪ و ضریب کاپای ۱ در مرحله آموزش رسید. که نشان دهنده طبقه بندی دقیق تمامی کلاس های اصلی است. هرچند در سطوح گروه بزرگ و زیرگروه دقت مدل کاهش یافت، اما همچنان عملکرد بهتری نسبت به مطالعات پیشین با (دقت های حدود ۵۸٪ و

ضریب کاپای ۰/۴۷ و ۰/۴۳ به ترتیب) نشان داد. در سطح فامیلی که بیشترین تنوع و تعداد کلاس‌ها (۲۶ کلاس) وجود داشت، دقت مدل به ۳۵٪ کاهش یافت که به دلیل پیچیدگی بالای طبقه‌بندی قابل پیش‌بینی بود. مقایسه مدل‌های با بوستینگ و بدون بوستینگ نیز تأیید کرد که استفاده از بوستینگ موجب افزایش چشمگیر دقت و ضریب کاپا می‌شود. تحلیل مؤلفه‌های اصلی با حذف متغیرهای همبسته، عملکرد مدل را بهبود داد و شاخص‌های توپوگرافی مرتبط با رسوب گذاری و شوری، تفسیر مدل را با واقعیت منطقه همسو کرده است. در همین راستا مطالعه‌ای مستقل در منطقه فاریاب-کهنوج نیستانی و همکاران (۱۳۹۹) نشان داد الگوریتم جنگل تصادفی نیز در طبقه‌بندی گروه‌های بزرگ خاک عملکرد قابل توجهی دارد؛ با دقت ۸۸٪ در منطقه مرجع و ۸۱٪ در منطقه پذیرنده، این نتایج بیانگر پایداری و قابلیت تعمیم‌پذیری بالای جنگل تصادفی در شرایط اقلیمی متنوع است و استفاده از آن به عنوان مدل مکمل در کنار C5.0 توصیه می‌شود. الگوریتم درخت تصمیم C5.0 همراه با بوستینگ موجب کاهش خطا و افزایش دقت شد و به دلیل قابلیت تفسیر ساده، برای کاربرد عملی مناسب است. نقشه‌های رقومی خاک با دقت مناسب در کشاورزی، مدیریت منابع و کنترل فرسایش کاربرد دارند. مدل ارائه شده کم هزینه، دقیق و قابل توسعه در مناطق مشابه است. این پژوهش ضمن تأیید اثر الگوریتم‌ها و اهمیت تحلیل مؤلفه‌های اصلی، چارچوبی برای مدلسازی مکانی خاک ارائه می‌دهد که الگویی برای پژوهش‌های آینده به شمار آید.

نتیجه گیری

ارتباط بین داده‌های خاک و متغیرهای محیطی به خوبی توسط مدل درخت تصمیم آشکار شد و به کمک روش بوستینگ این ارتباط تقویت گردید، به طوری که دقت پیش‌بینی‌های مدل به شکل محسوسی افزایش یافت. علاوه بر این، استفاده از نرم‌افزار R خطاهای ناشی از تبدیل عوارض را کاهش داد و امکان وارد کردن مستقیم داده‌های محیطی به نرم‌افزار را فراهم ساخت. بدلیل تنوع و غیریکنواختی در خاک‌ها تشخیص کلاس‌های خاک توسط مدل به سختی انجام می‌شود. در نتیجه اجرای تکنیک نقشه‌برداری رقومی با نتایج دقیق نیازمند اطلاعات ورودی دقیق با تراکم مناسب از منطقه مورد مطالعه می‌باشد. مدل درخت تصمیم در نقشه‌برداری رقومی خاک می‌تواند نقشه‌برداری خاک‌ها را در گستره‌ای وسیع به یکباره انجام داده و سبب بهبود نقشه‌های تولید شده نسبت به نقشه‌های سنتی گردد. همچنین این مدل با افزایش توان انتقال داده‌ها، امکان تولید نقشه‌های پیوسته از ویژگی‌ها و اطلاعات خاک در سراسر منطقه مورد مطالعه را فراهم می‌آورد.

منابع

۱. بنایی، م. ح. (۱۳۷۷). نقشه رژیم‌های رطوبتی و حرارتی خاک‌های ایران، مؤسسه تحقیقات خاک و آب کشور، وزارت کشاورزی، تهران، ایران.
۲. تقی زاده، ر. ا.، سرمیدیان، ف.، امید، م.، تومانیان، نورایر، روستا، رحیمیان. (۱۳۹۳). نقشه‌برداری رقومی کلاس‌های خاک با استفاده از انواع روش‌های داده کاوی در منطقه‌ی اردکان استان یزد. مهندسی زراعی، ۴(۴): ۱-۲۹.
۳. میرک زهی، خلیل اله، شهریار، پهلوان راد، بامری. (۲۰۱۷). کاربرد روش درختان تصمیم‌گیری تصادفی در پیش‌بینی کلاس‌های خاک در اراضی با پستی و بلندی کم (مطالعه موردی: شهرستان هیرمند). مجله پژوهش‌های حفاظت آب و خاک، ۲۴(۱): ۶۷-۸۴.
۴. نیستانی، م.، سرمیدیان، ف.، جعفری، ع.، کشاورزی، ع. (۱۳۹۹). برون‌یابی مکانی گروه بزرگ خاک با استفاده از روش جنگل تصادفی در منطقه خشک ایران مرکزی (فاریاب-کهنوج). منابع طبیعی ایران، ۷۲(۲): ۲۴۵-۲۵۸.
5. Au, T. C. (2018). Random forests, decision trees, and categorical predictors: the "absent levels" problem. *Journal of Machine Learning Research*, 19(45), 1-30.
6. Bui, E. N., Loughhead, A., & Corner, R. (1999). Extracting soil-landscape rules from previous soil surveys. *Soil Research*, 37(3), 495-508.
7. Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1), 37-46.

8. Grunwald, S. (2009). Multi-criteria characterization of recent digital soil mapping and modeling approaches. *Geoderma*, 152(3-4), 195-207.
9. Hastie, T., Tibshirani, R., Friedman, J., (2009). The elements of statistical learning: data mining, inference and prediction. 2nd ed. Springer, New York, NY (734pp.).
10. Hengl, T., Rossiter, D. G., & Stein, A. (2003). Soil sampling strategies for spatial prediction by correlation with auxiliary maps. *Soil Research*, 41(8), 1403-1422.
11. Jafari, A., Khademi, H., Finke, P. A., Van de Wauw, J., & Ayoubi, S. (2014). Spatial prediction of soil great groups by boosted regression trees using a limited point dataset in an arid region, southeastern Iran. *Geoderma*, 232, 148-163.
12. McBratney, A. B., Odeh, I. O., Bishop, T. F., Dunbar, M. S., & Shatar, T. M. (2000). An overview of pedometric techniques for use in soil survey. *Geoderma*, 97(3-4), 293-327.
13. Minasny, B., & McBratney, A. B. (2007). Incorporating taxonomic distance into spatial prediction and digital mapping of soil classes. *Geoderma*, 142(3-4), 285-293.
14. Minasny, B., & McBratney, A. B. (2016). Digital soil mapping: A brief history and some lessons. *Geoderma*, 264, 301-311.
15. Nauman, T. (2009). *Digital soil-landscape classification for soil survey using ASTER satellite and digital elevation data in Organ Pipe Cactus National Monument, Arizona*. The University of Arizona.
16. Pires, J. C. M., Martins, F. G., Sousa, S. I. V., Alvim-Ferraz, M. C., & Pereira, M. C. (2008). Selection and validation of parameters in multiple linear and principal component regressions. *Environmental Modelling & Software*, 23(1), 50-55.
17. Ramsey, F. L., & Schafer, D. W. (2002). *The Statistical Sleuth—A Course in Methods of Data Analysis* Duxbury. Boston, USA: Thomas Learn.
18. Saldaña, A., & Ibáñez, J. J. (2004). Pedodiversity analysis at large scales: an example of three fluvial terraces of the Henares River (central Spain). *Geomorphology*, 62(1-2), 123-138.
19. Taghizadeh-Mehrjardi, R., Nabiollahi, K., Minasny, B., & Triantafyllis, J. (2015). Comparing data mining classifiers to predict spatial distribution of USDA-family soil groups in Baneh region, Iran. *Geoderma*, 253, 67-77.
20. Wu, W., Li, A. D., He, X. H., Ma, R., Liu, H. B., & Lv, J. K. (2018). A comparison of support vector machines, artificial neural network and classification tree for identifying soil texture classes in southwest China. *Computers and Electronics in Agriculture*, 144, 86-93.
21. Yang, R. M., Zhang, G. L., Liu, F., Lu, Y. Y., Yang, F., Yang, F., ... & Li, D. C. (2016). Comparison of boosted regression tree and random forest models for mapping topsoil organic carbon concentration in an alpine ecosystem. *Ecological indicators*, 60, 870-878.

Digital Soil-Landscape Classification for Digital Soil Mapping using Satellite images and Digital Elevation Data in Abyek, Qazvin

Sadegh Farajollahi Samarein¹, Fereydoon Sarmadyan^{2*}, Sahel Deljavan¹, Matin Molodpour¹, Hakimeh Bahari¹

1- MSc student of Soil Science and Engineering, College of Agriculture & Natural Resources, University of Tehran

2- Professor Department of Soil Science and Engineering, College of Agriculture & Natural Resources, University of Tehran

Abstract

Nowadays, rapid population growth has led to the improper exploitation of vital resources, especially soil. Recent developments in remote sensing and information technology have made it possible to collect and update soil data. In this regard, digital soil mapping is promising because it can compensate for the lack of soil data in the world. The aim of this research is to prepare a digital soil map at five taxonomic levels using the Decision Tree Model and the C5.0 algorithm enhanced with boosting, using data from Landsat 8 satellite images, the ALOS digital elevation model, and geological maps in an area of 41,000 hectares of land in Abyek County. For this purpose, 128 soil profiles were selected, sampled and classified using a random grid method. Then, using principal component analysis, 20 environmental variables were selected as representatives of soil-forming factors. The valley flatness index was identified as the most important variable. The results showed that the overall accuracy of the model was 89% for order, 85% for suborder, 58% for great group, 58% for subgroup and 35% for family. Also, boosting increased the accuracy and

kappa coefficient at all levels. Therefore, the boosted C5.0 model is an efficient approach for digital soil mapping.

Keywords : Boosting, Remote Sensing, Kappa Coefficient, Decision Tree Model, Digital Soil Mapping