



محور مقاله: پدومتری و ارزیابی خاک‌ها

بررسی عملکرد جنگل تصادفی در نقشه‌برداری رقومی خاک یک منطقه نیمه‌خشک مدیترانه‌ای

آرمان بهمنی^۱، پرویز شکاری^{۲*}

^۱ دانش‌آموخته کارشناسی ارشد گروه علوم و مهندسی خاک، دانشکده کشاورزی، پردیس کشاورزی و منابع طبیعی، دانشگاه رازی، کرمانشاه، ایران.

^۲ استادیار گروه علوم و مهندسی خاک، دانشکده کشاورزی، پردیس کشاورزی و منابع طبیعی، دانشگاه رازی، کرمانشاه، ایران.

چکیده

نقشه‌های خاک از منابع عمده اطلاعات برای مدیریت پایدار سرزمین هستند. روش‌های نقشه‌برداری رقومی خاک برای تهیه نقشه‌های کم‌هزینه، دارای وضوح بالا و به‌روزرسانی آسان ابداع شده‌اند. روش‌های رایان‌آموختی به‌طور فزاینده‌ای در این زمینه به‌کار برده می‌شوند. در این پژوهش از الگوریتم جنگل تصادفی برای پیش‌بینی کلاس‌های خاک یک زمین‌چهر با مساحت حدود ۵۶۰۰۰ هکتار در منطقه میان‌دریوند استان کرمانشاه استفاده شد. با هدف اطمینان از پوشش مناسب نمونه‌ها در منطقه، یک شبکه مربعی با ۱۲۲ نقطه مطابق با روش نمونه‌برداری سیستماتیک ساده طراحی شد. همه پدان‌ها به دقت تشریح، تجزیه و سپس مطابق با سیستم مرجع جهانی (WRB) رده‌بندی شدند، که در پی آن ۵ گروه مرجع خاک در منطقه مطالعاتی شناسایی شد. برای محاسبه کواریت‌های محیطی، مدل رقومی ارتفاع و تصویر سنجنده OLI ماهواره لندست ۸ با تفکیک مکانی ۳۰ متر به‌کار برده شد. کواریت‌های محیطی به‌عنوان متغیرهای پیش‌بین برای پیش‌بینی و نقشه‌کردن الگوی پراکنش کلاس‌های خاک در سراسر منطقه مطالعاتی به‌کار رفت. مقادیر (OOB Error)، صحت عمومی نقشه (OA) و نمایه‌سازی (K) به‌عنوان سنج‌های از صحت پیش‌بینی گروه مرجع خاک به‌ترتیب ۰/۳۰، ۰/۷۰ و ۰/۵۷ محاسبه شدند. سنج‌های راستی‌آزمایی، توانایی چشم‌گیر جنگل تصادفی را در پیش‌بینی الگوی پراکنش خاک در منطقه مطالعاتی تایید نمود.

کلمات کلیدی: رایان‌آموزی، میان‌دریوند، کواریت‌های محیطی.

مقدمه

با توجه به نقش اساسی خاک در طبیعت، آگاهی از پراکنش مکانی و مدیریت درست خاک برای پایداری یک جامعه تولیدی و درک تعادل میان فرآیندهای فیزیکی و شیمیایی پیچیده‌ای که زندگی را بر روی زمین می‌سازند، ضروری است. در این راستا، دسترسی به نقشه‌هایی با قدرت تفکیک مکانی بالا برای برنامه‌ریزی کاربری زمین و دیگر فعالیت‌ها همچون جنگلداری، کشاورزی و حفاظت محیط زیست اهمیت بنیادین دارد. یک رویکرد مفید برای نقشه‌برداری خاک‌ها در سطح وسیعی از زمین‌های اصلاح نشده که با روش‌های سنتی می‌تواند زمان‌بر و پرهزینه باشد، نقشه‌برداری رقومی خاک است. نقشه‌برداری رقومی خاک برای استنتاج یا پیش‌بینی الگوی مکانی و ویژگی‌های خاک در سراسر یک زمین‌چهر میان مشاهدات صحرایی و داده‌های آزمایشگاهی با داده‌های سنجش از دور و روش‌های کمی ارتباط برقرار می‌کند. روش‌های زیادی در نقشه‌برداری رقومی خاک به‌کار می‌رود که عمدتاً در گروه روش‌های رایان‌آموختی^۱ قرار دارند. رایان‌آموختی یک اصطلاح کلی برای مجموعه‌ای گسترده از مدل‌هایی است که به منظور تحلیل الگوهای موجود در داده‌ها و انجام پیش‌بینی‌ها به‌کار می‌روند (Witten و همکاران ۲۰۱۱). اگر چه این روش‌ها اغلب به منظور داده‌کاوی در یک پایگاه داده‌ای بزرگ به‌کار می‌روند، در عین حال ابزاری جالب برای آموزش مدل‌های پیش‌بینی مکانی کلاس‌های خاک می‌باشند زیرا با روش سنتی، روابط بین کلاس‌های خاک و کواریت‌های محیطی، اغلب تا حد کمی قابل درک و فهم است (Grunwald, 2006). جنگل تصادفی^۲ (RF) یکی از روش‌های رایان‌آموختی است که در پیش‌بینی کلاس‌های خاک به‌کار رفته و عملکرد قابل قبولی داشته است. Barthold و همکاران (۲۰۱۳) نیز RF را برای پیش‌بینی پراکنش مکانی نوع خاک‌ها در چین مورد راستی‌آزمایی قرار دادند. در مطالعه‌ی آن‌ها خطای پیش‌بینی مدل برای سطح اول رده‌بندی سیستم WRB برابر ۵۱/۶٪ بود. Brungard و همکاران (۲۰۱۵) با استفاده از یازده مدل رایان‌آموختی، کلاس‌های خاک را در سه زمین‌چهر مختلف پیش‌بینی کردند و در هر سه منطقه، نتایج صحت بیشتر جنگل تصادفی را نسبت به دیگر مدل‌ها تایید نمود. انجام چنین مطالعاتی در استان کرمانشاه نیز به یک یا چند مورد محدود می‌شود. هم‌اکنون پژوهش‌هایی از این دست در ایران در مرحله آزمایشی آزمایشی بوده و چنانچه با دقت مطالعاتی و

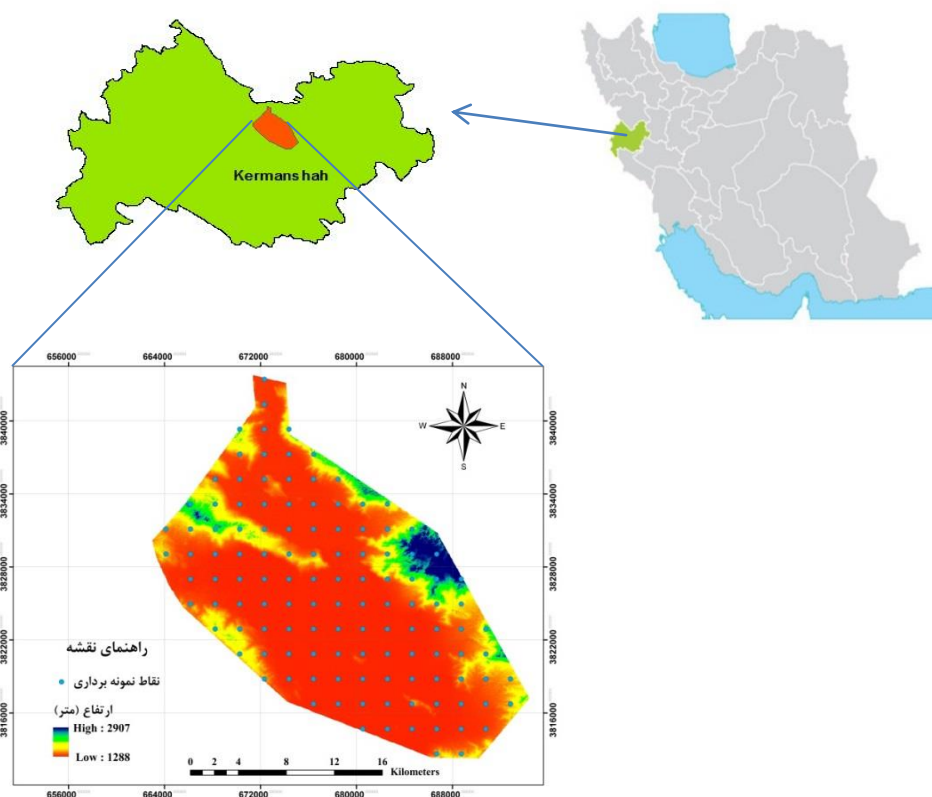
*ایمیل نویسنده مسئول: pshekaari@gmail.com

1 Machine learning
2 Random forests

مقیاس مناسب تهیه شوند می‌توان اطلاعات حاصله را در یک پایگاه داده سازمان‌دهی نموده و حسب نیاز، به عنوان اطلاعات پایه در اختیار کاربران نقشه‌های خاک و پژوهش‌گران در زمینه‌های پرشمار مانند ارزیابی سرزمین، آبیاری و زهکشی، جنگل‌کاری مصنوعی، آبخیزداری، آمایش سرزمین، پروژه‌های عمرانی و مانند آن قرار داد. پژوهش حاضر با هدف دستیابی به برآورد قابل قبولی از پراکنش مکانی گروه‌های مرجع خاک در دشت میان‌دریبد استان کرمانشاه با کاربرد مدل جنگل تصادفی انجام گردید. یکی دیگر از اهداف این پژوهش، تعیین مؤثرترین کوواریت‌های محیطی در تمایز خاک‌های منطقه مطالعاتی بود.

مواد و روش‌ها

دشت میان‌دریبد و بخشی از ارتفاعات پیرامونی آن با مساحت حدود ۵۶۰۰۰ هکتار و واقع در شمال غربی شهرستان کرمانشاه بین طول جغرافیایی ۴۶ درجه و ۴۵ دقیقه تا ۴۷ درجه و ۵ دقیقه شرقی و عرض جغرافیایی ۳۴ درجه و ۳۵ دقیقه تا ۳۴ درجه و ۲۳ دقیقه شمالی به‌عنوان منطقه مطالعاتی در نظر گرفته شد (شکل ۱). میانگین سالیانه‌ی دما، بارش و ارتفاع منطقه به ترتیب ۱۴ درجه سانتی‌گراد، ۴۶۲/۶ میلی‌متر و ۱۳۵۷/۷ متر است. رژیم رطوبتی و دمایی خاک در منطقه به ترتیب زیریک و ترمیک تعیین شده است (شرکت سهامی آب منطقه‌ای غرب، ۱۳۶۸).



شکل ۱- موقعیت منطقه مورد مطالعه و پراکنش نقاط نمونه‌برداری

با کاربرد روش سیستماتیک ساده^۱، موقعیت جغرافیایی تعداد ۱۲۲ نقطه مشاهداتی با فواصل ۲ کیلومتر در راستای طولی و عرضی تعیین شد. اما امکان نمونه‌برداری از ۷ نقطه فراهم نشد. از نقاط باقی‌مانده ۳۷ نقطه با رخنمون‌های سنگی در منطقه برخورد داشت و در مجموع ۷۸ پدان بررسی



شد. پس از یافتن مختصات نقاط مشاهداتی با استفاده از دستگاه موقعیت‌یاب جهانی^۱، در هر نقطه یک خاکرخ حفر، تشریح و از افق‌های آن نمونه‌برداری صورت گرفت. تشریح ریخت‌شناختی پدان‌ها و ریخت‌سنجی سطح^۲ زمین بر اساس بر اساس راهنمای تشریح صحرایی (Schoeneberger و همکاران ۲۰۱۲) انجام شد. ویژگی‌های فیزیکی و شیمیایی روتین مانند بافت، کربنات کلسیم معادل، pH، EC و نمونه‌های خاک ابتدا هوا خشک و پس از گذراندن از الک دو میلی‌متری مطابق با روش‌های استاندارد (مؤسسه تحقیقات خاک و آب، ۱۳۷۲ و ۱۳۷۶) اندازه‌گیری شد. سپس بر پایه داده‌های آزمایشگاهی و اطلاعات صحرایی، پدان‌های مطالعه شده مطابق با سیستم مرجع جهانی^۳ (FAO, 2015) رده‌بندی گردیدند. در این پژوهش، یک مدل رقومی ارتفاع و تصویر سنجنده OLI ماهواره لندست ۸ با تفکیک مکانی ۳۰ متر برای محاسبه و استخراج کوواریت‌های محیطی توپوگرافیک، آب‌شناختی و بازتابشی در محیط نرم افزار SAGA (Olaya, 2004) به کار برده شد. مشخصات کوواریت‌های محیطی مشتق‌شده از مدل رقومی ارتفاع و تصاویر ماهواره‌ای در جدول (۱) ارائه شده است. از آنجا که میزان اهمیت و تأثیر این کوواریت‌ها در هر منطقه متفاوت است، برای آگاهی از مؤثرترین این کوواریت‌های محیطی در منطقه مورد مطالعه نیاز به نوعی حساسیت‌سنجی داریم. در مدل جنگل تصادفی، نمایه میانگین کاهش جینی^۴ سنجه‌ای برای انتخاب کوواریت‌های محیطی مهم در فرآیند مدل‌سازی است.

جدول ۱. کوواریت‌های محیطی مورد استفاده در مدل‌سازی

ردیف	کوواریت	برابر فارسی	نوع کوواریت
۱	Elevation	ارتفاع	کمی
۲	Slope	تندی شیب	کمی
۳	Multi-Resolution Index of Valley Bottom Flatness (MRVBF)	نمایه همواری کف دره با درجه تفکیک متعدد	کمی
۴	Valley Depth	ژرفای دره	کمی
۵	Topographic Wetness Index (TWI)	نمایه خیزی توپوگرافیک	کمی
۶	Vertical Distance to Channel Network (VDCN)	فاصله عمودی تا شبکه آبراهه‌ها	کمی
۷	Convexity	کوژی	کمی
۸	Slope Length and Steepness Factor (LS Factor)	عامل تندی و درازای شیب	کمی
۹	Total Insolation (TI)	تابش دریافتی کل	کمی
۱۰	Multi-Resolution Ridge Top Flatness Index (MRRTF)	نمایه همواری بالای پشته با درجه تفکیک متعدد	کمی
۱۱	Difference Vegetation Index (DVI)	نمایه اختلاف پوشش گیاهی	کمی
۱۲	Normalized Difference Vegetation Index (NDVI)	نمایه اختلاف پوشش گیاهی نرمال‌شده	کمی
۱۳	Ratio Vegetation Index (RVI)	نمایه نسبت پوشش گیاهی	کمی
۱۴	Carbonate Index	نمایه کربنات	کمی
۱۵	Brightness Index	نمایه روشنایی	کمی
۱۶	Clay Index	نمایه رس	کمی

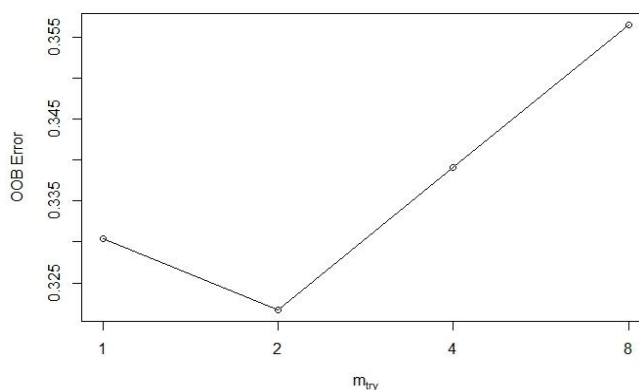
جنگل تصادفی گونه‌ای از روش‌های رایان‌آموختی است که شامل مجموعه‌ای از درختان رده‌بندی و رگرسیون^۵ (CART) تصادفی می‌باشد (Breiman, 2001). در این الگوریتم شمار زیادی درخت تصمیم‌گیری ساخته‌شده و جهت یک پیش‌بینی معین با هم ترکیب می‌شوند. برای پیش‌بینی

1 Global Positioning System
 2 Surface morphometry
 3 World Reference Base for Soil Resources (WRB)
 4 Mean Gini Decrease
 5 Classification and Regression Trees (CART)

پراکنش خاک‌ها در منطقه مطالعاتی، مقادیر کوواریت‌های محیطی (متغیر مستقل) و داده‌های مربوط به کلاس خاک (متغیر وابسته) در هر نقطه مشاهداتی به‌عنوان ورودی مدل وارد محیط نرم‌افزار (R-Studio 0.97.551) (R Development Core Team, 2013) گردید و با استفاده از بسته نرم‌افزاری جنگل تصادفی^۱ (Liaw and Wiener, 2002) مدل‌سازی انجام گرفت. برای راستی‌آزمایی عملکرد مدل جنگل تصادفی در پیش‌بینی کلاس‌های خاک از سنجه‌های^۲ (OOB Error)، صحت کاربر^۳ (UA)، صحت تولیدکننده^۴ (PA)، صحت عمومی نقشه^۵ (OA) و نمایه سازگاری کاپا^۶ (K) استفاده شد (Rossiter, 2014).

نتایج و بحث

در پیش‌بینی گروه‌های مرجع خاک توسط مدل جنگل تصادفی، کمترین مقدار OOB Error، ۰/۳۰ بود که با تعداد ۱۰۰۰ درخت و ۲ کوواریت در هر گره به‌دست آمد (شکل ۲، جدول ۲). همان‌طور که در جدول (۲) مشاهده می‌شود، پیش‌بینی در سطح گروه مرجع Calcisols دارای بیشترین و گروه‌های مرجع Cambisols و Kastanozems دارای کمترین صحت بود. فراوانی تعداد پدان‌ها در هر گروه مرجع و همچنین مؤثر بودن کوواریت‌های محیطی مورد استفاده در فرآیند مدل‌سازی، عامل‌های مهمی هستند که می‌توانند عملکرد مدل را به‌طور معنی‌داری تحت تأثیر قرار دهند. صحت بالای مدل در پیش‌بینی گروه مرجع Calcisols احتمالاً به دلیل فراوانی تعداد پدان‌های موجود در این گروه باشد که حدود ۳۳٪ مشاهدات را شامل می‌شود. Barthold و همکاران (۲۰۱۳) در پیش‌بینی گروه مرجع خاک با روش جنگل تصادفی مقدار OOB Error حدود ۵۱/۶٪ را گزارش کردند. Roecker و همکاران (۲۰۱۰) نیز در پیش‌بینی کلاس‌های خاک با روش جنگل تصادفی مقدار همین سنجه را ۵۱٪ محاسبه کردند. با وجود اینکه در پژوهش حاضر نسبت به سایر پژوهشگران از نقاط مشاهداتی کمتری استفاده شد اما نتایج این پژوهش در مقایسه با آن‌ها صحت بهتری را نشان می‌دهد. Pahlavan Rad و همکاران (۲۰۱۴) با استفاده از RF در به‌روزرسانی نقشه سری خاک‌های لسی شمال ایران، مقدار OOB Error را حدود ۵۲٪ گزارش کردند. با توجه به نتایج راستی‌آزمایی مدل جنگل تصادفی در پیش‌بینی گروه مرجع خاک بیشترین مقدار صحت کاربر و تولیدکننده به‌ترتیب برای گروه‌های مرجع Calcisols و Vertisols محاسبه شد. همچنین مقادیر صحت عمومی نقشه و نمایه سازگاری کاپا برای پیش‌بینی گروه مرجع خاک به‌ترتیب ۰/۷۰ و ۰/۵۷ به‌دست آمد (جدول ۳).



شکل ۲- نمودار تعیین تعداد بهینه کوواریت‌های محیطی به‌کاررفته در هر گره (mtry) برای پیش‌بینی گروه مرجع خاک

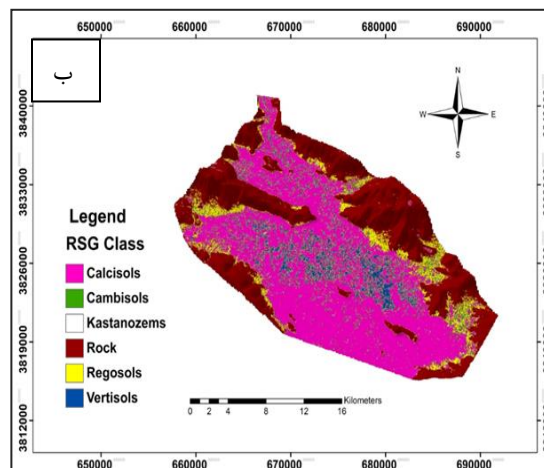
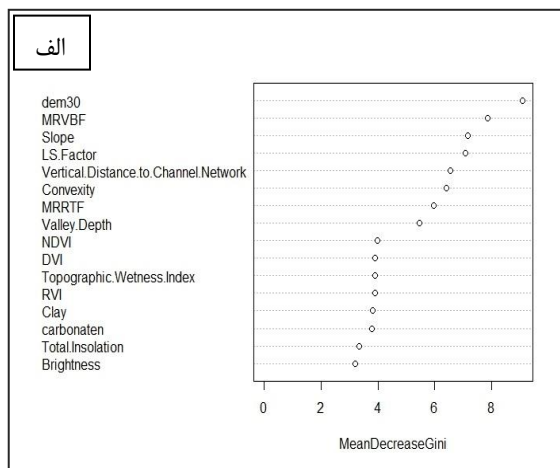
1 Random Forest package
 2 Out of bag Error (OOB Error)
 3 User's accuracy (UA)
 4 Producer's accuracy (PA)
 5 Overall map accuracy (OA)
 6 Kappa index of agreement (K)

مقدار صحت عمومی نقشه کاملاً قابل توجه بود، اما کاپا یک سنج قابل اعتماد آماری در برآورد صحت نقشه است (Rossiter, 2014) که مقدار آن در مقایسه با بسیاری از گزارش‌های ذکر شده در بالا قابل توجه بود و بر همین اساس می‌توان گفت به رغم گستردگی منطقه مطالعاتی و فاصله قابل توجه خاکرخ‌ها، عملکرد مدل رضایتبخش بود.

اهمیت کوواریت‌های محیطی به‌کاررفته در مدل با سنج میانگین کاهش جینی بررسی شد. هر بار که یک انشعاب از یک گره برای یک کوواریت در مدل جنگل تصادفی انجام می‌شود مقدار نمایه جینی در انشعابات حاصل برای آن کوواریت کاهش می‌یابد. مجموع کاهش نمایه جینی در تمام درختان برای یک کوواریت اهمیت آن را نشان می‌دهد. مقادیر این سنج برای هر یک از کوواریت‌های محیطی به‌کار برده شده در پیش‌بینی پراکنش گروه‌های مرجع خاک در منطقه محاسبه شد. همانطور که در شکل (۲) ملاحظه می‌شود از میان مهم‌ترین کوواریت‌ها در پیش‌بینی نقشه پراکنش گروه‌های مرجع خاک ارتفاع، نمایه همواری کف دره با درجه تفکیک متعدد، تندی شیب و فاکتور LS قابل اشاره هستند. این موضوع می‌تواند مؤید این مطلب باشد که در این منطقه، پستی و بلندی از مهم‌ترین فاکتورهای خاک‌سازی است و تأثیر به‌سزایی در متمایز کردن کلاس‌های خاک از هم دارد. Jafari و همکاران (۲۰۱۲) نیز در منطقه خشک اقدام به پیش‌بینی کلاس‌های خاک کردند و نمایه خیزی و همواری کف دره با درجه تفکیک متعدد را به‌عنوان کوواریت‌های مهم در مدل‌سازی معرفی نمودند. افزون بر آنچه گفته شد، با توجه به مشاهدات صحرایی از نظر موقعیت ژئومورفیک قرارگیری کلاسهای مختلف خاک می‌توان گفت نقشه پیش‌بینی (شکل ۳) هماهنگی قابل قبولی با واقعیت داشت.

جدول ۲- راستی‌آزمایی پیش‌بینی گروه مرجع خاک با استفاده از مدل جنگل تصادفی

گروه مرجع خاک	صحت کاربر	صحت تولیدکننده	صحت عمومی	کاپا
Calcisols	۰/۸۷	۰/۶۴		
Cambisols	۰	۰		
Kastanozems	۰	۰	۰/۷۰	۰/۵۷
Regosols	۰/۳۴	۰/۳۴		
Vertisols	۰/۳۹	۰/۷۰		



شکل ۳- الف) اهمیت نسبی کوواریت‌های محیطی در پیش‌بینی گروه مرجع خاک و ب) نقشه پیش‌بینی پراکنش مکانی گروه‌های مرجع خاک که بر DEM سه بعدی منطقه نشانده شده است.



مقدار سنجه‌های ارزیابی عملکرد مدل نشان داد که جنگل تصادفی توانست با به‌کارگیری کوواریت‌های محیطی، کلاس‌های خاک سیستم مرجع جهانی را با صحت مکانی قابل قبولی پیش‌بینی کند. از موثرترین کوواریت‌ها در منطقه می‌توان به ارتفاع، نمایه همواری کف دره با درجه تفکیک متعدد و تندی شیب و فاکتور LS اشاره نمود. ضمناً فراوانی مشاهدات در هر گروه مرجع و همچنین اندازه تأثیر کوواریت‌های محیطی در فرآیند مدل‌سازی، عامل‌های مهمی هستند که می‌توانند عملکرد جنگل تصادفی را به‌طور معنی‌داری تحت تأثیر قرار دهند. نقشه پیش‌بینی‌شده با الگوی پراکنش خاک‌ها در سطح منطقه که هنگام مطالعات صحرایی دیده شد، هماهنگی بالایی نشان داد.

منابع

گزارش مطالعات خاک‌شناسی نیمه‌تفصیلی دقیق طرح سد گاوشان- دشت میان‌دریند باختران. ۱۳۶۸. شرکت سهامی آب منطقه‌ای غرب. مؤسسه تحقیقات خاک و آب. ۱۳۷۲. شرح روش‌های تجزیه شیمیایی خاک (جلد اول). نشریه شماره ۸۹۳. سازمان تحقیقات، آموزش و ترویج کشاورزی. وزارت کشاورزی.

مؤسسه تحقیقات خاک و آب. ۱۳۷۶. شرح روش‌های تجزیه شیمیایی خاک (جلد دوم). نشریه شماره ۱۰۲۴. سازمان تحقیقات، آموزش و ترویج کشاورزی. وزارت کشاورزی.

- Barthold, F., Wiesmeier, M., Breuer, L., Frede, H.-G., Wu, J., & Blank, F. (2013). Land use and climate control the spatial distribution of soil types in the grasslands of Inner Mongolia. *Journal of Arid Environments*, 88, 194-205.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Brungard, C. W., Boettinger, J. L., Duniway, M. C., Wills, S. A., & Edwards, T. C. (2015). Machine learning for predicting soil classes in three semi-arid landscapes. *Geoderma*, 239, 68-83.
- Grunwald, S. (2006). *Environmental soil-landscape modeling: Geographic information technologies and pedometrics*: CRC/Taylor & Francis
- Hartemink, S. Kienast-Brown (Eds.), *Digital Soil Mapping: Bridging Research, Environmental Application and Operation*. Progress in Soil Science. Springer, New York.
- IUSS Working Group WRB. G. 2015. World Reference Base for Soil Resources 2014, update 2015 International soil classification system for naming soils and creating legends for soil maps. World Soil Resources Reports No. 106. FAO, Rome.
- Jafari, A., Finke, P., Vande Wauw, J., Ayoubi, S., & 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 of Soil science*, 63(2), 284-298.
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, 2(3), 18-22.
- Olaya, V. (2004). *A gentle introduction to SAGA GIS*. The SAGA User Group eV, Gottingen, Germany, 208.
- Reza Pahlavan Rad, M., Toomanian, N., Khormali, F., Brungard, C. W., Bayram Komaki, C., & 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(97-106), 232-234.
- Roecker, S.M., Howell, D.W., Haydu-Houdeshell, C.A., and Blinn, C. 2010. A qualitative comparison of conventional soil survey and digital soil mapping approaches, P 369-384. In: J.L. Boettinger, D.W. Howell, A.C. Moore, E.A. Rossiter, D. G. 2014. Statistical methods for accuracy assessment of classified thematic maps. Dept. of Earth systems Analysis. University of Twente, Faculty of Geo-Information Sciences & Earth observation (ITC), Enschede, The Netherlands.
- Schoeneberger, P. J., D.A. Wysocki, E.C. Benham, and Soil Survey Staff. 2012. (2012). Field book for describing and sampling soils, Version 3.0. Lincoln: Natural Resources Conservation Service, National Soil Survey Center.
- Team, R. (2013). R Development Core Team. *RA Lang Environ Stat Comput*, 55, 275-286.
- Witten, I. H., Frank, E. Hall, M. A. (2011). *Data Mining: Practical machine learning tools and techniques*: Morgan Kaufmann.



16th Iranian Soil Science Congress

University of Zanjan, Iran, August 27-29, 2019



Topic for submission: Pedometry and Soil Evaluation

A study on random forest performance in digital soil mapping of a semi-arid Mediterranean area

Bahmani, A.¹, Shekaari, P.^{2*}

¹ M. Sc. graduate, Department of Soil Science and Engineering, Faculty of Agriculture, Razi University, Kermanshah, Iran.

² Assistant Prof., Department of Soil Science and Engineering, Faculty of Agriculture, Razi University, Kermanshah, Iran.

Abstract

Soil resource inventories contain the essential information required in sustainable land management. Digital soil mapping methods are devised to make high-resolution, low cost, and readily updatable soil maps. Machine learning techniques are increasingly adopted in this way. In the present work, Random Forest (RF) algorithm used to predict soil classes across a 56,000 ha. landscape of Miandarband region, in Kermanshah province. To ensure proper coverage, an equidistant grid of 122 points designed through simple systematic sampling method. All pedons were described, analysed and classified according to WRB systems, which led to identification of 5 RSGs. Several environmental covariates calculated from 30-meter resolution DEM and OLI sensor image of Landsat 8. The covariates used as predictors in RF model to predict and map, distribution of soil classes over the study area. As a measure of prediction accuracy, OOB error, overall accuracy (OA) and kappa index of agreement (K) values of 0.30, 0.70 and 0.57 calculated for RSGs of WRB, respectively. Validation metrics confirmed considerable capability of RF in prediction of soil distribution pattern over the study area.

Keywords: Machine Learning, Miandarband, Environmental Covariates.

* Corresponding author Email: pshekaari@gmail.com