شانزدهمین کنگره علوم خاک ایران

دانشگاه زنجان، ۵ تا ۷ شهریور ۱۳۹۸

محور مقاله: پدومتری و ارزیابی خاک‌ها

کاربست الگوریتم K-نزدیکترین همسایه در بازشناسی الگوی گروه‌های مرجع WRB در منطقه سنجابی کرمانشاه

نوا کیانیان[۱]، پرویز شکاری[۲*]

[۱] دانش‌آموخته کارشناسی ارشد گروه علوم و مهندسی خاکخاک، دانشکده کشاورزی، پردیس کشاورزی و منابع طبیعی، دانشگاه رازی، کرمانشاه.

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

چکیده

کاربست روش‌های نوین مطالعات خاکشناسی برای دستیابی به اطلاعات کمی و آسان به‌روزشونده خاک در راستای برنامه‌ریزی و مدیریت پایدار سرزمین ضروری به نظر می‌رسد. Kنزدیک‌ترین همسایه (KNN) یکی از شناخته‌شده‌ترین مدل‌های ناپارامتری در دسته‌بندی و مدل‌کردن داده‌ها است. برای ارزیابی عملکرد KNN در تهیه نقشه پیش‌بینانه خاک، بخشی از دشت سنجابی به مساحت حدود ۴۸ هزار هکتار در استان کرمانشاه برگزیده شد. تعداد ۹۳ پدان مطالعه و در سطح اول سیستم WRB رده‌بندی شد. ارزیابی پیش‌بینی‌ها به صحت عمومی ۰/۵۲ و ضریب سازگاری کاپا ۰/۳۲ انجامید. عملکرد KNN به فراوانی مشاهدات حساسیت داشت به همین دلیل، گروه مرجع چیره در منطقه (Calcisols) را با صحت خوب، اما گروه‌های با فراوانی کمتر (Kastanozems, Vertisols, Leptosols) را ضعیف پیش‌بینی نمود. این ضعف در همه روش‌های رایان‌آموختی دیده می‌شود. یکی از راه حل‌های این مشکل طراحی و کاربست روشی برای نمونه‌گیری است که توان سازگارشدن از پیش با پراکنش خاک در هر منطقه معین مطالعاتی را داشته باشد. در مجموع، دست‌بالای عملکرد KNN در منطقه مطالعاتی میانه بود و در یک نمونه معین، نشان از نیاز به ترکیب این مدل با دیگر روش‌های رایان‌آموختی به عنوان راه دیگر بهبود عملکرد است.

کلمات کلیدی: رایان‌آموختی، الگوی پراکنش خاک، نقشه برداری رقومی خاک.

مقدمه

نقشه خاک منبع با ارزشی از اطلاعات برای برنامه‌ریزی و ارزیابی کاربری و بسیاری از فعالیت‌های مرتبط با خاک در مقیاس ملی و محلی است. شناسایی یا نقشه‌برداری خاک به عنوان روشی برای تعیین الگوی پراکنش خاک، توصیف و نمایش آن به شکل قابل فهم و تفسیر برای کاربران مختلف، پایه و اساس اطلاعات خاک برای مدل‌سازی‌های محیطی است (Zhu و همکاران ۲۰۰۱). به طور کلی، روش‌های سنتی شناسایی خاک تجربی بوده و براساس مدل ذهنی نقشه‌بردار از روابط خاک-زمین‌ریخت تهیه استوار هستند. اما خاک متأثر از عوامل محیطی است که بر تشکیل و تکامل آن تأثیر می‌گذارند، در نتیجه همان‌طور که تغییر در عوامل محیطی[۱] به صورت پیوسته و تدریجی صورت می‌گیرد، ویژگی‌های خاک نیز در پهنه سرزمین در مقیاس‌های مختلف، تغییراتی تدریجی و پیوسته دارند. بنابراین، پیدایش خاک در یک منطقه را نمی‌توان حاصل اثر مطلق یک فرآیند خاک‌شناختی دانست. اما روش‌شناسی نقشه برداری سنتی خاک بر این باور بنا شده است. به همین دلیل ، برخی پژوهشگران از جمله Webو Lilburne (۲۰۰۵) اعتقاد دارند که عدم قطعیت زیاد در نقشه‌های سنتی خاک[۲]، به شدت از ارزش عملی کاربرد آنها کاسته است . از سوی دیگر، مطالعات میدانی و آزمایشگاهی رایج خاک به علت زمان‌بر بودن و هزینه‌های زیاد آنها، دیگر مقرون‌به‌صرفه نیستند.

از این رو، با توجه به محدودیت‌های موجود برای نقشه‌برداری چندگوشه‌ای خاک و با افزایش روزافزون گرایش به مطالعات محیطی، مطالعات شناسایی خاک‌ها در حال گذر از یک حالت کیفی و ذهنی به یک حالت کمی با درجه اعتماد بالاتر است (McBratney و همکاران ۲۰۰۳). یکی از روش‌های نوین ارائه شده در این ارتباط، نقشه‌برداری رقومی خاک[۳] (DSM) نام دارد. نقشه‌برداری رقومی خاک عبارت است از ایجاد و جمع‌آوری سیستم‌های اطلاعات مکانی خاک با استفاده از روش‌های صحرایی و آزمایشگاهی که با سیستم‌های استنتاج مکانی خاک همراه شده و برون‌داد آن را می‌توان به صورت نقشه‌های برآورد با فرمت رستری به دست آورد (Lagachery and McBratney, 2006). روش‌های رایان‌آموختی از زیرشاخه‌های داده‌کاوی است که با ایجاد همبستگی بین داده‌های استفاده شده در نقشه‌های چندگوشه‌ای و تلفیق آن با پارامترهای کمکی محیطی به گسترش مجموعه داده‌های آموزشی کمک می‌کند (McBratney و همکاران ۲۰۰۳). سپس بر اساس واقعیت‌های آموزش داده شده، یک تصویردر قالب مدل ایجاد

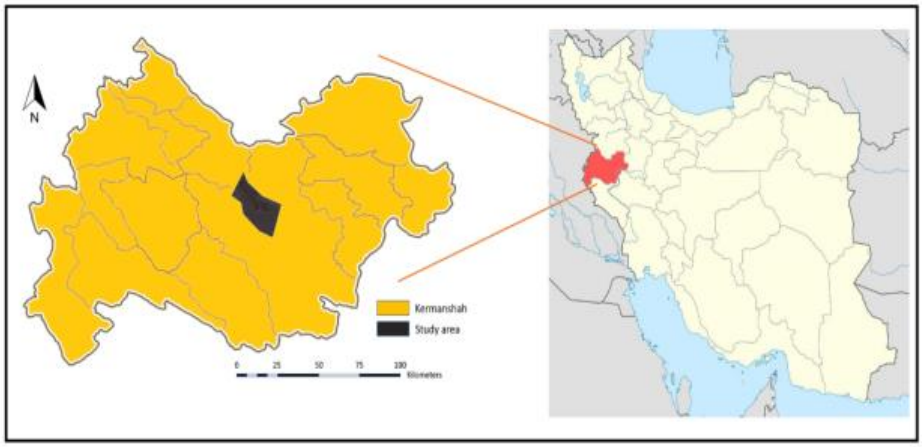
* ایمیل نویسنده مسئول: pshekaari@gmail.com

[1] Environmental Covariates
[2] Traditional Soil maps
[3] Digital Soil mapping

می‌شود که جهت پیش‌بینی ویژگی‌ها و کلاس‌های خاک در مکان‌هایی که اطلاعاتی از آن‌ها در دسترس نیست، به کار برده می‌شود. تکنیک کای نزدیکترین همسایه[1] (KNN) ، یکی از مهم‌ترین و توسعه‌یافته‌ترین رویکردهای ناپارامتریک است که برای نخستین بار در سال ۱۹۵۱ توسط Ficks و Hugges ارائه شد و از سال ۱۹۷۰ در بسیاری از برنامه‌های تخمین آماری و بازشناسی الگو مورد استفاده قرار گرفته است. به طور کلی KNN روشی برای طبقه‌بندی یک فرد در یک مجموعه است که این کار را بر اساس میزان نزدیکی ویژگی‌های آن با دیگر اعضای موجود در مجموعه انجام می‌دهد. در این روش تعداد بهینه‌ای از نمونه‌های موجود در یک جامعه که دارای شبیه‌ترین ویژگی‌ها به نمونه هدف یعنی K باشند، در نظر گرفته شده و سپس با تعیین فاصله‌ها و وزن‌دهی عناصر آموزش دهنده، جایابی و طبقه‌بندی عنصر هدف در این مجموعه از داده‌ها انجام می‌شود. اولین گام درKNN تعیین فاصله بین نمونه هدف با هر یک از داده‌های موجود در بانک داده است. استفاده از تابع فاصله اقلیدسی وزن‌دارشده[2] در انجام پیش‌بینی‌ها توسط KNN بسیار معمول است (Sorjamaa و همکاران ۲۰۰۵). استفاده از روش KNN در علوم زمین و خاک‌شناسی در ایران بسیار اندک گزارش شده است و در سطح جهانی نیز در مرحله مطالعه است. Nemes و همکاران (۲۰۰۹) به تفسیر توزیع اندازه ذرات خاک با استفاده از روش KNN پرداختند که در واقع یکی از اولین موارد استفاده این تکنیک در علوم خاک بوده است. Sharif و همکاران (۲۰۰۶) با استفاده از روش KNN شبیه‌سازی طرح‌های اقلیمی را برای ارزیابی آسیب‌پذیری یک حوضه آبریز در برابر رویدادهای شدید انجام دادند. در این پژوهش مهم‌ترین عوامل، دما، میزان بارش و میزان رواناب معرفی شد. آنها گزارش کردند که بارش‌های شدید و بی‌سابقه یا خشکسالی توسط این روش به آسانی قابل پیش‌بینی و شبیه‌سازی است. Nemes و همکاران (۲۰۰۹) با مقایسه کارآیی توابع انتقالی پارامتریک و تکنیک KNN، توانایی روش KNN در تخمین توابع هیدرولیکی خاک را در برخی خاک‌های ایالات متحده، بهتر از توابع پارامتریک ارزیابی نمودند. در مطالعه Heung و همکاران (۲۰۱۶) از ده روش رایان‌آموختی جهت کلاس‌بندی خاک در منطقه‌ای به مساحت ۵۴۷۲ کیلومترمربع استفاده شد . نتایج نشان داد که روش KNN و SVM جزو برترین روش‌ها می‌باشند. با توجه به اهمیت مناطق مهم کشاورزی در بخش نیمه‌خشک کشور تمرکز این پژوهش بر کاربست روش رایان‌آموختی Kنزدیکترین همسایه در پیش‌بینی الگوی پراکنش گروه‌های مرجع خاک در دشت سنجابی کرمانشاه بود.

مواد و روش‌ها

منطقه مطالعاتی با مساحت حدود ۴۸ هزار هکتار و میانگین ارتفاع ۱۳۳۰ متر بین عرض‌های جغرافیایی ۳۴ درجه و ۲۰ دقیقه تا ۳۴ درجه و ۳۱ دقیقه شمالی و طول‌های جغرافیایی ۴۶ درجه و ۴۴ دقیقه تا ۴۶ درجه و ۵۹ دقیقه شرقی قرار داشت (شکل ۱).



شکل ۱. موقعیت منطقه مطالعاتی در استان کرمانشاه.

از نظر زمین‌شناختی، منطقه مطالعاتی قسمتی از زون خرد شده و گسله زاگرس است بیشتر ارتفاعات حاشیه منطقه، مربوط به سازندهای آهکی است. موقعیت جغرافیایی ۱۰۰ نقطه به عنوان نقاط مشاهده‌ای در دشت سنجابی تعیین شد. مطالعه‌ی چند پدان که در معادن فعال قرار داشتند، ممکن نشد. در مجموع ۹۳ خاکرخ مشاهداتی تا عمق ۱۲۰ سانتی‌متری حفر، تشریح و از تمامی افق‌های ژنتیک نمونه‌برداری صورت گرفت. نمونه‌ها به به آزمایشگاه منتقل و پس از هواخشک شدن از الک ۲ میلی‌متری گذر داده شدند. پارامترهای فیزیکی و شیمیایی آنها به روش‌های استاندارد (Page, 1996 و

[1]. K Nearest Neighbors
[2]. Weighted Euclidean distance function

Klutet,1996) اندازه‌گیری شد. هر پدان بر اساس داده‌های آزمایشگاهی و صحرائی مطابق با سیستم مرجع جهانی (FAO, 2014) در سطح گروه‌های مرجع خاک رده‌بندی گردید. برای محاسبه متغیرهای محیطی توپوگرافیک و آب‌شناختی مدل رقومی ارتفاع منطقه با تفکیک مکانی ۳۰ متر به کار برده شد. دیگر شاخص‌های محیطی نیز با استفاده از تصویر ماهواره‌ای مربوط به سنجنده Landsat8 (OLI) به دست آمد. مراحل آموزش، بهینه‌سازی، راستی‌آزمایی و تهیه نقشه در محیط نرم افزار R-Studio انجام گردید. راستی‌آزمایی عملکرد مدل K نزدیکترین همسایه نیز بر پایه مقادیر محاسبه شده صحت عمومی[1]، صحت کاربر[2]، صحت تولیدکننده[3] و شاخص کاپا[4] (Lu و همکاران ۲۰۰۴) انجام شد.

نتایج و بحث

با توجه به نتایج به‌دست آمده از مطالعات صحرائی خاکرخ‌ها و تجزیه‌های آزمایشگاهی، پدان‌های مطالعه شده در منطقه مطالعاتی بر اساس سطح اول سیستم رده‌بندی مرجع جهانی خاک در ۶ گروه مرجع خاک قرار گرفت که نام و فراوانی آنها در جدول ۱ ارائه شده است.

جدول ۱. گروه‌های مرجع سیستم مرجع جهانی خاک (WRB) و فراوانی آن‌ها در منطقه مطالعاتی.

گروه مرجع خاک	Calcisols	Cambisols	Kastanozems	Leptosols	Regosols	Vertisols
فراوانی	۳۷	۱۴	۱۷	۴	۱۵	۶

در بیشتر پدان‌های مورد بررسی، درصد بالای کربنات کلسیم، سنگ بستر آهکی و مواد Calcaric مشاهده شد. به طور کلی میزان کربنات کلسیم معادل از سطح به عمق افزایش داشت. به نظر می‌رسد روند نامنظم میزان کربن آلی خاک با عمق، در برخی پدان‌ها، به دلیل دریافت آبرفت و اضافه شدن مواد جدید در طول زمان بر روی خاک‌های منطقه باشد.

در بیشتر پدان‌های مورد بررسی، درصد بالای کربنات کلسیم، سنگ بستر آهکی و مواد Calcaric مشاهده شد. به طور کلی میزان کربنات کلسیم معادل از سطح به عمق افزایش داشت. به نظر می‌رسد روند نامنظم میزان کربن آلی خاک با عمق، در برخی پدان‌ها، به دلیل دریافت آبرفت و اضافه شدن مواد جدید در طول زمان بر روی خاک‌های منطقه باشد.

در جدول ۲ می‌توان عملکرد مدل KNN را در پیش‌بینی کلاس‌های خاک را در منطقه‌ی مطالعاتی مشاهده کنید. صحت عمومی نقشه گروه مرجع خاک ۰/۵۲ محاسبه گردید. صحت ۰/۸۵ برای گروه مرجع Calcisols (جدول ۲) می‌تواند به دلیل فراوانی بالای مشاهدات این گروه باشد. این گروه مرجع حدود ۴۰٪ از مشاهدات صحرائی را در بر می‌گرفت. گروه‌های مرجع Kastanozems، Leptosols، Vertisols و رخنمون‌های سنگی با کمینه صحت توسط مدل KNN برآورد شدند. ضریب کاپای ۰/۳۲ و همچنین صحت کاربر و تولید کننده برای این گروه‌های مرجع این مطلب را تایید می‌کند. میزان خطا در Leptosols، Vertisols و رخنمون‌های سنگی احتمالاً به دلیل تعداد پایین مشاهدات و نقص در مرحله‌ی آموزش مدل است. اما خطای برآورد مدل برای گروه مرجع Kastanozems می‌تواند به دلیل چیرگی کربنات کلسیم و نوع کاربری منطقه باشد.

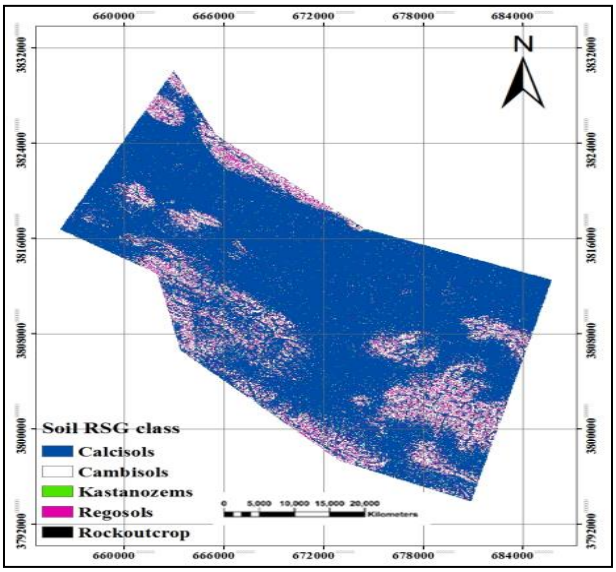
جدول ۲. نتایج راستی‌آزمایی پیش‌بینی گروه مرجع خاک در منطقه مطالعاتی با مدل K نزدیک‌ترین همسایه.

گروه مرجع خاک	Calcisols	Cambisols	Kastanozems	Leptosols	Regosols	Vertisols
صحت کاربر	۰/۵	۰/۵	۰	۰	۰/۶۶	۰
صحت تولیدکننده	۰/۵	۰/۵	۰	۰	۱	۰

نقشه پیش‌بینی پراکنش مکانی گروه‌های مرجع خاک که در شکل ۲ ارائه شده است. گروه مرجع Calcisols بیش‌ترین پوشش را در منطقه مطالعاتی داشت. این مناطق دارای ارتفاع کم، نمایه خیسی و همواری کف دره بالا بوده و در واحد زمین‌ریخت‌شناسی دشت آبرفتی رودخانه‌ای و دامنه‌ای قرار گرفته‌اند. گروه مرجع Cambisols در نقاطی با ارتفاع نسبتا کم مانند دشت‌های دامنه‌ای که مستعد فرسایش‌هستند، به درستی پیش‌بینی شده است. مدل KNN گروه مرجع Cambisols را در بعضی مناطق مانند قسمت‌های مرتفع جنوب تا جنوب‌غرب دشت به اشتباه پیش‌بینی نموده است. گروه مرجع Regosols به درستی در بخش‌های کوهستانی با شیب بین۳۰ تا ۶۰ درجه و پوشش خاکی کم‌عمق برآورد شده است. گروه مرجع Kastanozems کمترین مساحت نقشه پیش‌بینی شده را تشکیل داده است. می‌توان گفت مدلKNN نتوانسته است گروه مرجع Kastanozems را به

[1] Overall accuracy
[2] User's accuracy
[3] Producer's reliability
[4] Kappa Index

درستی برآورد کند و آنها را در گروه مرجع Calcisols پیش‌بینی نموده است. گروه مرجع Vertisols خاک‌هایی با بافت سنگین هستند که عمدتاً در نواحی پست با شیب کم و نمایه هموارى کف دره زیاد شکل می‌گیرند. مدل این گروه مرجع را به اشتباه در گروه مرجع Calcisols طبقه‌بندی کرده است. همچنین مدل برآورد Leptosols را با خطای ۱۰۰ درصد پیش‌بینی نموده است. مانند دیگر الگوریتم‌های رایان‌آموختی شمار نقاط مشاهداتی در هر یک از سطوح رده‌بندی رابطه مستقیمی با صحت پیش‌بینی مدل دارد.



شکل ۲. نقشه پیش‌بینی پراکنش مکانی گروه‌های مرجع خاک در منطقه مطالعاتی با مدل کای نزدیک‌ترین همسایه

نتیجه‌گیری

با توجه به موارد ذکر شده ، می‌توان گفت که هماهنگی میان نقشه پیش‌بینی شده در سطح گروه مرجع با مشاهدات صحرایی الگوی پراکنش خاک‌ها در منطقه حداکثر در اندازه متوسط بود. در نتیجه مدل K نزدیک‌ترین همسایه برای پیش‌بینی کلاس‌های خاک موجود در منطقه در سطح گروه مرجع عملکردی در همان سطح داشت. با توجه به تغییرات تدریجی که زیر اثر چیرگی کربنات کلسیم و نبود شرایط اقلیمی مناسب برای ایجاد تفاوت‌های چشم‌گیر در خاک‌ها در منطقه مطالعاتی هستند، انتخاب سیستم نمونه‌برداری، شمار نمونه‌ها و همچنین مقدار بهینه K در عملکرد روش K نزدیکترین همسایه، بسیار حائز اهمیت است. اساس روش KNN بر مبنای شباهت داده‌ها است، بنابراین کاربست آن به عنوان یک روش ساده و کارآمد در علوم خاک شایان بررسی است. کاربست چنین الگوریتم‌هایی که به جای برازش معادلات به داده‌ها تشخیص الگو انجام می‌دهند می‌تواند در بهبود و تسریع استخراج دانش از داده‌ها سودمند واقع شود.

منابع

Fix, E., Hodges, J. 1951. Dicriminatory analysis, non parametric discrimination: consistency properties. Technical report 4, USA, School of aviation medicine Randolph field texas.

Heung, B., Ho, H. C., Zhang, J., Knudby, A., Bulmer, C. E., & Schmidt, M. G. (2016). An overview and comparison of machine-learning TECHNIQUES for classification purposes in digital soil mapping. Geoderma, 265, 62-77.

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.

Lagacherie, P., McBratney, A.B., 2006. Chapter 1 spatial soil information systems and spatial soil inference systems: perspectives for digital soil mapping. Developments in Soil Science. 3 –22.

Klute, A., 1996. Methods of soil analysis. Part 1. Chemical and mineralogical properties. Monograph No. 9. 2nd. ed. ASA monograph No. 9. SSSA, Madison, WI, USA.

Lu, D., Mausel, P., Brondizio, E. and Moran, E. 2004. Change detection techniques. International Journal of remote sensing, 25(12), 2365-241.

McBratney A.B., Mendonça Santos M.L., and Minasny B. 2003. On digital soil mapping. Geoderma, 117:3-52.

Nemes, A., D.J. Timlin, A. Pachepsky Ya and W.J. Rawls. 2009. Evaluation of the pedotransfer functions for their Applicability at the U.S. National Scale. Soil Sci. Soc. Amer. J. 73:1638-1645.

Page, A.L., 1996. Methods of soil analysis. Part 2. Chemical and mineralogical properties. Monograph No. 9 . 2nd. ed. ASA monograph No. 9. SSSA, Madison, WI, USA.

Sharif, M, and Burn D.H. 2006.Simulating climate change scenarios using an improved K-nearst neighbor model.Journal of Hydrology 325,179-196.

Sorjamaa, A., Reyhani, N., and Lendasse, A. (2005). Input and structure selection for K-NN approximator. 8th. International Conference on Artificial Neural Networks, Lecture Notes in Computer Science Springer. IWANN. Berlin . 958-992.

Webb, T. H. and Lilburne, L. R. 2005. Concequence of soil map unit uncertainty on environmental risk assessment. Australian journal of soil research, 43: 119-126.

Zhu, A. B., Hudson, J. Burt, K. Lubich and D. Simonson. 2001. Soil mapping using GIS, expert knowledge, and fuzzy logic, Soil Sci. Soc. Amer. J. 65: 1463-1472.

16th Iranian Soil Science Congress

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

Topic for submission: Pedometery and Soil Evaluation

Using K-nearest neighbor algorithm for pattern recognition of WRB RSGs in Sandjabi area of Kermanshah

N., Kianian[1], P., Shekaari[2*]

[1] MSc. 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

Applying new methods of soil survey to obtain quantitative information and easily updatable maps seems necessary in the direction of sustainable land management. K-nearest neighbor (KNN) is popular a non-parametric algorithm for classifying bulks of data. To evaluate KNN performance in predictive soil mapping, a 48,000 ha. extent of Sandjabi area in Kermanshah province selected. A number of 93 pedons described, sampled, and classified as RSGs of WRB. Assessing model predictions led to overall map accuracy (OA) of 0.52 and kappa index of agreement (K) of 0.32. KNN showed sensitivity to RSGs frequency, since dominant RSG (Calcisols) predicted accurately, while that was not the case with less frequent RSGs (Kastanozems, Vertisols, and Leptosols). This is the problem with every machine learning method, since there is no sampling method to guarantee sufficient frequency of observations in every soil class, a priori. Generally, KNN performed moderate at its best. Capability of the algorithm could be probably enhanced through be combined with other machine learning methods.

Keywords: machine learning, K-nearest neighbors, soil pattern, digital soil mapping.

[*] Corresponding author, Email: pshekaari@gmail.com