محور مقاله: فناوریهای نوین در علوم خاک

مقایسه برون‌یابی مکانی گروه بزرگ خاک با استفاده از روش‌های طبقه بندی درخت تصمیم و رگرسیون لاجستیک چندجمله ای در منطقه آبیک، قزوین

مهرناز نیستانی*[۱]، فریدون سرمدیان[۲]، اعظم جعفری[۳]، علی کشاورزی[۱]

[۱] دانشجوی دکتری دانشگاه تهران، دانشکده مهندسی علوم خاک، پردیس کشاورزی و منابع طبیعی دانشگاه تهران

[۲] استاد دانشکده مهندسی علوم خاک، پردیس کشاورزی و منابع طبیعی دانشگاه تهران

[۳]استادیار بخش علوم و مهندسی خاک دانشکده کشاورزی دانشگاه شهید باهنر کرمان

[۴]استادیار دانشکده مهندسی علوم خاک، پردیس کشاورزی و منابع طبیعی دانشگاه تهران

چکیده

نقشه برداری رقومی روش مناسبی را برای دستیابی به نقشه خصوصیات و کلاس‌های خاک در نقاط فاقد اطلاعات فراهم آورده‌است. هدف از این مطالعه انجام برون‌یابی و تعیین کارایی این روش با استفاده از مقایسه دو الگوریتم پیش بینی کننده درخت تصمیم و رگرسیون لاجستیک چندجمله ای در مناطق با اطلاعات محدود می‌باشد. برای این منظور منطقه‌ای به وسعت ۲۵۱۹۷ هکتار تحت عنوان منطقه مرجع جهت مدلسازی اولیه برای پیش‌بینی گروه بزرگ خاک انتخاب گردید. سپس مدل‌های استخراج شده به جهت پیش‌بینی گروه بزرگ خاک‌های شاهد در منطقه پذیرنده فاقد اطلاعات استفاده شد. دقت برون‌یابی از این مدل با استفاده از مقایسه خاک های پیش‌بینی شده با خاک‌های شاهد و تشکیل ماتریس ابهام تعیین گردید. نتایج برون‌یابی این مدل‌ها در منطقه پذیرنده برای مدل درخت تصمیم دقت کلی ۶۲٪ و مدل رگرسیون لاجستیک چندجمله ای ۵۲٪ نشان می‌دهد که این دقت در هر دو الگوریتم درحد متوسط قابل قبول می‌باشد اگرچه که کارایی مدل درخت تصمیم بالاتر از رگرسیون لاجستیک چندجمله‌ای می باشد. لذا از این روش می‌توان به عنوان یک روشی سریع و کم‌هزینه در پیش‌بینی الگوی کلاس خاک در مناطق بااطلاعات محدود استفاده کرد.

کلمات کلیدی: نقشه برداری رقومی خاک، مدلسازی، برون یابی درخت تصمیم، رگرسیون لاجستیک چندجمله ای.

مقدمه

نقشه‌برداری رقومی خاک به‌طور گسترد ای امروزه در سطح جهانی برای پهنه‌بندی خصوصیات و کلاس‌های خاک استفاده می‌شود (Arrouays et al., 2014). نقشه‌برداری رقومی به‌دلیل تولید نقشه‌های با قابلیت اعتماد بالا برای مدیریت پایدار سرزمین توسط FAO تائیدیه دریافت کرده‌اند. اگرچه که این روش نقشه‌بردرای امروزه به‌صورت اجرایی است اما نکات ناشناخته برای نمایش بهتر فرایندهای الگوی سرزمین و خاک وجود دارد. این تکنیک نقشه‌برداری توسط مک برتنی در سال ۲۰۰۳ به‌صورت فرمول ارائه شده‌است. نقشه‌برداری رقومی فرمول پیشرفته‌تری از مدل CLORPT است که توسط ینی در سال ۱۹۴۱ فرموله شده است و فاکتورهای تشکیل دهنده خاک را نشان می‌دهد. به علت محدودیت در منابع و عدم دسترسی به برخی منابع اطلاعاتی در بسیاری از مناطق جهان، به‌دست آوردن اطلاعات خاک دشوار است. از طرفی مهمترین هدف نقشه‌برداری رقومی خاک تهیه نقشه با استفاده از متغیرهای کم هزینه برای مناطقی است که اطلاعات خاکی از آن‌ها در دسترس نمی‌باشد (Abbaszade Afshar et al., 2018). برای دستیابی به این هدف پیشنهاد می‌شود که مدل در منطقه‌ای که اطلاعات خاک در آن به تفصیل در دسترس است ایجاد شود و از مدل ساخته شده برای مناطق فاقد اطلاعات استفاده شود (Zeraatpisheh et al., 2017). بنابراین می‌توان این مطلب را برداشت نمود که مدل منتج شده برای پیش‌بینی خصوصیات و یا کلاس خاک درصورتی که فرایندهای خاکشناسی، هیدرولوژی و ژئومورفولوژی مشابهی رخ دهد و یا اگر توزیع مکانی متغیرهای محیطی در دو منطقه یکسان باشد، کارایی خواهد داشت (Ten Caten et al., 2011; Legros, 2006; Minasny et al.,

* ایمیل نویسنده مسئول: neyestani_mehr@yahoo.com

2010) به عبارت دیگر، اگر مدلی برای منطقه‌ای ایجاد شد این مدل می‌تواند برای دیگر مناطق مشابه با داشتن تعداد کم ویا بدون نیاز به داشتن متغیرهای کمکی و نمونه‌های خاک، برازش داده شود از این روش تحت عنوان برون‌یابی یاد می‌شود. بنابراین پروسه برون‌یابی می‌تواند به عنوان ابزاری مفید در کاهش هزینه و زمان عمل نماید. برون یابی و یا با نام دیگر آن هوموسول در سال ۲۰۱۰ توسط مالاوان و همکاران ارائه شده‌است. برون‌یابی متغیرهای محیطی قبلاً در شناسایی اقلیم‌های مشابه در جهان برای پیش بینی محصول استفاده شده‌است. به‌منظور برون‌یابی مناطق مشابه لازم است علاوه بر اقلیم، خصوصیات زمین شناسی، پستی و بلندی نیز دارای تشابه باشند که برای این منظور از شاخص‌های تشابه برای تعیین تشابه دو منطقه استفاده می‌شود. در برون یابی نکته قابل توجه این است که هرچه منطقه پذیرنده به مناطق مرجع تشابه بیش‌تری داشته باشد، برون‌یابی در این مناطق دقت بالاتری خواهد داشت و نتایج حاصل از منطقه مرجع با اطمینان بیش‌تری قابل برازش میباشند. عباس‌زاده افشار و همکاران (۲۰۱۸) با استفاده از برون‌یابی از منطقه مرجع دقت کلی ۵۹٪ و کاپا۰/۴۷ را برای استفاده از مدل رگرسیون لاجستیک در سطح گروه بزرگ خاک ارائه داده‌اند. آرودا و همکاران (۲۰۱۶) با استفاده از روش شبکه هوشمند عصبی اطلاعات حاصل از مدلسازی را در یک منطقه کوچک به مناطق اطراف پیرامونی برای تعیین روابط سیمای اراضی استفاده نمودند. نتایج برون یابی مطالعات آن‌ها دقت کلی ۸۳٪ و کاپای ۰/۶۹ را نشان داد که دستیابی به این اطلاعات همراه با صرف هزینه و زمان کم‌تر بوده‌است. از آن‌جایی که بیشتر مناطق در ایران فاقد نقشه‌های اطلاعاتی مربوط به خاک می‌باشند و از طرفی دسترسی به آن‌ها برای دستیابی به اطلاعات خاک دشوار است، تحت چنین شرایطی برون‌یابی می‌تواند به عنوان روشی کارامد در جهت پیش‌بینی کلاس و خصوصیات خاک به‌کار رود (Abbaszadeh Afshar et al., 2018). هدف ما در این مطالعه ارزیابی و مقایسه روش برون‌یابی برای پیش‌بینی کلاس خاک در منطقه پذیرنده با استفاده از مدل به‌دست آمده از روش درخت تصمیم و رگرسیون لاجستیک چندجمله‌ای در سطح گروه بزرگ منطقه مرجع می‌باشد.

مواد و روش‌ها

وسعت و منطقه مطالعاتی

منطقه مورد مطالعه با وسعت ۶۲۷۴۰ هکتار در شهرستان آبیک از توابع استان قزوین واقع شده‌است (شکل ۱. الف). به منظور تعیین اقلیم منطقه از داده‌های هواشناسی دوره آماری ۲۴ ساله ایستگاه‌های کلیماتولوژی نیروگاه شهید رجائی و باغ کوثر و کرج استفاده گردید. میانگین بارندگی سالیانه در حوزه مطالعاتی ۳۱۰ میلی‌متر بوده و رژیم بارش، مدیترانه‌ای است. به این مفهوم که فصل بارندگی منطبق بر نیمه سرد سال و فصل خشک منطبق بر تابستان است. میانگین سالانه دمای روزانه در حوزه مورد نظر از کمتر از ۵ درجه سانتی‌گراد در مناطق مرتفع تا بیش از ۱۷/۵ درجه سانتی‌گراد در بخش‌های جنوبی متغیر است. متوسط دمای سالیانه منطقه ۱۴/۳ درجه سانتی‌گراد می‌باشد. رژیم حرارتی خاک ترمیک و رژیم رطوبتی خاک در منطقه مورد مطالعه در دامنه زریک خشک و اریدیک ضعیف تعیین گردید .

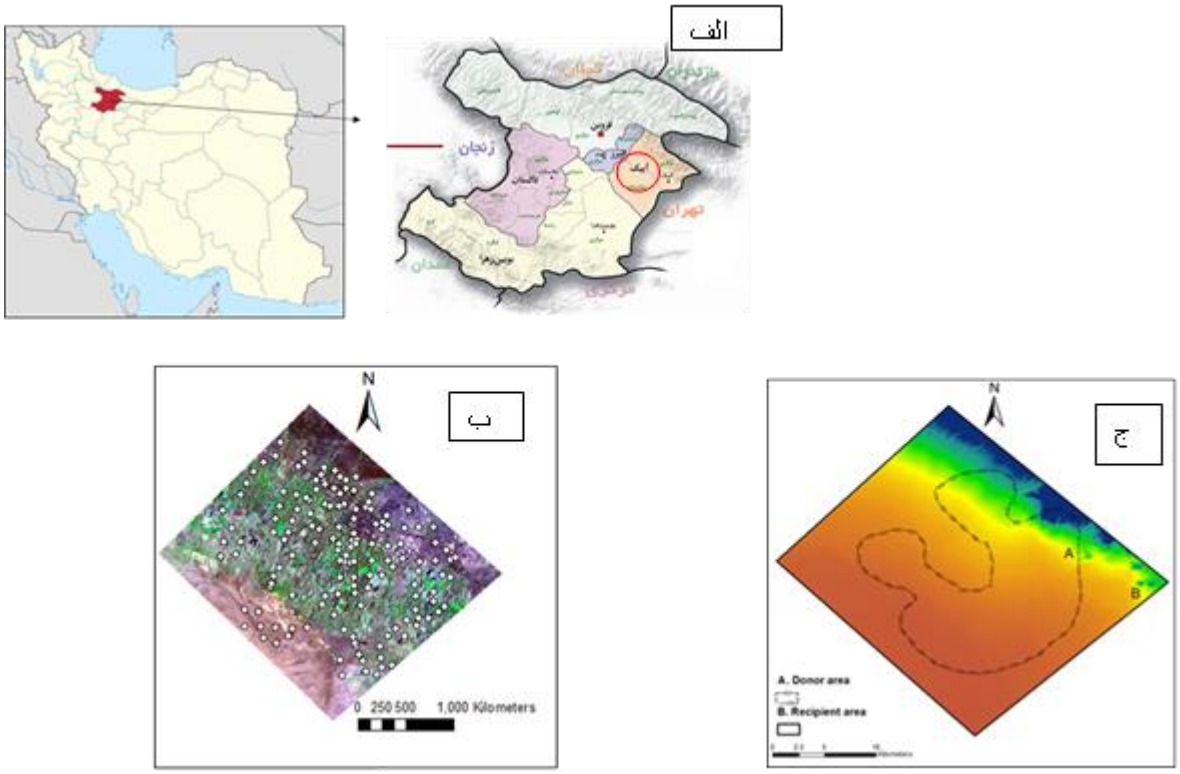
طرح نمونه‌برداری و تشریح پروفیل

الگوی حفر نمونه‌ها به‌صورت شبکه‌بندی نامنظم می‌باشد به‌نحوی که شبکه‌هایی با فواصل ۱۷۵۰ متر ایجاد می‌شود و سپس با ابزار ET Geowizard در نرم‌افزار آرک جی‌ای‌اس در هر شبکه یک نقطه به صورت تصادفی انتخاب می‌شود. با این روش ۱۶۰ نمونه تصادفی حفر و نمونه‌برداری شد و سپس بااستفاده از کلید تاکسونومی(Soil Survey Staff, 2014) تا سطح گروه بزرگ طبقه بندی شدند که شامل ۱۰ گروه بزرگ که عبارتند از زراورتنتز، کلسیزرپتز، هاپلوزرپتز، توریفلوونتز، هاپلوکلسیدز، کلسی جیپسیدز، هاپلوجیپسیدز، ناترجیپسیدز، هاپلوسالیدز، اکویی سالیدز می‌باشد. شکل ۱. ب محل نمونه‌های خاک را در کل منطقه نمایش می‌دهد.

متغیرهای کمکی:

برای آموزش مدل در منطقه مرجع نیاز به وارد کردن متغیرهای محیطی است تا بتواند به پیش‌بینی کلاس خاک بپردازد. به جهت به‌دست آوردن متغیرهای کمکی مربوط به توپوگرافی از مدل رقومی ارتفاعی با رزلوشن ۱۲/۵ متر استفاده شد و همچنین از تصاویر ماهواره‌ای لندست ۸ برای نشان دادن پوشش گیاهی و مواد مادری استفاده می‌شود که با استفاده از نرم افزار ساگا جی‌آی‌اس متغیرهای مربوطه از آن استخراج گردید. تمامی این متغیرها به سیستم مختصات جهانی WGS 84 UTM 39N تبدیل و با استفاده از نرم‌افزار ساگا مقدار تمامی متغیرها در جدولی برای هر نقطه مشاهده‌ای وارد شدند تا برای مدلسازی در نرم‌افزار R (R. Development Core Team, 2013) مورد استفاده قرار گیرند.

تکنیک های داده کاوی مثل رگرسیون درختی به میزان زیادی برای نقشه برداری خصوصیات خاک استفاده می شوند. در این روش به‌جای این‌که مدل را به تمامی داده‌ها برازش دهند، با استفاده از یک ساختار درختی و تقسیم نمودن داده‌ها، به‌صورت بازگشتی به تعدادی گروه ایجاد می‌شود.



شکل ۱. (الف)موقعیت جغرافیایی منطقه مطالعات و(ب) محل حفر پروفیل‌های شاهد خاک و (ج) مرزبندی مناطق مرجع و پذیرنده در مناطق مطالعاتی.

در هر تقسیم بندی برخی پارامترهای اندازه گیری می شود و پارامترهایی که تفاوت در میزان متغیر مسؤل را در دو گروه به حداکثر می‌رساند، به منظور انشعاب بندی درخت انتخاب می‌شود. در این روش هم داده های پیوسته و هم داده های دسته بندی شده را می تواند مورد برازش قرار دهد تا به پیش بینی کلاس‌های خاکی مجزا از هم و یا خصوصیات خاکی پیوسته بپردازد. از مزایای رگرسیون درختی بر مدل خطی این است که می‌تواند با روابط غیرخطی سر و کار داشته باشد. به علاوه آن ها به هیچ فرضی درباره داده نیاز ندارند و توانایی این را دارد که با رفتارهای غیرتجمعی سروکار داشته باشد در حالی که دیگر مدل ها نیاز دارند که این روابط شناسایی شوند (McBratney et al., 2003). از دیگر مزایای انالیزهای رگرسیون درختی تفسیر آسان و توانایی آن ها در یکی کردن داده های کمکی دسته ای و پیوسته می باشد که استفاده از این مدل را برای مکان هایی مثل ایران مناسب می کند (Taghizadeh et al., 2015). رگرسیون درختی ایجاد شده شامل گره‌هایی می‌باشد که هرکدام بیانگر یک خصوصیت و عارضه‌اند، هم‌چنین شامل شاخه‌هاست که هر شاخه ارزش هر خصوصیت را نمایش می‌دهد و شامل برگ می‌باشد که برگ ها یک خصوصیت و یا کلاس خاک پیش بینی شونده را نمایش می‌دهد (McBratney et al., 2003).

رگرسیون لاجیستیک نوعی از مدل های طبقه بندی احتمالات آماری است که به خانواده مدل‌های خطی جنرال تعلق دارد (Taghizadeh et al., 2015). این مدل ها برای متغیرهای وابسته دسته‌ای به خوبی کارایی دارند (Heung et al., 2016) و روابط بین یکسری از متغیرهای پیش بینی کننده را با یک متغیر وابسته مورد پیش بینی را به خوبی بیان می کنند. رگرسیون لاجستیک چندجمله‌ای با موقعیت هایی سر و کار دارند که نمونه‌های

مشاهده شده می‌توانند به سه و یا تعداد بیش تری گروه طبقه بندی شوند (Taghizadeh et al., 2015) و مدل کلی شده‌ای از رگرسیون لاجستیکی دوجمله‌ای است که خروجی‌های ممکن چندگانه از یک سری پیش بینی کننده‌ها با یکسری ضرایب به کار برده شده بر آن‌ها به دست می آیند و با استفاده از تخمین بیش ترین احتمال حضور آن، محاسبه می شود و یک کلاس به عنوان مرجع قبل از برازش مدل باید انتخاب شود و سپس توابع لگاریتمی مربوط به سایر کلاس‌ها نیز محاسبه می‌شوند وکلاس‌ها را با آن مقایسه می‌کند. در حقیقت تابع لگاریتمی از نسبت احتمالات بین قرارگیری پیکسلی در کلاس A به قرار گیری آن در کلاسی به جز کلاس A می‌باشد. این معادلات برای رگرسیون لاجیستیک دو جمله ای سرو کار دارد و در مورد رگرسیون های لاجیستیک چندجمله ای معادله به شکل زیر بیان می شود:

$$p(A) = \frac{e^{Bo+BA1X1+\cdots+BAkXk}}{e^{B0+BA1X1+\cdots+BAkXk} + e^{B0+BB1X1+\cdots+BBkXk} + \ldots + e^{B0+BE1X1+\cdots BEkXk}}$$

$$p(A) = \frac{\exp(A)}{\exp(A)+\exp(B)+\cdots+\exp(E)}$$

در این معادله B0 محل تقاطع منحنی رگرسیونی می باشد وB1 تا Bk ضرایب رگرسیونی هستند و با استفاده از داده های کمکی برای هر کلاس طبقه بندی به صورت مجزا محاسبه می شوند.

اجرای برون‌یابی

منطقه مطالعاتی به دو بخش که دارای روند تغییرات و فاکتورهای خاکساز مشابهی می‌باشند با عناوین منطقه مرجع و پذیرنده تقسیم بندی شد. جداسازی این دو سطح به گونه‌ای انجام شد که شاهد تغییرات مشابه و تنوع یکسانی باشند و از لحاظ تمامی فاکتورهای خاکساز مشابه باشند. منطقه A (شکل ۱ .ج) به عنوان منطقه مرجع با ۲۵۱۹۷ هکتار مساحت، که مدل پیش‌بینی کننده نقشه خاک براساس اطلاعات نمونه‌های خاک و متغیرهای محیطی مربوط به آن ساخته می‌شود و منطقه دوم (B) با مساحت ۳۷۵۴۳ هکتار که در اطراف آن قرار دارد،به عنوان منطقه پذیرنده انتخاب شده‌است. مدل درخت تصمیم و مدل رگرسیون لاجستیک چندجمله ای در منطقه مرجع با ۸۸ پروفیل به جهت تعیین روابط خاک و سیمای اراضی اجرا شد و خاک‌های این منطقه در سطح گروه بزرگ مورد پیش‌بینی قرار گرفتند. ۸۰٪ داده‌های این منطقه برای آموزش مدل استفاده شدند و ۲۰٪ مانده داده‌ها در جهت تعیین اعتبار مدل ساخته شده استفاده گردیدند. تمامی مراحل مدل سازی در نرم‌افزار R 3.4.3 و با بسته "C5.0" جهت مدلسازی به روش درخت تصمیم و با استفاده از بسته "nnet" برای مدلسازی به روش رگرسیون لاجستیک چندجمله ای اجرا شد. بعد از ایجاد مدل در منطقه مرجع از این مدل به جهت برون‌یابی و پیش بینی گروه بزرگ‌های شاهد در منطقه پذیرنده استفاده شد بدین ترتیب که از این مدل ۷۲ نمونه موجود در منطقه پذیرنده مورد پیش‌بینی قرار گرفتند و پس از ایجاد ماتریس ابهام دقت مدل در منطقه پذیرنده مشخص گردید. در این ماتریس ستون‌ها سطوح گروه بزرگ مشاهده شده در مطالعات زمینی و سطرها سطوح پیش بینی شده توسط مدل را تشکیل می‌دهند. عناصر قطری ماتریس مشاهداتی را که در هر سطح به درستی پیش‌بینی شده اند را نشان می‌دهند و عناصر غیرقطری بیانگر خطا می‌باشند. در نتیجه معیاری برای ارزیابی پیش‌بینی‌های صورت گرفته با استخراج شاخص‌های مناسب ازجمله دقت کلی و ضریب کاپا ارایه می‌دهد.

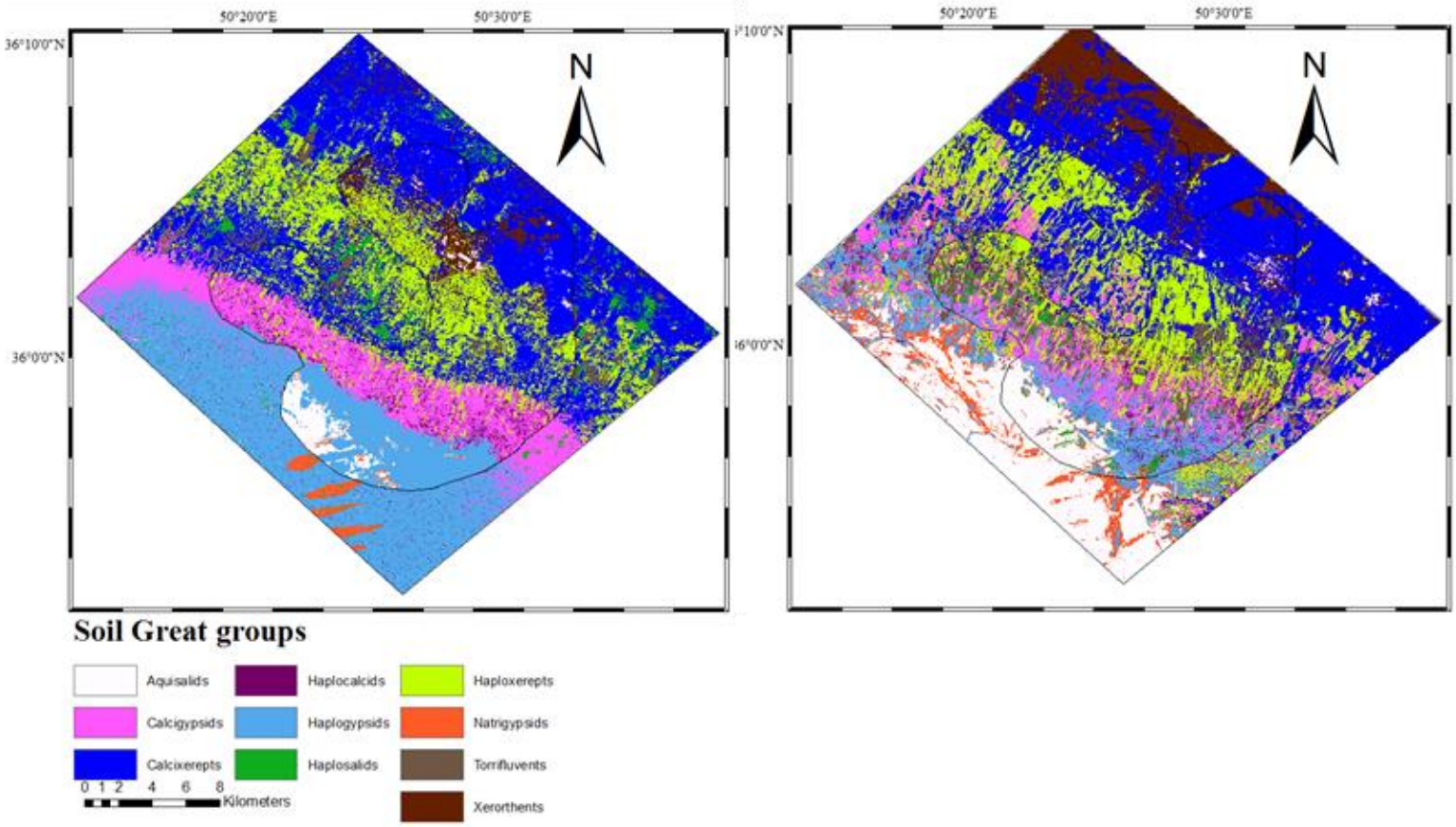
نتایج و بحث

پس ازبررسی، منطقه مطالعاتی از نظر تغییرات فاکتورهای خاکساز به دو سطح مرجع جهت مدلسازی و سطح پذیرنده جهت برون‌یابی تقسیم‌بندی شد. به‌طوری که شکل ۳. نقشه حاصل از مدلسازی نشان می‌دهد، هر دو سطح از روند تغییرات مشابهی تبعیت می‌کنند. نمونه‌های شاهد موجود در منطقه مرجع تا سطح گروه‌بزرگ نامگذاری شدند و سپس مدلسازی با استفاده از مدل درخت تصمیم الگوریتم c5.0 و رگرسیون لاجستیک چندجمله ای انجام شد. دقت روش درخت تصمیم در منطقه مرجع برابر با ۷۸٪ با ضریب کاپا ۰/۶۸ می باشد در حالی‌که دقت کلی رگرسیون لاجستیک برای منطقه مرجع ۵۲٪ همراه با ضریب کاپا ۰/۳۰ می باشد. نتایج برون یابی در منطقه پذیرنده برای مدل درخت تصمیم دقت کلی ۶۲٪ و مدل رگرسیون لاجستیک چندجمله ای ۵۲٪ را نشان داد (شکل ۳ و جدول ۱). با توجه به نتایج حاصل از این مطالعه، توانایی مدل ها در پیش بینی کلاس خاک متفاوت از یکدیگر می‌باشد و همچنین قابلیت پیش‌بینی هر مدل در فرایند برون‌یابی از فرایند درون‌یابی متفاوت می‌باشد. مدل درخت تصمیم به علت داشتن قدرت بالاتر و الگوریتم پیشرفته‌تر، نتایج صحیح‌تری را نسبت به مدل رگرسیون لاجستیک چندجمله‌ای در هر دو مورد درون‌یابی و برون‌یابی پیش‌بینی می‌نماید و نقشه حاصله از درخت تصمیم از همگنی بیش‌تری برخوردار می‌باشد. تفاوت در برون‌یابی و درون‌یابی منطقه مرجع، در

عدم حضور گروه بزرگ‌های توریفلوونتز و هاپلوسالیدز می‌باشد و از آن‌جا که هیچ نمونه‌ای با این گروه بزرگ در منطقه پذیرنده موجود نبوده‌است و هم‌چنین تعداد نمونه‌های تحت این عنوان در منطقه مرجع نیز کم می‌باشد لذا چنین گروه بزرگ‌هایی احتمال کم‌تری برای پیش‌بینی نیز داشته‌اند. از آن‌جایی که

شکل۳. نقشه درون‌یابی منطقه مرجع و نقشه حاصل از برون‌یابی منطقه پذیرنده (الف) مدل درخت تصمیم، (ب) مدل رگرسیون لاجستیک چندجمله‌ای

هاپلوزرپتز و کلسیزرپتز همبستگی بالایی با یکدیگر دارند و در سطوح بالاتر تاکسونومیک مشابه می‌باشند و نواحی مشابهی را در سطوح ژئومورفیک اشغال کرده‌اند و واحدهای ژئوفرم نقش مهمی را در شناسایی گروه بزرگ خاک در منطقه مطالعاتی داشته‌اند در نتیجه احتمال اینکه این خاک‌ها به اشتباه به یکدیگر پیش‌بینی شوند، زیاد می‌باشد و دقت تولید کننده در آن ها کم‌تر می‌شود. در روند تغییرات کلاس‌های خاک منطقه پذیرنده و جداسازی واحدهای مربوط به آن، همانند منطقه مرجع، ارتفاع و شکل زمین نقش مهمی را داشته‌اند. نتایج ارزیابی برون‌یابی با توجه به جداول مربوط به ماتریس ابهام در فرایند برون‌یابی (جداول ۲و ۳) نشان می‌دهد که خاک‌هایی که مشابه بودند وموقعیت‌های مشابهی را در دو منطقه مرجع و پذیرنده، اشغال کرده بودند با قدرت بالاتری مورد پیش‌بینی قرار گرفتند (Abbaszade Afshar et al., 2018). ازطرفی به علت شباهت در توزیع خاک‌های کلسیزرپتز و هاپلوزرپتز در طول سیمای اراضی و مقادیر مشابه متغیرهای کمکی مربوطه برای آن‌ها، آموزش مدل برای جداسازی خاک‌های مشابه با چالش مواجه می‌شود و به اشتباه این دو سطح را به‌یکدیگر پیش‌بینی می‌کند درنتیجه در ارزیابی دقت برون‌یابی این سطوح بیش ترین تعداد را از لحاظ پیش بینی اشتباه به یکدیگر را به‌خود اختصاص داده‌اند. با پایین رفتن سطح تاکسونومیکی به علت بالارفتن تنوع گروه بزرگ‌های خاک، پیچیدگی‌ها افزایش می‌یابد و احتمال خطا و کاهش دقت نیز بیش‌تر می‌شود. تفاوت در توانایی پیش‌بینی هر مدل ناشی از تفاوت در عملیات محاسباتی و الگوریتم‌های متفاوت استفاده شده در هر مدل می‌باشد که بالتبع نتایج متفاوتی را خواهد داشت. الگوریتم‌های ساده قادر نخواهند بود روابط پیچیده مشاهده شده طی پروسه آموزش مدل را شناسایی کند درنتیجه پیش‌بینی‌های آنها قابلیت اعتماد کم‌تری از مدل‌های پیش‌بینی کننده پیچیده خواهند داشت. روند پیش‌بینی حضور خاک‌ها در منطقه مطالعاتی برای هر دو مدل، مشابه می‌باشد اما دقت هر مدل با یکدیگر متفاوت می‌باشد. از آنجایی که



جدول۱. دقت حاصله از مدل درخت تصمیم و رگرسیون لاجستیک چندجمله ای برای مناطق مرجع و پذیرنده

	DT		MLR	
	OA (%)	Kappa	OA (%)	Kappa
Donor area	78	0.68	52	0.304
Recipient area	62	0.51	52	0.36

DT= Decision Tree, MLR= Multinomial Logistic Regression, OA= Overall Accuracy

جدول۲. ماتریس ابهام مربوط به برون‌یابی مدل منطقه مرجع در منطقه پذیرنده با روش درخت تصمیم

UNIT	Aquisalids	Calcigypsids	Calcixerepts	Haplocalcids	Haplogypsids	Haploxerepts	Natrigypsids	Xerorthents	%UA
Aquisalids	۴	۰	۰	۰	۰	۰	۱	۰	۸۰
Calcigypsids	۰	۳	۰	۳	۱	۰	۰	۰	۴۳
Calcixerepts	۰	۰	۱۶	۱	۰	۴	۰	۳	۶۷
Haplocalcids	۰	۰	۰	۳	۱	۰	۰	۰	۵۰
Haplogypsids	۰	۰	۰	۰	۷	۰	۰	۰	۱۰۰
Haploxerepts	۰	۰	۴	۰	۰	۱۳	۰	۳	۶۵
Natrigypsids	۰	۰	۰	۰	۰	۰	۰	۰	-
Xerorthents	۰	۰	۳	۰	۰	۳	۰	۰	۰
%PA	۱۰۰	۱۰۰	۷۰	۲۰	۷۸	۶۵	۰	۰	%OA=62 KA=0.51

%PA= % Producer's Accuracy, %UA=% User Accuracy, %OA= %Overall Accuracy, KA= KAPPA

جدول۳. ماتریس ابهام مربوط به برون‌یابی مدل منطقه مرجع در منطقه پذیرنده با روش رگرسیون لاجستیک چندجمله ای

UNIT	Aquisalids	Calcigypsids	Calcixerepts	Haplocalcids	Haplogypsids	Haploxerepts	Natrigypsids	Xerorthents	%UA
Aquisalids	۴	۰	۰	۱	۰	۰	۰	۰	۸۰
Calcigypsids	۰	۱	۰	۰	۱	۰	۰	۰	۵۰
Calcixerepts	۰	۰	۱۵	۰	۰	۹	۰	۳	۵۶
Haplocalcids	۰	۱	۱	۱	۳	۰	۰	۰	۱۷
Haplogypsids	۰	۰	۰	۰	۳	۰	۰	۰	۷۵
Haploxerepts	۰	۰	۷	۲	۰	۱۰	۰	۲	۴۸
Natrigypsids	۰	۰	۰	۰	۲	۰	۱	۰	-۳۴
Xerorthents	۰	۰	۰	۰	۰	۱	۰	۱	۵۰
%PA	۱۰۰	۳۴	۶۶	۲۵	۳۴	۵۰	۱۰۰	۱۷	%OA=52 KA=0.36

%PA= % Producer's Accuracy, %UA=% User Accuracy, %OA= %Overall Accuracy, KA= KAPPA

مدل درخت تصمیم نتایج درون‌یابی با دقت بالاتری را نشان داد این انتظار میرفت که برای پروسه برون‌یابی نیز عملکرد بالاتری نسبت به مدل رگرسیون لاجستیک چندجمله‌ای داشته باشد. مقایسه نتایج حاصل از این تحقیق با سایر مطالعات نشان می‌دهد که نتایج حاصله از این مطالعه از دقت مناسبی برخوردار می‌باشد. عباس زاده افشار و همکاران (۲۰۱۸) برون یابی را در منطقه پذیرنده با استفاده از مدل رگرسیون لاجستیک چندجمله‌ای در سطح گروه بزرگ خاک برازش دادند و نتایج آن‌ها دقت کلی ۵۹٪ و کاپا ۰/۴۷ را گزارش داد. والسکی و همکاران (۲۰۱۷) تکنیک درخت تصمیم را در برون‌یابی استفاده نمودند دقت کلی که گزارش دادند برابر با ۷۲/۹٪ همراه با ضریب کاپا ۰/۵۴ بوده است. لمرسیر و همکاران (۲۰۱۲) با استفاده از تکنیک برون‌یابی به الگوریتم درخت تصمیم بوست شده دقت کلی ۴۹٪ و کاپا ۰/۴۰را به‌دست آوردند و نتیجه‌گیری نمودند که روش برون‌یابی می تواند نقش مهمی در پیش‌بینی مکانی خصوصیات خاک ایفا نماید. تن کاتن و همکاران (۲۰۱۱) نیز برون یابی را در منطقه‌ای که از نظر پستی و بلندی تنوع داشته انجام دادند و دقت کلی ٪۶۱/۷۹ و کاپا ۰/۴۶ را در مطالعاتشان بیان نمودند.

نتیجه‌گیری

در این مطالعه، برون‌یابی برای تهیه نقشه رقومی کلاس خاک در سطح گروه بزرگ در مناطق فاقد اطلاعات با استفاده از مقایسه مدل درخت تصمیم و مدل رگرسیون لاجستیک چندجمله‌ای در منطقه پذیرنده که از نظر فاکتورهای خاکساز مشابه با منطقه مرجع بوده است بررسی می شود، از

آنجایی که تکنیک برون یابی دقت قابل قبولی را نشان داد (برای مدل درخت تصمیم دقت کلی ۶۲٪ و مدل رگرسیون لاجستیک چندجمله ای ۵۲٪) می‌توان از آن به عنوان روشی کارآمد با حفظ کاهش هزینه و سرعت بالا در تهیه نقشه رقومی استفاده شود و رابطه سیمای اراضی و خاک را در مناطق فاقد اطلاعات را با استفاده از مناطق مرجع مشابه تعیین نمود و این‌گونه مناطق با محدودیت اطلاعاتی را تحت پوشش اطلاعات قرار داد. از سویی در این مطالعه با مقایسه دو روش متفاوت پیش بینی کننده مشخص نمود که الگوریتم هایی که ساختار درختی دارند در مقایسه با رگرسیون لاجستیک چندجمله ای از قدرت بالاتری برای پیش بینی کلاس خاک برخوردار هستند که ناشی از توانایی بالاتر آنها در تعیین روابط پیچیده خاک می‌باشد در حالیکه مدل رگرسیون لاجستیک چند جمله‌ای به علت ساده‌تر بودن مدل از توانایی کم‌تری برای برقراری روابط خاک سیمای اراضی و پیش بینی کلاس خاک برخوردار بوده است.

منابع

Abbaszade Afshar, F., Ayoubi, S., Jafari, A. 2018. The extrapolation of soil great groups using multinomial logistic regression at regional scale in arid regions of Iran. Geoderma, 315, 36-48.

Arrouays, D., McBratney, A., Minasny, B., Hempel, J., Heuvelink, G., MacMillan, R., Hartemink, A., Lagacherie, P., andMcKenzie, N. 2014. The GlobalSoilMap project specifications, GlobalSoilMap: Basis of the global spatial soil information system, 9.

Arruda, G.P., Dematte, J.A.M., Chagas, Cd.S., Fiorio, P.R., Souza, A.B., Fongaro, C.T. 2016. Digital soil mapping using reference area and artificial neural networks. Scientia Agricola. 73(3), 266-273.

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

Legros, J.P. 2006. Mapping of the Soil. Science Publishers, Enfield.

Lemercier, B., Lacoste, M., Loum, M., Walter, C., 2012. Extrapolation at regional scale of local soil knowledge using boosted classification trees: a two-step approach. Geoderma 171-172, 75–84.

Mallavan, B.P., Minasny, B., McBratney, A.B., 2010. Homosoil, a methodology forquantitative extrapolation of soil information across the globe. In: Janis, L.,Boettinger, D.W., Howell, A.C., Moore, A.E., Hartemink, Kienast-Brown, S. (Eds.),Digital Soil Mapping Bridging Research, Environmental Application, and Operation. Springer, Dordrecht, Heidelberg, London, New York, pp. 137–150.McBratney, A.B., Santos, M.M., Minasny, B., 2003. On digital soil mapping. Geoderma 117(1), 3-52.

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

Minasny, B., McBratney, A.B. 2010. Methodologies for global soil mapping. In: Boettinger, J. l., Howell, D. W., More, A. C., Hartemink, A. E., Kienast-Browns, S., (Eds.), Digital Soil Mapping: Bridging Research, Environmental Application, and Operation. Springer, London, pp. 429-436.

R. Development Core Team, 2017. A language and environment for statistical computing. R foundation for statistical computing. Vienna, Austria

Soil Survey Staff. 2014. Keys to Soil Taxonomy, 12th ed. USDA-Natural Resources ConservationService, Washington, DC.

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.

Ten Caten, A., Dalmolin, R.S.D., Pedron, F.A., Mendonça Santos, M.L. 2011. Extrapolation of soil-landscape relationship from a reference area. Ciência Rural, Santa Maria 41(5), 812-816.

Wolski, M.S., Dalmolin, R.S.D., Flores, C.A., Moura-Bueno, J.M., Caten, A.T., Kaiser, D.R. 2017. Digital soil mapping and its implications in the extrapolation of soil-landscape elationshps in detailed scale. Pesq. Agropec. Bras., Brasília 52(8), 633-642.

Zeraatpisheh, M., Ayoubi, S., Jafari, A., Fink, P. 2017. Comparing the efficiency digital and conventional soil mapping to predict soil types in a semi-arid region in Iran. Geomorphology, 285(15), 186:204.

16th Iranian Soil Science Congress

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

Topic for submission: Novel Technologies in Soil Science

Comparing soil great group extrapolation by application of decision tree classification and multinomial logistic regession in Abyek, Qazvin

Neyestani[*1], M., Samadian[2], F., Jafari, A.[3] Keshavazi, A. [4]

[1] Ph.D Student, Soil Science Department, Faculty of Agriculture University of Tehran, Iran

[2] Associate Prof., Soil Science Department, Faculty of Agriculture University of Tehran, Iran

[3] Assistant Prof., Soil Science Department, Faculty of Agriculture Shahid Bahonar University of Kerman, Iran

[4] Assistant Prof., Soil Science Department, Faculty of Agriculture University of Tehran, Iran

Abstract

Digital soil mapping provides an efficient way to access soil properties and classes maps in regions with limited information. Our object by this study was considering and assessing extrapolation by applying two different classification algorithms as decision tree and multinomial logistic regression in areas with limited available information. To cover this purpose a region with approximate area of 25197 Km^2 as donor area for gaining predictive soil class model at great group level was selected. Produced model then was applied to regions with limited data to predict soil great group of soil samples in recipient area. Extrapolation accuracy was determined based comparing observed and predicted soil samples in confusion matrices. The results show us that overall accuracy of decision tree and multinomial logistic regression was 62% and 52% respectively. The accuracy was reasonably acceptable although the efficiency of decision tree was higher than multinomial logistic regression. As a result we can conclude that extrapolation could be an efficient, fast method with low cost to predict soil classes at regions with limited information.

Keywords: Digital soil mapping, modelling, extrapolation, decision tree, multinomial logistic regressio

[*] Corresponding author, Email: Neyestani_mehr@yahoo.com