

ترکیب فاصله تاکسونومیکی خاک‌ها و درخت تصمیم برای پیش‌بینی مکانی کلاس‌های خاک در اردکان

۱ - روح‌الله تقی‌زاده مهرجردی، استادیار دانشکده کشاورزی و منابع طبیعی، دانشگاه اردکان

۲ - فریدون سرمیدیان، استاد پردیس کشاورزی و منابع طبیعی، دانشگاه تهران
fsarmad@ut.ac.ir

۳ - محمود امید، استاد پردیس کشاورزی و منابع طبیعی، دانشگاه تهران

۴ - نورایر تومانیان، استادیار مرکز تحقیقات کشاورزی و منابع طبیعی اصفهان

۵ - محمد جواد روستا، استادیار مرکز ملی تحقیقات شوری

۶ - محمد حسن رحیمیان، مربی مرکز ملی تحقیقات شوری

دریافت: ۱۳۹۲/۰۶/۰۳

پذیرش: ۱۳۹۲/۱۲/۱۰

چکیده

نقشه‌برداری رقومی با مطالعه پروفیل‌های خاک و طبقه‌بندی کردن آن‌ها شروع می‌شود. مسئله اصلی در این روش ارتباط دادن کلاس‌های خاک با تعدادی متغیرهای کمکی است. روش‌های کنونی نقشه‌برداری، کلاس‌های خاک را تنها به عنوان‌هایی مستقل در نظر گرفته و در نهایت سعی در حداقل کردن خطای طبقه‌بندی دارند. این در حالی است که خاک در سطوح تاکسونومیکی دارای روابطی است. با استفاده از درخت تصمیم می‌توانیم الگوریتمی را تعریف کنیم که فواصل تاکسونومیکی را به جای خطای طبقه‌بندی حداقل کند. این پژوهش، برای پیش‌بینی مکانی کلاس خاک در منطقه‌ای به وسعت ۷۲۰ کیلومتر مربع در اردکان از درخت تصمیم استفاده شد. نخست، بر اساس روش مربع لاتین، مکان جغرافیایی ۱۸۷ پروفیل خاک مشخص، تشریح و نمونه‌برداری گردید و بر اساس سیستم آمریکایی طبقه‌بندی شد. متغیرهای محیطی یا نمایندگان عوامل خاک‌سازی که در این پژوهش استفاده شد شامل اجزاء سرزمین، داده‌های تصویر ETM+ ماهواره لندست و نقشه سطوح ژئومورفولوژی است. برای محاسبه فواصل تاکسونومیکی از آنالیز تشخیصی استفاده شد. نتایج نشان داد که استفاده از فواصل تاکسونومیکی منجر به دقت کلی مدل به میزان ۷۰٪ شده است. همچنین، نتایج نشان داد که برای پیش‌بینی کلاس خاک متغیرهای شاخص خیلی، سطوح ژئومورفولوژی و شاخص صافی قسمت میانی دره با درجه تفکیک بالا با اهمیت‌ترین هستند. در کل نتایج نشان داد که مدل‌های درختی به همراه فواصل تاکسونومیکی دارای دقت قابل قبولی است. بنابراین پیشنهاد می‌شود در مطالعات آینده جهت تهیه نقشه رقومی خاک از مدل‌های درختی به همراه فواصل تاکسونومیک استفاده شود.

واژگان کلیدی: درخت تصمیم؛ آنالیز تشخیص؛ دقت کلی؛ اردکان.

مقدمه

ها، به واسطه ارائه چند وجهی پراکنش انواع خاک‌ها و انجام ذهنی و دستی تجزیه و تحلیل‌ها به صورت فاحشی کاهش می‌یابد. افزون بر این، نمایش توزیع جغرافیایی خاک‌ها که بر مبنای تفکیک واحدهای نقشه بر روی عکس‌های هوایی و یا نقشه‌های توپوگرافی به همراه کنترل زمینی است، بدون خطا نیست، زیرا خطا در مشاهده و ترسیم واحدها به علت این که متکی بر قابلیت‌های فردی

نقشه‌برداری سنتی دارای مشکلات زیادی است. [۳۸] فهرست کاملی از محدودیت‌های روش مرسوم خاک شناسی را ارائه نمودند. از نظر مفهومی روش مرسوم خاک شناسی بر مدل منقطع یا مدل سطحی-کلاسی خاک‌ها استوار است که اساس و زیربنای مدل پلی‌گونی تغییرات متغیرهای طبیعت است. بر این اساس، توانایی محققین خاک شناس در ارائه توزیع مکانی حقیقی اطلاعات خاک-

را دارا است. دانشمندان خاک شناس استرالیایی در استفاده از مدل درختی پیش قدم بوده‌اند. [۵] یک مجموعه از قوانین پیش‌بینی کننده را با استفاده از نقشه‌های زمین‌شناسی، مدل رقومی ارتفاع و پارامترهای اراضی از نقشه ۱:۱۰۰۰۰۰ استخراج کردند. به طور مشابه، [۲۸] از تصاویر ماهواره‌ای، نقشه‌های زمین‌شناسی، مدل رقومی ارتفاع و مؤلفه‌های اراضی به عنوان ورودی (متغیرهای کمکی) مدل درخت تصمیم در نظر گرفته شده و اقدام به پیش‌بینی گروه‌های بزرگ خاک در استرالیا کردند. نتایج نشان دهنده کارایی این مدل با دقت ۵۳٪ است. همچنین، درخت تصمیم را برای تولید نقشه‌های واحدهای خاک با مقیاس ۱:۲۵۰۰۰ به کار برده و کارایی بالای درخت تصمیم را با دقت ۵۸٪ تأیید نمودند [۶]. [۱۳] درخت تصمیم را برای پیش‌بینی خصوصیات خاک از قبیل فسفر، ازت، ضخامت، بافت، درصد رس، واکنش خاک و کربن آلی به کار بردند. نتایج نشان داد که درخت تصمیم دارای دقت بالا جهت پیش‌بینی خصوصیات کمی خاک با ضریب تبیین ۰/۴۵ دارد. [۲۱] همچنین درخت تصمیم و مدل‌های خطی را برای تهیه نقشه‌های خاک استفاده کردند. نتایج این پژوهشگران نشان داد که متغیرهای ورودی مدل از تکنیک‌های به کار گرفته شده مهم‌تر هستند. [۳۴] از این مدل برای توسعه نقشه خاک در بیابان موه‌اوی کالیفرنیا استفاده کردند. ایشان برای نواحی کوهستانی و دشت دو مدل مختلف ارائه نمود و بیان داشت که انتخاب ورودی‌ها تأثیر کمی بر روی برازش کلی مدل داشته، ولی می‌تواند نوع قوانین را متأثر کند. همچنین ایشان نقشه ژئومورفولوژی را به عنوان متغیر کمکی مناسب در مناطق خشک معرفی کردند. [۳۳] از درخت تصمیم جهت پیش‌بینی کلاس‌های خاک در منطقه‌ای از آمریکا (یومینگ) استفاده کردند. ایشان از قوانین درخت تصمیمی که در منطقه‌ای متفاوت به دست آمده بود استفاده و نتایج را با روش طبقه‌بندی نظارت نشده مقایسه کردند. نتایج نشان داد که درخت تصمیم به خوبی توانسته است عمل برون-یابی را انجام داده و خاک‌ها را پیش‌بینی کند. [۱۲] برای پیش‌بینی کلاس‌های خاک در منطقه‌ای از نوادا، آمریکا از تکنیک درختی استفاده کرد. ایشان به این نتیجه رسید که تکنیک درخت تصمیم با دقت بالا (۵۴٪) گروه‌های خاک

و تجربی است اجتناب‌ناپذیر است [۸]. از دیگر مشکلات روش معمول نقشه‌برداری خاک‌ها، وابستگی تعیین حدود و مرز واحدهای خاک به قدرت درک و فهم کارشناس از روابط خاک و محیط است. به گونه‌ای که می‌توان گفت احتمال این که چند کارشناس مرز واحدی را از خاک‌های یک ناحیه ارائه دهند کم است [۱۱]. همچنین تفکیک واحدهای خاک بر اساس تفسیر چشمی عکس‌ها، ارتباط دادن لایه‌های مختلف داده‌های محیطی به صورت ذهنی، به کارگیری علم و دانش تجربی و مدل ارتباطی خاک-زمین‌نما، در تجزیه و تحلیل داده‌ها منجر به نادیده گرفته شدن بعضی از واقعیت‌های طبیعی و بروز خطا در به نقشه درآوردن تمام جزئیات می‌شود [۱۹]. افزون بر این، به روز کردن نقشه‌ها و اطلاعات خاک به روش دستی برای کشور پهناوری مانند ایران، به زمان زیادی نیاز دارد که عملاً آن را نیز غیر ممکن می‌سازد.

برای به تصویر کشیدن تغییرات مکانی خاک‌ها با هزینه کم‌تر و در مدت زمان کوتاه‌تر به گونه‌ای که نقشه‌ای قابل قبول از خاک ارائه شود، روش‌های تخمین مکانی خاک با تأکید بر منابع داده‌های قابل دسترس ارائه شده است [۲۰]. پایه و اساس نقشه‌برداری نوین مبتنی بر معادله اسکورپن^۱ است [۲۰]. در این معادله خصوصیات^۲ یا کلاس خاک^۳ با یک سری متغیرهای محیطی (متغیرهای کمکی^۴) ارتباط داده می‌شود. به این صورت که در هر موقعیت مشاهده‌ی خاک، یک سری متغیرهای محیطی وجود دارد که همبستگی بالایی با خصوصیات یا کلاس‌های خاک دارند که یافتن این ارتباط و پیش‌بینی خصوصیات یا کلاس‌های خاک در نقاط دیگر، منجر به تهیه نقشه رقومی خاک در تمام منطقه مورد مطالعه می‌گردد. توابعی که برای پیش‌بینی کلاس خاک می‌توان استفاده کرد عبارت است: رگرسیون لاجیستیک^۵ [۱۵]، شبکه‌های عصبی مصنوعی^۶ [۷]، طبقه‌بندی درختی یا تصمیم‌گیری درختی^۷ [۲۵]، آنالیز تطبیقی متعارفی^۸ [۱۸]. از این میان، درخت تصمیم در بین محققین جایگاه مناسب‌تری

1-SCORPAN

2-Soil attributes

3-Soil class

4-Auxiliary variables

5-Logistic Regression

6-Artificial Neural Networks

7-Decision Tree

8-Canonical Correspondence Analysis (CCA)

مسطح بوده و تنها رشته کوه‌هایی در شمال شرق منطقه مشاهده می‌شود. بلندترین ارتفاعات منطقه ۱۹۴۴ متر و کم‌ارتفاع‌ترین قسمت منطقه که در نواحی میانی قرار گرفته دارای ارتفاع متوسط ۹۴۴ متر است. لندفرم‌های اصلی این منطقه با روند کلی شرق به غرب شامل کوه، مخروط افکنه‌های جدید، سطوح رسی و تپه‌های گچی است.

نمونه‌برداری هایپرکیوب

این روش، یک طرح تصادفی طبقه‌بندی شده‌ای است که باعث نمونه‌برداری مؤثری به کمک توزیع چند متغیر می‌شود. یک شبکه مربع حاوی موقعیت‌های نمونه، یک شبکه لاتین است اگر و تنها اگر فقط یک واحد نمونه-برداری در هر ردیف و هر ستون وجود داشته باشد. یک هایپرکیوب لاتین، تعمیم این مفهوم به تعداد اختیاری از ابعاد است. به طوری که هر واحد نمونه‌برداری فقط در یک صفحه آن قرار گرفته باشد. LHS شامل نمونه‌برداری n مقدار از توزیع تشریح شده هر متغیر است. توزیع تجمعی هر متغیر به n فاصله با احتمال مساوی تقسیم می‌شود و یک مقدار از هر فاصله به طور تصادفی انتخاب می‌شود. سپس، n مقدار به دست آمده برای هر متغیر با متغیرهای دیگر جفت می‌شود. این روش یک پوشش کاملی از هر متغیر را ایجاد می‌کند [۲۴]. بر اساس تکنیک هایپرکیوب، محل ۱۸۷ پروفیل در منطقه اردکان به وسعت ۷۲۰۰۰ هکتار انتخاب شد (شکل ۱).

تجزیه آزمایشگاهی

از لایه‌های ژنتیکی همه مقاطع نمونه‌برداری و برای تجزیه به آزمایشگاه منتقل شد. تعدادی از ویژگی‌های فیزیکی و شیمیایی نمونه‌ها به شرح زیر اندازه‌گیری شد: بافت خاک به روش هیدرومتری [۹]، واکنش خاک نمونه‌های خاک در گل اشباع به وسیله pH متر، قابلیت هدایت الکتریکی عصاره اشباع به وسیله هدایت‌سنج، کربنات کلسیم معادل خاک به روش تیتراسیون برگشتی با اسیدکلریدریک ۲ نرمال، کاتیون و آنیون‌های محلول بر اساس روش‌های استاندارد [۳۵]، درصد کربن آلی خاک به روش اکسایش تر [۳۷]، و درصد گچ [۲۹] تعیین شد.

را پیش‌بینی می‌کند. [۱۰] توانایی رگرسیون درختی توسعه یافته را برای پیش‌بینی دقیق واحدهای ژئومورفولوژی در نواحی بازدید نشده بررسی کردند. آن‌ها دریافتند توانایی تخمین مدل در موارد برون‌یابی پایین است، ولی برای درون‌یابی توانایی بالایی دارد. [۱۷] جهت تهیه نقشه رقومی کلاس‌های خاک در منطقه زرد کرمان از تکنیک درخت تصمیم توسعه یافته استفاده و اذعان داشتند که این تکنیک دارای کارایی بالایی است.

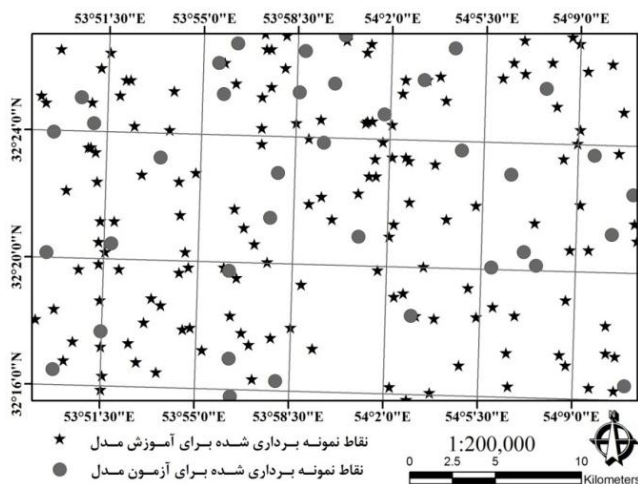
مشکل اساسی درخت تصمیم این است که الگوریتم مورد استفاده در روش، سعی در حداقل کردن خطای طبقه‌بندی دارد [۱۷]. این در حالی است که بین سطوح تاکسونومیک خاک ارتباط‌هایی وجود دارد. بنابراین، دخالت دادن فاصله‌های تاکسونومیک خاک در درخت تصمیم حائز اهمیت است. به عنوان مثال [۲۵] از ترکیب فاصله تاکسونومیک و تصمیم‌گیری درختی برای نقشه برداری رقومی استفاده نمود. نتایج نشان داد که فواصل تاکسونومیک توانسته دقت مدل درختی را افزایش دهد.

این پژوهش نیز سعی بر این دارد تا از فواصل تاکسونومیکی که بین گروه‌های بزرگ تاکسونومیکی خاک وجود دارد، استفاده کرده و آن‌ها را به همراه مدل درختی جهت پهنه‌بندی رقومی واحدهای خاک تا سطح گروه بزرگ در منطقه اردکان به کار گیرد.

مواد و روش‌ها

معرفی منطقه مورد بررسی

منطقه مورد مطالعه در ایران مرکزی و در استان یزد قرار دارد. مساحت آن برابر با ۷۲۰۰۰ هکتار از اراضی دشت یزد-اردکان را شامل می‌شود. این منطقه بین طول جغرافیایی 50° تا 53° و عرض جغرافیایی 16° تا 27° شمالی قرار دارد. این محدوده دارای مشکلات شوری منابع آب و خاک است. میانگین تبخیر و تعرق مرجع، دما، و بارندگی سالانه منطقه مورد مطالعه به ترتیب برابر ۳۴۸۳ میلی‌متر، $18/5^{\circ}\text{C}$ و ۷۵ میلی‌متر است. بارندگی بسیار کم بوده و محدود به بارش‌های زمستانی است. بنابراین، رژیم رطوبتی و حرارتی خاک به ترتیب اریدیک و ترمیک طبقه‌بندی می‌شود. منطقه مورد مطالعه از نظر توپوگرافی بیشتر



شکل ۱- پراکنش نقاط نمونه برداری شده بر حسب مربع لاتین (در شکل مکان هندسی نقاط آموزشی و آزمون مدل مشخص شده است، این داده‌ها به صورت تصادفی انتخاب شده‌اند).

تهیه داده‌های محیطی یا کمکی

نقشه ژئومورفولوژی

در محیط سامانه جغرافیایی ساگا محاسبه و استخراج شد [۱۴].

تصاویر ماهواره‌ای

خاک‌های گوناگون خصوصیات طیفی متفاوتی دارند [۲]. با توجه به این نکته که بیشتر منطقه مورد مطالعاتی فاقد پوشش گیاهی بوده، بنابراین تغییرات خصوصیات و نوع خاک‌ها به راحتی توسط اطلاعات ماهواره‌ای قابل تشخیص است [۲۳]. در این مطالعه از تصاویر ماهواره‌ای لندست سنجد $ETM+$ استفاده شد. برای کاهش تعداد داده‌های خام (تعداد باندها) تصویر ماهواره‌ای از تجزیه به مولفه‌های اصلی بر اساس ماتریس همبستگی استفاده شد. بر اساس مؤلفه‌های خاک‌سازی در منطقه مجموعه‌ای از شاخص‌ها شامل شاخص گیاهی نرمال شده (نشان‌دهنده وضعیت پوشش گیاهی)، شاخص رس، شاخص کربنات، شاخص گچ، شاخص شوری و شاخص روشنایی (نشان-دهنده ترکیبات زمین‌شناسی مواد سطحی) محاسبه شد.

آماده‌سازی داده‌ها برای مدل‌سازی

برای پهنه‌بندی کلاس‌های خاک در سطح گروه بزرگ، داده‌های کلاس خاک کد گذاری شده و به عنوان متغیر وابسته در نظر گرفته شد. همچنین، تمام لایه‌های

بر روی عکس‌های هوایی ۱:۵۵۰۰۰ منطقه روکش طلق مانندی کشیده شد تا خطوط سطوح ژئومورفیک روی آن‌ها ترسیم شود. برای تفکیک واحدهای اراضی از سیستم استریوسکوپی برای تشخیص تفاوت‌ها استفاده شد. سپس عکس‌های هوایی و روکش مربوطه اسکن شده و وارد محیط سامانه اطلاعات جغرافیایی (نرم افزار پی‌سی-آی ژئوماتیکا) شد. در مراحل بعد، عکس‌ها و روکش‌های وارد شده به سامانه جغرافیایی با استفاده از تصاویر ماهواره‌ای به روش ارتوفوتوژئورفرنسینگ [۳۱] زمین مرجع شدند. با رقومی کردن مرزهای ترسیم شده بر روکش عکس‌ها، نقشه ژئومورفولوژی تهیه شد.

استخراج پارامترهای زمین‌نما

در این تحقیق، از مدل رقومی سازمان نقشه‌برداری کشور (قدرت تفکیک مکانی ۱۰ متر) استفاده شد. بعد از آماده کردن مدل ارتفاعی رقومی، از آن برای استخراج اطلاعات کمکی یا تصاویر تخمین‌زننده متغیرهای ژنتیکی خاک استفاده شد. متغیرهای زمین‌نما از قبیل شیب، ارتفاع، ارتفاع بالای شبکه زهکشی، مساحت حوزه‌های اصلاح شده، موقعیت میانی شیب، عمق دره، شاخص خیزی، شاخص صافی قسمت میانی دره با درجه تفکیک بالا، شاخص بالای پشته با درجه تفکیک بالا و شیب حوزه

که مشاهدات هر گروه با یکدیگر شبیه باشند و مشاهدات گروه‌های مختلف نسبت به یکدیگر کمترین شباهت را داشته باشند. همچنین در آنالیز تشخیصی توابعی بر حسب مقادیر متغیرهای مورد اندازه‌گیری، امتیازی برای عضویت در هر گروه محاسبه می‌شود. با قرار گرفتن خاک‌ها در گروه‌های مختلف می‌توان فاصله برداری بین آن‌ها را محاسبه نمود. برای انجام این آنالیز، کلاس‌های خاک به عنوان متغیر مستقل و اطلاعات محیطی (به‌جز اطلاعات کیفی مثل ژئومورفولوژی) وارد نرم‌افزار JMP گردید. پس از گروه‌بندی، فواصل به صورت کمی به دست آمد. سپس از این فواصل به جای الگوریتم کمینه‌سازی خطای درخت تصمیم استفاده گردید.

بررسی عملکرد تکنیک نقشه‌برداری رقومی

برای بررسی کارایی مدل مورد استفاده و مقایسه آن با درخت تصمیم معمولی، در برآورد متغیرهای وابسته (گروه بزرگ خاک) از ضریب کاپا و دقت کلی استفاده شد. لازم به توضیح است که برای بررسی دقت مدل برای پیش‌بینی کلاس خاک، داده‌ها به دو دسته آموزش (۸۰٪) از داده‌های نمونه‌برداری شده) و آزمون (۲۰٪) از داده‌های نمونه‌برداری شده) به طور تصادفی تقسیم شدند.

نتایج و بحث

طراحی الگوی نمونه‌برداری

در این تحقیق از روش طراحی الگوی نمونه‌برداری مربع لاتین برای انتخاب ۱۸۷ محل نمونه‌برداری از بین ۸۴۲۷۶۶ فضای ممکن استفاده شد. ورودی‌های این مدل-ها بر اساس تکنیک آنتروپی شانون انتخاب شد (جدول ۱). با توجه به جدول، ۱۰ پارامتری که دارای وزن بیشتری هستند، از قبیل نقشه ژئومورفولوژی، ارتفاع، شاخص خیزی، شاخص صافی قسمت میانی دره با درجه تفکیک بالا، جهت حوزه، تجزیه به مولفه‌های اول و دوم تصویر ETM+، شیب، موقعیت شیب میانی و انحنای شیب وارد مدل مربع لاتین شد.

همان‌طور که در جدول ۱ ملاحظه می‌شود مجموع اوزان یک نمی‌شود که به این دلیل است که پارامترهای دیگر مقدار وزن کمتری را به خود اختصاص داده و در

اطلاعاتی یا کمکی به فرمت رستری با اندازه سلولی ۳۰×۳۰ متری تهیه شد. این اطلاعات به عنوان متغیرهای وابسته در مدل اسکورپن شناخته می‌شوند. تمامی اطلاعات (داده‌های خاک و متغیرهای محیطی) به فرمت نهایی متنی (txt) تبدیل و برای انجام فرآیند مدل‌سازی در نرم‌افزارهای مربوطه آماده شد.

مدل‌سازی خاک

درخت تصمیم

مدل درختی طبقه‌بندی و رگرسیون، یک روش پارامتری الگوریتمی است که قادر به پیش‌بینی متغیرهای کمی یا متغیرهای طبقه‌بندی شده بر اساس مجموعه‌ای از متغیرهای پیش‌بینی کننده کمی و کیفی است. در این روش، مجموعه‌ای از شرط‌های منطقی به صورت یک الگوریتم با ساختار درختی برای طبقه‌بندی یا پیش‌بینی کمی یک متغیر به کار می‌رود. ایجاد درخت تصمیم شامل دو مرحله است. مرحله اول ایجاد و رشد درخت است. این مرحله شامل پیوند و انشعاب است. مرحله دوم، مرحله توقف و هرس است. هدف از این مرحله به حداقل رساندن خطای پیش‌بینی است. در تحقیق حاضر از نرم‌افزار سی ۵ [۳۰] برای ساختن درخت تصمیم و در نهایت پیش‌بینی کلاس‌های خاک استفاده شد.

فاصله تاکسونومیک

برای محاسبه فاصله تاکسونومیکی که بین گروه‌های خاک وجود دارد، روش آنالیز تشخیصی در این مطالعه استفاده گردید. آنالیز تشخیصی، راهکاری است برای آن‌که متغیرها را قالب گروه‌های مجزا از هم تفکیک کنیم، به صورتی که هر گروه با وجود این‌که با گروه دیگر شباهت و همبستگی دارد، از انسجام لازم نیز برخوردار باشد. در واقع آنالیز تشخیصی، اعضای جامعه مورد تحقیق را در قالب گروه‌های مجزا از هم طبقه‌بندی و تفکیک می‌کند. افزون بر این تابع تشخیص، معادله‌ای است که با داشتن مشخصات هر فرد از جامعه، می‌توان با قرار دادن این مشخصات در آن معادله، پیش‌بینی کرد که وی به کدام گروه تعلق دارد. به عبارت دیگر، منظور از آنالیز تشخیصی، گروه‌بندی داده‌ها به گروه‌های متجانس است، به گونه‌ای

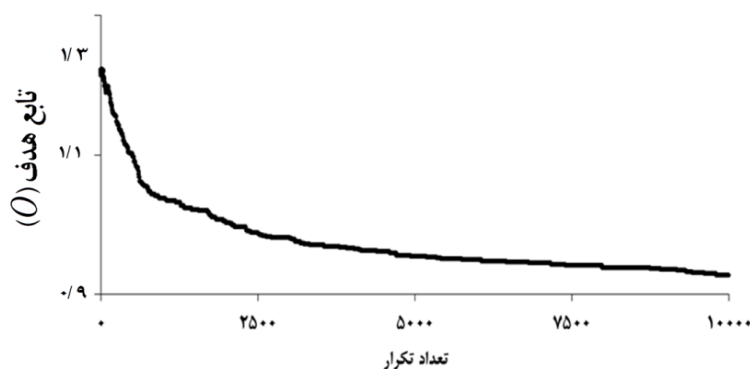
نتیجه وارد مدل نشده‌اند.

جدول ۱- پارامترهای ورودی مدل‌های مربع لاتین و فازی کی‌مینز و وزن‌های مربوطه بر اساس روش آنترپوی شانون

پارامتر	ژئومورفولوژی	ارتفاع	شاخص خیزی	ام‌آر بی‌اف	جهت	مولفه اول	مولفه دوم	شیب	میداسلوپ	انحنای شیب	وزن
	۰/۲۸	۰/۱۱	۰/۱۷	۰/۲۵	۰/۰۸	۰/۰۹	۰/۰۶	۰/۰۰۱	۰/۰۰۱	۰/۰۰۱	۰/۰۰۱

آشفستگی از خود نشان می‌دهد. برای این تحقیق از تعداد تکرار ۱۰۰۰۰ برای یافتن جواب بهینه استفاده شد. محل-هایی که با استفاده از مربع لاتین انتخاب شده در شکل ۱ ارائه شده است. باید توجه شود که مناطق شهری از نقشه اولیه حذف شد.

تکامل معادله هدف با ۱۰۰۰۰ تکرار در شکل ۲ ارائه شده است. این الگوریتم شروع به نمونه‌برداری تصادفی کرده تا بتواند نقاطی که مربع لاتین را تشکیل می‌دهد، پیدا کند. مقدار تابع هدف (O) با افزایش تکرار کاهش می‌یابد و در شرایط تکرارهای کم‌تر از ۱۰۰۰ قدری



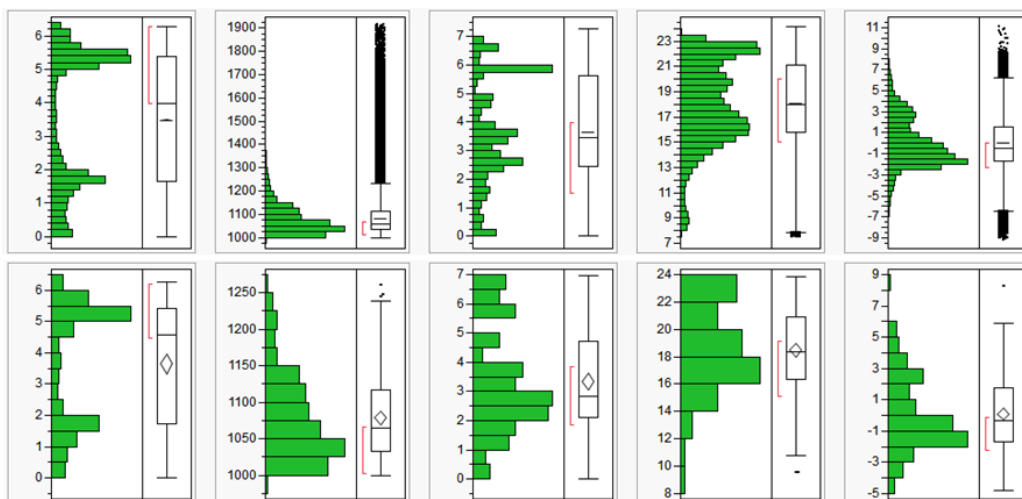
شکل ۲- تکامل معادله هدف با تعداد تکرار ۱۰۰۰۰

کرده است. [۳۶] نیز برای مشخص کردن محل‌های نمونه-برداری در منطقه‌ای به وسعت ۲۳۰۰۰ هکتار در کشور فیلیپین، اقدام به مقایسه روش‌های تصادفی، شبکه‌ای و مربع لاتین نمود. نتایج حاکی از برتری مربع لاتین داشت. [۳۲] نیز به نتایج مشابه‌ای در مورد مربع لاتین دست یافت.

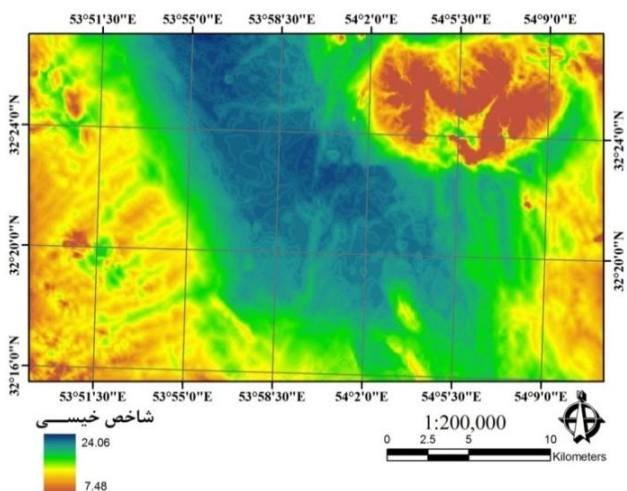
متغیرهای محیطی

با بازدیدهای صحرایی، تفسیر عکس‌های هوایی و تطابق آن‌ها با واقعیت و استفاده از تصاویر ماهواره‌ای، چهار نوع زمین‌نما و ۲۵ سطح ژئومورفیک در منطقه مورد مطالعه تشخیص داده شد. تفکیک سطوح ژئومورفیک متفاوت در مرحله تفسیر عکس‌های هوایی براساس دانش ژئومورفولوژی و اطلاعات قابل درک از روی عکس‌های هوایی منطقه بود و طی بازدیدهای صحرایی براساس شیب، میزان تجمع مواد دانه ریز و دانه درشت، تراکم آبراهه‌ها، میزان رطوبت خاک و سطح آب زیرزمینی صحت تقسیمات مشخص شد.

برای بررسی دقت روش نمونه‌برداری، توزیع هیستوگرامی متغیرهای استفاده شده مدل در محل‌های منتخب رسم شد (شکل ۳). به نظر می‌رسد که توزیع متغیرهایی انتخاب شده با استفاده از مربع لاتین بیشترین تبعیت را از توزیع جامعه دارند. در حقیقت، الگوی نمونه-برداری مربع لاتین طوری نقاط نمونه را از جامعه انتخاب کرده که توزیع نمونه‌گیری در هیستوگرام یکنواخت باشد. [۲۴] نیز در منطقه‌ای به وسعت ۱۴ کیلومتر مربع در کشور استرالیا اقدام به مقایسه چهار روش نمونه‌برداری مربع لاتین، میانگین کا، مولفه‌های اصلی و تصادفی نمود. ایشان از لایه رقومی ارتفاع، تصویر ماهواره‌ای ETM+ و واحدهای یکنواخت خاک/چشم‌انداز به عنوان ورودی‌های مدل مربع لاتین استفاده کرد. نتایج نشان داد که روش مربع لاتین به خوبی هیستوگرام متغیرهای کمکی در منطقه را در بر گرفته و نمونه‌برداری میانگین کا تنها توزیع جغرافیایی مناسبی دارند. همچنین، [۴] نیز شباهت توزیع فراوانی روش مربع لاتین به متغیرهای جامعه را دلیل برتری نسبت به سایر روش‌های نمونه‌برداری گزارش



شکل ۳- هیستوگرام روش مربع لاتین از بالا به پایین: داده‌های اصلی و نمونه‌برداری مربع لاتین؛ متغیرهای مختلف از چپ به راست: جهت شیب حوزه، ارتفاع، شاخص خیسی، شاخص صافی قسمت میانی دره با درجه تفکیک بالا و مؤلفه درجه اول تصویر ماهواره‌ای.



شکل ۴- نقشه شاخص خیسی با استفاده از مدل رقومی ارتفاع.

مدل سازی مکانی

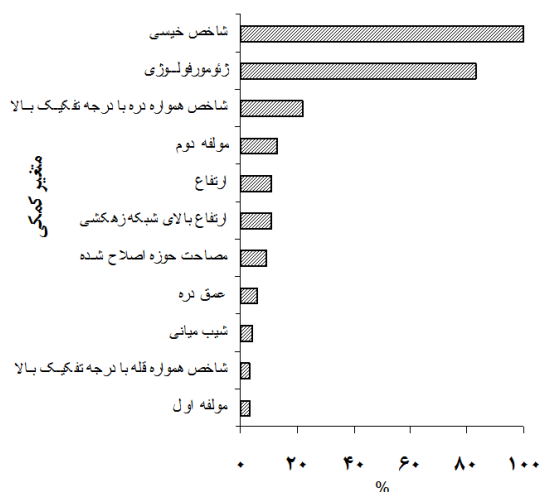
آنالیز درخت تصمیم نشان داد که بعضی از متغیرهای کمکی مانند شاخص خیسی، سطوح ژئومورفولوژی، شاخص صافی قسمت میانی دره با درجه تفکیک بالا، مؤلفه درجه دوم، ارتفاع و ارتفاع بالای شبکه زهکشی بالاترین تأثیر را بر روی پیش‌بینی گروه‌های بزرگ خاک را دارد (شکل ۶). این شکل تأثیر هر متغیر کمکی به صورت درصد را بر روی مدل درختی نشان می‌دهد. با توجه به این شکل، استنباط می‌شود که شاخص خیسی مؤثرترین متغیر کمکی در مدل بوده و این شاخص ۱۰۰ درصد در مدل استفاده شده است. همچنین، متغیر کمکی مهم دیگر، شاخص صافی قسمت میانی دره با درجه تفکیک بالا است که ۲۷ درصد در مدل وارد شده است. این موضوع

دومین متغیر محیطی مورد استفاده در این تحقیق، تصویر ماهواره‌ای ETM+ است. در این پژوهش، از مؤلفه‌های اصلی به جای باندهای ۱ تا ۶ تصویر استفاده شد. نتایج مؤلفه‌های اصلی نشان داد که سه مؤلفه اول ۹۹ درصد تغییرات را شامل می‌شود. بنابراین، سه مؤلفه اول به عنوان ورودی مدل انتخاب شد. اولین مؤلفه ۸۹/۸۱ درصد تغییرات باندهای تصویر را در بر می‌گیرد و بالاترین مقدار ویژه متعلق به باند ۲ و کمترین آن مربوط به باند ۴ است.

سومین متغیر محیطی مورد استفاده در این پژوهش، از آنالیز مدل رقومی ارتفاع به دست آمد. شکل ۴ تغییرات مکانی شاخص خیسی را نشان می‌دهد. مقادیر کم شاخص خیسی (>۷) به طور عموم، در مناطق کوهستانی واقع شده است. مقادیر حدواسط که در حدود ۱۴ است، قسمت‌هایی از اراضی باهادا را پوشش می‌دهد و بالاترین مقادیر این شاخص با شکل اراضی پلایا همخوانی دارد. در حقیقت، این ناحیه توان بالقوه بالای تجمع مواد شور را نشان می‌دهد. پلایا به خصوص واحد P1113 در انتهای حوزه دشت یزد اردکان قرار گرفته است و در نتیجه مواد با قابلیت حلالیت بالا را دریافت می‌کند که این پدیده منجر به شور شدن خاک‌های این واحد شده است. بین شاخص خیسی و شوری خاک ارتباط زیادی وجود دارد که این موضوع با یافته‌های پژوهشگرانی مانند [۲۷] و [۱۷] مطابقت دارد.

شاخص خیسی را تایید کرده‌اند. [۲۷] نیز در منطقه خشک اقدام به پیش‌بینی کلاس‌های خاک نموده و شاخص خیسی و همواری دره با درجه تفکیک بالا را به عنوان متغیر مهم برای مدل خود معرفی کردند. سطوح ژئومورفولوژی در رده دوم قرار دارد و ۸۴ درصد در مدل استفاده شده است. سطح ژئومورفولوژی نیز یک متغیر بسیار مهم است. این موضوع را می‌توان این‌گونه توضیح داد که سطوح ژئومورفیک در طول یک دوره زمین‌شناسی همراه با شرایط تشکیل خاک نزدیک به فرآیندهای خاکساز کنونی در مناطق خشک تشکیل شده است، چرا که سطوح ژئومورفیک آخرین و نزدیک‌ترین سطح به شرایط خاک در مقیاس بزرگ است. اثر فرآیندهای ژئومورفولوژی روی الگوی توزیع خاک‌ها و توسعه آن‌ها توسط تعدادی از محققین بررسی شده است. به عنوان مثال [۳۴] و [۱۷] تأکید کردند که سطوح ژئومورفولوژی مهم‌ترین متغیر کمکی برای شناسایی خاک‌ها در مناطق بیابانی است. نتایج همچنین نشان داد که سایر متغیرهای کمکی تأثیر چندانی بر روی مدل برای پیش‌بینی گروه‌های بزرگ خاک ندارند.

می‌تواند تأیید کننده این مطلب باشد که در این منطقه، پستی و بلندی از مهم‌ترین مؤلفه‌های خاک سازی است. اجزاء سرزمین در توزیع مکانی خاک و خصوصیات آن بسیار مؤثر هستند، چرا که توزیع مکانی تابش خورشیدی، درجه حرارت، رطوبت و جریان مواد را تحت تأثیر قرار می‌دهد. یکی از اجزاء سرزمین، شاخص خیسی یا رطوبتی است که برای تعیین توزیع مکانی نفوذ و جریان آب سطحی به کار می‌رود. در واقع تأثیر توپوگرافی بر مکان و میزان تجمع رطوبت در خاک و یا آب در سطح زمین‌نما را نشان می‌دهد. این متغیر با متغیرهای وابسته‌ای که با رطوبت ارتباط دارند و تحت تأثیر رطوبت قرار دارند، همبستگی بیشتری نشان می‌دهد. به‌طوری‌که در این مطالعه، شاخص خیسی به عنوان یک متغیر پیش‌بینی-کننده معنی‌دار در مدل‌های افق مشخصه سالیک وارد شد. شاخص خیسی و شاخص صافی قسمت میانی دره با درجه تفکیک بالا مناطق مستعد برای دریافت رسوبات را (نواحی میانی) نشان می‌دهند و این قسمت از مناطق بیشتر از خاک‌های شور تشکیل شده‌اند. در حقیقت این شاخص‌ها در شناسایی گروه بزرگ هاپلوسالیدز کمک می‌کنند. به طور مشابه [۱۷] ارتباط معنی‌دار بین خاک‌های شور و



شکل ۶- نتایج اهمیت متغیرهای محیطی در پیش‌بینی گروه‌های بزرگ خاک بر حسب درصد.

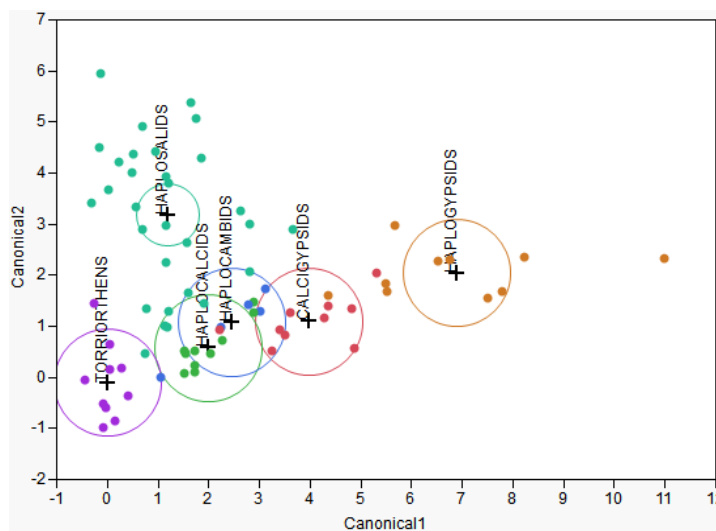
برای سطح گروه بزرگ مشخص شود. نتایج ماتریس خطا برای مرحله آزمون با ۳۷ داده نشان داد که بالاترین دقت متعلق به گروه هاپلوسالیدز (۰/۸۸) بوده و کم‌ترین دقت طبقه‌بندی مربوط به کلاس هاپلوکمبیدز خاک است که

بعد از به‌دست آوردن متغیرهای محیطی مهم‌تر که در بالا توضیح داده شد، مدل درختی با الگوریتم حداقل کردن خطا را بر روی داده‌های آزمون (۲۰ درصد از داده‌ها) آزمایش شد تا دقت این مدل برای پیش‌بینی کلاس خاک

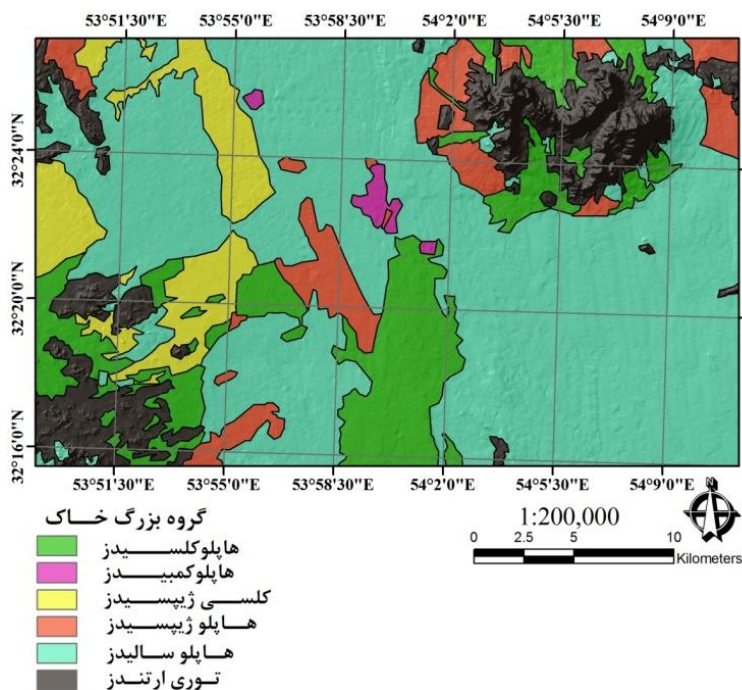
حدود اطمینان ۹۵٪ در شکل مشخص است. هاپلوژنوسایدز، توری‌ارتندز و هاپلوسالیدز به خوبی از گروه‌های دیگر جدا شده‌اند. که احتمالاً به علت خصوصیت خاص این سه گروه و تعداد بسیار زیاد داده در این گروه‌ها است. هاپلوکمبیدز و هاپلوکلیدز تقریباً همپوشانی دارند که این مسئله شباهت خصوصیات خاک این دو گروه را نشان می‌دهد. اما باید دقت شود به علت تعداد کم هاپلوکمبیدز این گروه دارای عدم قطعیت فراوانی است. بر اساس مراکز این گروه‌ها می‌توان فاصله‌ها را بین گروه‌های مشاهداتی محاسبه کرد. سپس فواصل بدست آمده را به عنوان ماتریس کاهش دهنده به همراه درخت تصمیم به کار گرفته شد، که باعث بهبود دقت کلی مدل در حدود ۱۸٪ گردید. به این صورت که دقت کلی مدل از ۵۲/۶٪ (مدل درختی با الگوریتم حداقل کردن خطا) به ۷۰٪ (مدل درختی با الگوریتم حداقل کردن فاصله تاکسونومیک) افزایش پیدا کرد. [۲۵] نیز نشان دادند که به کار بردن فواصل تاکسونومیک باعث افزایش دقت مدل نهایی شده است. در نهایت با استفاده از قوانین به دست آمده، نقشه رقومی واحدهای خاک در سطح گروه بزرگ تهیه شد (شکل ۳).

دارای دقت پیش‌بینی صفر درصد است. دقت کلی مدل ۵۲ درصد برای پیش‌بینی کلاس خاک برای سطح گروه بزرگ است. محققین مختلفی کارآیی درختان تصمیم را در پیش‌بینی کلاس‌های خاک تأیید کرده‌اند [۵، ۶، ۱۳، ۲۲، ۲۵، ۲۶، ۲۸]. اما مشکل درخت تصمیم و یا حتی همه تکنیک‌های داده‌کاوی این است که الگوریتم مورد استفاده در روش سعی در حداقل کردن خطای طبقه‌بندی دارد [۳]. در حالی که در سطوح تاکسونومیک، خاک‌ها به یکدیگر مرتبط هستند و در بعضی از فواصل تاکسونومیک خطا در پیش‌بینی نوع خاصی از کلاس بسیار مهم‌تر است (مثلاً فاصله هاپلوسالیدز با ژنوسایدز و توری‌ارتندز در سطح گروه بزرگ). بنابراین، دخالت دادن فاصله تاکسونومیک در طبقه‌بندی خاک حائز اهمیت می‌باشد. یکی از روش‌ها برای محاسبه فواصل تاکسونومیک استفاده از آنالیز تشخیصی بر اساس خصوصیات فیزیکی-شیمیایی خاک است [۱۶].

در این تحقیق از خصوصیات درصد رس، آهک، گچ و قابلیت هدایت الکتریکی و نسبت سدیم جذب شده برای محاسبه آنالیز تشخیصی استفاده گردید. شکل ۲ نتایج آنالیز تشخیصی را نشان می‌دهد. میانگین هر کلاس با



شکل ۲- آنالیز تشخیصی برای محاسبه فاصله تاکسونومیکی بین گروه‌ها.



شکل ۳- نقشه رقومی خاک در سطح بزرگ خاک بر اساس مدل درخت تصمیم.

نتیجه‌گیری کلی

همچنین نشان داد که مهمترین متغیر کمکی شاخص خیزی و سطوح ژئومورفولوژی است. به‌طور کلی نتایج نشان داد که تکنیک‌های خاک‌آماری به ویژه درخت تصمیم می‌تواند فرآیند نقشه‌برداری خاک‌ها را در سطح وسیع و متشکل از هر گونه عوارض طبیعی به یک‌باره انجام دهد و روش سنتی را ارتقاء بخشیده، سرعت عمل و کارآمدی نقشه‌ها را در انتقال داده‌ها و اطلاعات افزایش داده و قابلیت استفاده آن‌ها را برای قشر وسیعی از شاخه‌های علمی امکان‌پذیر نماید. بنابراین پیشنهاد می‌گردد که جهت تهیه نقشه رقومی خاک از مدل‌های درختی در مطالعات آینده استفاده شود.

در این مطالعه از فن درخت تصمیم با دو رویکرد (حداقل کردن خطای طبقه‌بندی و حداقل کردن فاصله تاکسونومیک) جهت پیش‌بینی کلاس‌های خاک در منطقه اردکان استفاده گردید. نتایج نشان داد که با به کار بردن فاصله تاکسونومیکی که از آنالیز تشخیصی قابل استخراج هستند، دقت مدل افزایش پیدا کرده است. به‌طوری‌که مدل نهایی گروه‌های خاک با دقت ۷۰٪ پیش‌بینی کرده است. لازم به ذکر است که این دقت در بحث‌های نقشه‌برداری رقومی خاک قابل قبول است. همچنین وجود ارتباط قوی بین داده خاک و متغیرهای محیطی نیز یکی دیگر از مؤلفه‌های تأثیرگذار بر روی دقت است. نتایج

References

- [1]. Allison, G. W., & Bauder, J. W. (1986). Particle size analysis. p. 383-411, In: A. Klute. (ed), Methods of Soil Analysis. Part 1. American Society of Agronomy. Madison, WI.
- [2]. Andronikov, V. L., & Dorbrolov'skiy, G. V. (1991). Theory and methods for the use of remote sensing in the study of soils. *Remote Sensing*, 28, 92-101.
- [3]. Behrens, T., & Scholten, T. (2007). A comparison of data-mining techniques in predictive soil mapping. In: Lagacherie, P., McBratney, A. B., Voltz, M. (Eds.), Digital Soil Mapping. An Introductory Perspective. Developments in Soil Science. Elsevier, Amsterdam, 31, 353- 364.
- [4]. Brungard, C. W., & Boettinger, J. L. (2010). Conditioned Latin Hypercube Sampling: Optimal Sample Size for Digital Soil Mapping of Arid Rangelands in Utah, USA.
- [5]. Bui, E. N., Loughhead, A., & Corner, R. (1999). Extracting soil-landscape rules from previous soil surveys. *Australian Journal of Soil Research*, 37, 495- 508.

- [6].Bui, E. N., & Moran, C. J. (2001). Disaggregation of polygons of surficial geology and soil maps using spatial modelling and legacy data. *Geoderma*, 103, 79–94.
- [7].Du, C., Linker, R., Shaviv, A. (2008). Identification of agricultural soils using mid-infrared photo acoustic spectroscopy. *Geoderma*, 143, 85–90.
- [8].Elnaggar, A. A. (2007). Development of predictive mapping techniques for soil survey and salinity mapping. (Doctoral dissertation, Oregon State University, Corvallis, Oregon). Retrieved from <http://hdl.handle.net/1957-5754>.
- [9].Gee, G. W., & Bauder, J. W. (1986). Particle size analysis. p. 383-411, In: A. Klute. (ed), *Methods of Soil Analysis. Part 1*, American Society of Agronomy Madison, WI.
- [10].Grinand, C., Arrouays, D., Laroche, B., & Martin, M. P. (2008). Extrapolating regional soil landscapes from an existing soil map: Sampling intensity, validation procedures, and integration of spatial context. *Geoderma*, 143, 180–190.
- [11].Grunwald, S. (2009). Multi-criteria characterization of recent digital soil mapping and modeling approaches. *Geoderma*, 152, 195–207.
- [12].Hash, S. J. (2009). Use of Decision Tree Analysis for Predictive Soils Mapping and Implementation on the Malheur County, Oregon Initial Soil Survey. Oregon State University. 171pp.
- [13].Henderson, B. L., Bui, E. N., Moran, C. J., & Simon, D. A. P. (2005). Australia-wide predictions of soil properties using decision trees. *Geoderma*, 124, 383-398.
- [14].Hengl, T., Huvelink, G. B. M., & Stein, A. (2004). A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma*, 120, 75–93.
- [15].Hengl, T., Toomanian, N., Reuter, H., & Malakouti M. J. (2007). Methods to interpolate soil categorical variables from profile observations: Lessons from Iran. *Geoderma*, 140, 417-427.
- [16].Hole, F. D., & Hironaka, M. (1960). An experiment in ordination of some soil profiles. *Soil Science society of American Journal*, 24, 309–312.
- [17].Jafari, A., Finke, P. A., de Wauw, J. V., 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, 284–298.
- [18].Liu, J., Pattey, E., Nolin, M. C., Miller, J. R., & Ka, O. (2008). Mapping within-field soil drainage using remote sensing, DEM and apparent soil electrical conductivity. *Geoderma*, 143, 261–272.
- [19].McBratney, A. B., Odeh, I. O. A., Bishop, T. F. A., Dunbar, M. S., & Shatar, T. M. (2000). An overview of pedometric techniques for use in soil survey. *Geoderma*, 97, 293–327.
- [20].McBratney, A. B., Mendonça-Santos, M. L., & Minasny, B. (2003). On digital soil mapping. *Geoderma*, 117, 3–52.
- [21].McKenzie, N. J., & Ryan, P. J. (1999). Spatial prediction of soil properties using environmental correlation. *Geoderma*, 89, 67–94.
- [22].Mendonça-Santos, M. L., McBratney, A.B., & Minasny, B. (2006). Soil Prediction with Spatially Decomposed Environmental Factors. *Digital Soil Mapping- An Introductory Perspective*, 31, 269–278.
- [23].Metternicht, G., & Zinck J. A. (2003). Remote Sensing of Soil Salinity: Potentials and Constraints. *Remote Sensing Environment*, 85, 1-20.
- [24].Minasny B., & McBratney, A. B. (2006). A conditioned Latin hypercube method for sampling in the presence of ancillary information, *Geology*, 32, 1378-1388.
- [25].Minasny, B., & McBratney, A. B. (2007). Incorporating taxonomic distance into spatial prediction and digital mapping of soil classes. *Geoderma*, 142, 285–293.
- [26].Minasny, B., McBratney, A. B., Mendonca-Santos, M. L., Odeh, I. O. A., & Guyon, B. (2006). Prediction and digital mapping of soil carbon storage in the Lower Namoi Valley. *Australian Journal of Soil Research*, 44, 233–244.
- [27].Moore, I. D., Turner, A. K., Wilson, J. P., Jenson, S. K., & Band, L. E. (1993). GIS and landsurface-subsurface process modelling. In: Goodehild, M. F., Parks, B. & L. T. Steyaert, (Eds.), *Environmental Modeling with GIS*. Oxford University. Press, Oxford, PP. 196–230.

- [28]. Moran, C. J., & Bui, E. N. (2002). Spatial data mining for enhanced soil map modelling. *International Journal of Geo-Information Science*, 16, 533-549.
- [29]. Nelson, R. E. (1982). Carbonate and gypsum. In: A. L. Page, R. H. Miller and R. Keeny, *Methods of soil analysis, Part II-chemical and microbiological properties*, Madison, WI, PP. 181-196.
- [30]. Quinlan, J. R. (2001). Cubist: An Informal Tutorial. <http://www.rulequest.com>.
- [31]. Rossiter, D. G., & Hengl, T. (2001). Technical note: Creating geometrically-correct photo-interpretation, photo-mosaics, and base maps for a projects GIS. <http://www.itc.nl/rossiter>
- [32]. Roudier, P., Hewitt, A. E., & Beaudette, D. E. (2012). A conditioned Latin hypercube sampling algorithm incorporating operational constraints. *Digital Soil Assessments and Beyond*. Taylor and Francis Group, 227-231
- [33]. Saunders, A. M., & Boettinger, J. L. (2007). Incorporating classification trees into a pedogenic understanding raster classification methodology, Green River Basin, Wyoming, USA. In: P. Lagacherie, A. B. McBratney, and M. Voltz (eds.), *Digital Soil Mapping: An introductory perspective*. *Developments in Soil Science*. Elsevier, Amsterdam, 31, 389-399.
- [34]. Scull, P., Franklin, J., Chadwick, O. A. (2005). The application of classification of tree analysis to soil type prediction in a desert landscape. *Ecological Modelling*, 181, 1-15.
- [35]. Sparks, D. L., Page, A. L., Helmke, P. A., Leppert, R. H., Soltanpour, P. N., Tabatabai, M. A., Johnston, G. T., & Summer, M. E. (1996). *Methods of Soil Analysis*, SSSJA, Madison, Wisconsin.
- [36]. Thomas, M., Odgers, N. P., Ringrose-Voase, A., Grealish, G., Glover, M., & Dowling, T. (2012). Soil survey design for management-scale digital soil mapping in a mountainous southern Philippine catchment. *Digital Soil Assessments and Beyond*. Taylor and Francis Group, 233-238.
- [37]. Walkey, A., & Black, I. A. (1934). An examination of Degtjareff method for determining soil organic matter and a proposed modification of the chromic acid in soil analysis. 1. Experimental. *Soil Science society of American Journal*, 79, 459-465.
- [38]. Zhu, A. X., Hudson, B., Burt, J., Lubich, K., & Simonson, D. (2001). Soil mapping using GIS, expert knowledge, and fuzzy logic. *Soil Science Society of American Journal*, 65, 1463-1472.

Incorporating soil taxonomic distance and decision tree for spatial prediction of soil classes in Ardakan, Yazd

- 1- R. Taghizadeh-Mehrjardi, Assistant Faculty of Agriculture and Natural Resources, University of Ardakan
- 2- F. Sarmadian, Professor University College of Agriculture & Natural Resources, University of Tehran
fsarmad@ut.ac.ir
- 3- M. Omid, Professor University College of Agriculture & Natural Resources, University of Tehran
- 4- N. Toomanian, Assistant Agricultural and Natural Resources Research Center, Isfahan
- 5- M. J. Rousta, Assistant Member of National Salinity Research Center
- 6- M. H. Rahimian, Member of National Salinity Research Center

Received: 25 Aug 2013

Accepted: 01 Mar 2014

Abstract

Mapping soil classes digitally generally starts with soil profile description with observed soil classes at a taxonomic level in a particular classification system. At each soil observation location there is a set of co-located environmental variables and a challenge is to correlate soil classes with environmental variables. The current methodology treats soil classes as 'labels' and their prediction only considers the minimization of the misclassification error. Soil classes at any taxonomic level have taxonomic relationships between each others. Using classification trees, we can specify an algorithm that minimises the taxonomic distance rather than misclassification error. Therefore, in this research, we have attempted to develop decision tree model for spatial prediction of soil taxonomic classes in an area covering 720 km² located in arid region of central Iran. In this area, using the conditioned Latin hypercube sampling method, location of 187 soil profiles were selected, which then described, sampled, analyzed and allocated in taxonomic classes according to soil taxonomy of America. Auxiliary data used in this study to represent predictive soil forming factors were terrain attributes, Landsat 7 ETM+ data and a geomorphologic surfaces map. Discriminant analysis was applied to calculate taxonomic distances. Results showed using the taxonomic distances led to achieve overall accuracy up to 70%. Results also showed some auxiliary variables had more influence on predictive soil class model which included: wetness index, geomorphology map and multi-resolution index of valley bottom flatness. General results showed that incorporating taxonomic distance into decision tree model had reliable accuracy. Therefore, it is suggested using of decision tree model with taxonomic distance for spatial prediction of soil classes in the future studies.

Keywords: Decision tree; Discriminant analysis; Overall accuracy; Ardakan.