



دانشگاه گیلان

نشریه پژوهش‌های حفاظت آب و خاک
جلد بیست و چهارم، شماره دوم، ۱۳۹۶
<http://jwsc.gau.ac.ir>

ارزیابی عملکرد رگرسیون چندمتغیره، شبکه عصبی مصنوعی و برنامه‌ریزی بیان ژن در تخمین برخی خصوصیات خاک

ابراهیم محمودآبادی^۱، علیرضا کریمی^۲، غلامحسین حق‌نیا^۳ و عادل سپهر^۴

^۱ دانشجوی دکتری گروه علوم خاک، دانشکده کشاورزی، دانشگاه فردوسی مشهد، دانشیار گروه علوم خاک،
^۲ دانشکده کشاورزی، دانشگاه فردوسی مشهد، ^۳ استاد گروه علوم خاک، دانشکده کشاورزی، دانشگاه فردوسی مشهد،
^۴ استادیار گروه مدیریت مناطق خشک و بیابانی، دانشکده منابع طبیعی و محیط زیست، دانشگاه فردوسی مشهد
تاریخ دریافت: ۹۵/۶/۲۷؛ تاریخ پذیرش: ۹۶/۳/۲۹

چکیده

سابقه و هدف: با ظهور سیستم‌های کامپیوتری در کنار سامانه اطلاعات جغرافیایی و دسترسی به داده‌های رقمی مکانی، روش‌های مختلف داده‌کاوی، مدل‌سازی و تخمین خصوصیات خاک، جایگاه خود را در علوم خاک و پدومتری باز کرده است. داده‌کاوی خصوصیات خاک با استفاده از روش‌های آماری کامپیوتر-محور به کشف الگوهای پنهان در بانک اطلاعاتی داده‌ها می‌پردازد که در نهایت منجر به برازش مدل به‌منظور استفاده و تخمین خصوصیات خاک می‌گردد. یکی از کاربردهای مهم این روش‌ها استفاده در معادله اسکورپن می‌باشد. دو جزء اصلی معادله اسکورپن شامل متغیرهای محیطی و برنامه یادگیری می‌باشد. این مطالعه با هدف ارزیابی و مقایسه سه مدل عددی شامل روش رگرسیون چندگانه خطی، شبکه عصبی مصنوعی و برنامه‌ریزی بیان ژن به‌عنوان برنامه یادگیری (تابع f) در معادله اسکورپن با استفاده از داده‌های دورسنجی، توپوگرافی و پوشش گیاهی به‌عنوان داده‌های کمکی به‌منظور تخمین خصوصیات خاک از جمله کربنات کلسیم معادل، رس، چگالی ظاهری، نیتروژن کل، کربن آلی، شن، سیلت و ظرفیت رطوبت اشباع انجام گرفت.

مواد و روش‌ها: این پژوهش، در مراتع بخش باجگیران در استان خراسان رضوی و با مساحت ۱۲۲۵ هکتار انجام شد. به‌منظور بررسی پوشش گیاهی و خاک، تعداد ۱۳۷ واحد مورد بررسی قرار گرفت. در هر واحد کاری ۳ تا ۵ پلات با فاصله ۱۰ متر و در امتداد یک برش انتخاب شدند و نوع و تعداد گونه گیاهی و درصد پوشش گیاهی درون پلات‌ها ثبت گردید. سپس یک نمونه خاکی در هر برش و در مجموع ۱۳۷ نمونه خاکی از سطح منطقه مورد مطالعه برداشته شد. داده‌های توپوگرافی منطقه از نقشه DEM، داده‌های طیفی و باندهای مختلف از تصاویر سنجنده ETM و شاخص‌های تنوع گیاهی و درصد پوشش گیاهی اندازه‌گیری شد و به‌عنوان متغیرهای کمکی در پیش‌بینی کربنات کلسیم معادل، رس، چگالی ظاهری، نیتروژن کل، کربن آلی، شن، سیلت و ظرفیت رطوبت اشباع به‌کار گرفته شدند. به‌منظور کاهش تعداد داده ورودی در شبکه عصبی مصنوعی و برنامه‌ریزی بیان ژن از نتایج PCR استفاده گردید سپس عملیات نرمال‌سازی و استانداردسازی بر روی داده‌ها صورت گرفت.

* مسئول مکاتبه: karimi-a@um.ac.ir

یافته‌ها: نتایج حاصل از ارزیابی مدل‌های رگرسیون چندگانه، شبکه عصبی مصنوعی و برنامه‌ریزی بیان ژن براساس آماره‌های ارزیابی شامل میانگین اریبی خطا (MBE)، ریشه میانگین مربعات خطا (RMSE) و ضریب تبیین (R^2) در فاز آزمون نشان داد که شبکه عصبی مصنوعی پرسپترون، با توجه به مقادیر ضریب تبیین بالاتر برای کربنات کلسیم، رس، نیتروژن کل، کربن آلی، شن، سیلت، ظرفیت رطوبتی و چگالی ظاهری به ترتیب با مقادیر ۰/۷۷، ۰/۶۷، ۰/۴۶، ۰/۷۲، ۰/۶۲، ۰/۷، ۰/۸۵ و ۰/۶۹ و همچنین مقادیر خطای RMSE کم‌تر با مقادیر به ترتیب ۰/۴۶، ۰/۰۳، ۰/۲۷، ۰/۵/۶، ۳/۵۵ و ۳/۴ درصد برای کربنات کلسیم معادل، درصد رس، نیتروژن کل، کربن آلی، درصد شن، درصد سیلت، ظرفیت رطوبت اشباع و ۰/۰۸ گرم بر سانتی‌مترمکعب برای چگالی، بهترین نتایج را از بین روش‌های مورد مقایسه نشان داد. روش شبکه عصبی مصنوعی توانست ۶۰ تا ۸۵ درصد تغییرپذیری خصوصیات مورد بررسی را نشان دهد که از بین خصوصیات مختلف، بهترین تخمین برای ظرفیت رطوبت اشباع خاک با $R^2=0/85$ و کربن آلی با $R^2=0/77$ بود.

نتیجه‌گیری: نتایج ارزیابی تخمین خصوصیات خاک از طریق سه مدل عددی که بهترین نتایج به دست آمده برای مدل شبکه عصبی مصنوعی پرسپترون به دست آمد. نتایج اعتبارسنجی مدل شبکه عصبی مصنوعی نشان داد که مقدار MBE مدل برای متغیرها نزدیک به صفر بوده و این امر بیانگر این مطلب است که برازش، توسط مدل ایجاد شده ناریب بوده است. مقدار RMSE پایین مدل نیز نشان‌دهنده دقت مناسب و قابل قبول برآورد برای متغیرهای خاک می‌باشد. نتایج الگوریتم بیان ژن نیز بیانگر دقت بالاتر این روش نسبت به رگرسیون خطی برای اکثر خصوصیات خاک بود.

واژه‌های کلیدی: شبکه عصبی مصنوعی، سنجش از دور، مدل اسکورپن، الگوریتم بیان ژن، مدل‌سازی خاک

مقدمه

داده‌کاوی^۱ یک حوزه میان‌رشته‌ای و با رشد سریع است که حوزه‌های گوناگونی مانند پایگاه داده، آمار، یادگیری ماشین و سایر زمینه‌های مرتبط را با هم تلفیق می‌کند تا به استخراج اطلاعات نهان و یا الگوها و روابط مشخص در حجم زیادی از داده‌ها در یک بانک اطلاعاتی بزرگ دسترسی یابد. مدل‌های آماری، الگوریتم‌های ریاضی و ماشین یادگیری^۲ می‌توانند به منظور کشف و استخراج اطلاعات و دانش نهفته در حجم بزرگی از داده‌ها مورد استفاده قرار گیرند. در واقع داده‌کاوی منحصر به گردآوری و مدیریت داده‌ها نبوده و تجزیه و تحلیل اطلاعات و پیش‌بینی را نیز شامل می‌شود (۲۰). در علوم خاک، به دلیل توسعه سامانه‌های اطلاعاتی جغرافیایی، دسترسی داده‌های رقمی مکانی و رشد و توسعه تکنیک‌های رایانه‌ای،

روش‌های داده‌کاوی و ماشین‌های یادگیری در مجموعه پدومتری مورد استفاده قرار گرفته است (۲۲). ارتباط بین متغیرهای محیطی و خاک در نقاط اندازه‌گیری شده با استفاده از روش داده‌کاوی کشف می‌شود سپس از این مدل توسعه یافته به منظور پیش‌بینی مقادیر خصوصیات خاک در نواحی که اطلاعات خاک در دسترس نیست، مورد استفاده قرار می‌گیرد (۲۲). در سال‌های اخیر پژوهشگران زیادی برای شناسایی خاک و مدل کردن تغییرات مکانی خاک با استفاده از متغیرهای محیطی، از روش‌های مختلف داده‌کاوی بهره گرفته‌اند (۲، ۹، ۳۸ و ۴۶). یک از کاربردی‌ترین روش‌های تخمین ویژگی‌های خاک استفاده از مدل اسکورپن (رابطه ۱) می‌باشد (۲۹) که در این مدل ویژگی‌ها یا کلاس خاک با یک سری متغیرهای محیطی ارتباط داده می‌شود به این صورت که در هر موقعیت مشاهده خاک، یک سری متغیرهای محیطی وجود دارد که همبستگی بالایی با

1- Data mining
2- Machine learning methods

ویژگی‌ها یا کلاس‌های خاک دارند که با یافتن این ارتباط و تبدیل آن به مدل منجر به پیش‌بینی خصوصیات در نقاط فاقد نمونه‌برداری با استفاده از داده کمی می‌گردد.

$$S_a = f(s, c, o, r, p, a, n) \quad (1)$$

که در آن، S_a خصوصیات خاک، s مربوط به اطلاعات خاک که از نقشه یا پایگاه داده یا دانش کارشناس به دست می‌آید. c اقلیم، o موجودات زنده، r پستی و بلندی، p مواد مادری، a زمان و در نهایت n موقعیت مکانی می‌باشد. تابع f یک برنامه یادگیری نظارت شده را نشان می‌دهد. بنابراین، دو جز اصلی معادله اسکورپن شامل متغیرهای محیطی و برنامه یادگیری می‌باشد. از مدل‌های به کار برده شده به عنوان تابع f در معادله اسکورپن می‌توان به گستره زیادی از مدل‌ها اشاره نمود. نگاهی به منابع علمی نشان می‌دهد که روش‌های رگرسیون خطی و شبکه عصبی مصنوعی در مطالعات زیادی به عنوان تابع تخمین به کار گرفته شده‌اند؛ به طور مثال، لئو و همکاران (۲۰۰۸) در پیش‌بینی کربن آلی از مدل رگرسیونی و داده‌های سنجش از دور استفاده نمودند که توانست ۷۶ درصد از تغییرات کربن آلی را توجیه کند (۲۷). محمودآبادی و همکاران (۲۰۱۵) با استفاده از معادله رگرسیونی چندمتغیره توانستند ۵۲ درصد از تغییرات رس را تبیین نمایند (۲۸). کلومن و همکاران (۱۹۹۳) در مطالعه خود مقدار ضریب تبیین ۰/۱ تا ۰/۴ را برای معادلات تخمین مقادیر رس، شن و سیلت گزارش نمودند (۱۳). همچنین رگرسیون چندگانه در مطالعه دمانته (۲۰۰۹) برای پیش‌بینی چندین ویژگی خاک مورد استفاده قرار گرفت (۱۵). بن-دور و همکاران (۱۹۹۵) از روش رگرسیونی به منظور تخمین خصوصیات خاک از جمله ظرفیت تبادل کاتیونی، کربن آلی، رس و کربنات کلسیم استفاده نمودند (۹). کاربرد شبکه عصبی مصنوعی نیز در مطالعاتی از

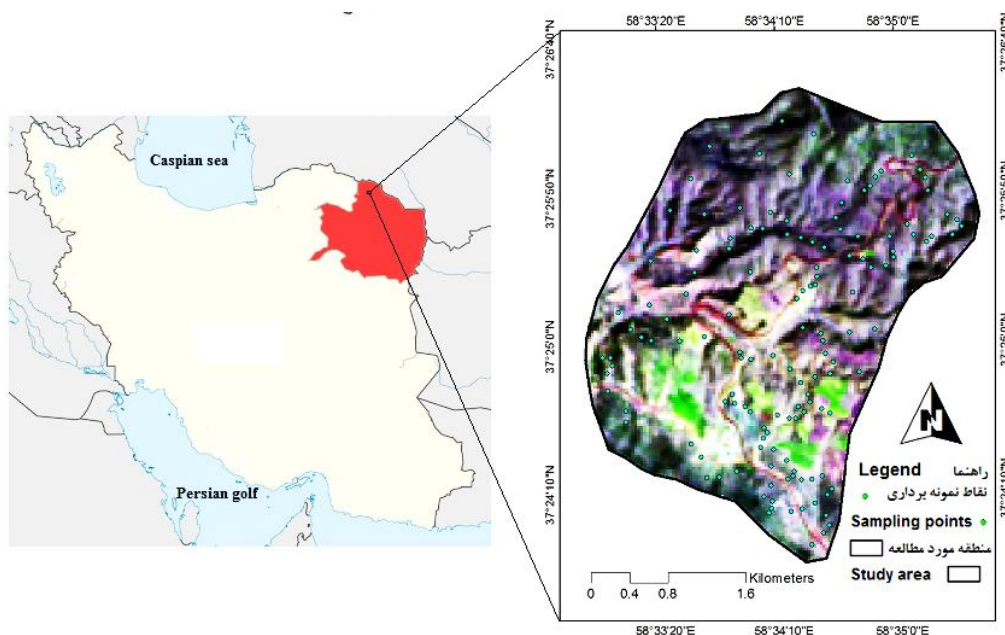
جمله باقری و همکاران (۲۰۱۵) با استفاده از داده‌های توپوگرافی با استفاده از شبکه عصبی مصنوعی به مدل‌سازی و تهیه نقشه رقومی واحدهای خاک پرداختند (۸). مدل‌سازی و تخمین مقدار کربن، چگالی و عمق خاک از طریق شبکه عصبی مصنوعی در اسکاتلند توسط ایکنهد و همکاران (۲۰۱۶) به طور موفقیت‌آمیزی استفاده شد (۲). ایکیناد و همکاران (۲۰۱۳) نیز به منظور تخمین خصوصیات از طریق رنگ خاک از مدل شبکه عصبی مصنوعی به عنوان بهترین مدل عددی تخمین استفاده نمودند (۱). مدل شبکه عصبی (ساختار پرسپترون چندلایه) به عنوان مدلی مناسب به منظور تخمین کربن آلی در مطالعه پرویزی و همکاران (۲۰۱۰) ارزیابی شد (۳۵). از سوی دیگر، اخیراً مدل‌های الگوریتم ژنتیک نیز مورد توجه برخی کارشناسان منابع طبیعی و خاک‌شناسی قرار گرفته است و به طور بسیار محدودی در برخی از مطالعات خاک به کار گرفته شده است. از جمله این مطالعات می‌توان به مکان‌یابی نقاط با آلودگی بالا به فلزات سنگین با استفاده از الگوریتم ژنتیک اشاره کرد (۴۹). همچنین امام‌قلی‌زاده و همکاران (۲۰۱۵) برای تخمین ظرفیت تبادل کاتیونی خاک از برنامه‌ریزی بیان ژن به عنوان تابع f بهره بردند (۱۶). تقی‌زاده و همکاران (۲۰۱۶) با استفاده از برنامه‌ریزی ژنتیک به مدل‌سازی و تخمین شوری سطحی خاک در منطقه مرکزی ایران پرداختند (۴۵). با این وجود، مطالعات در حوزه مقایسه مدل‌های عددی به منظور استفاده در معادله اسکورپن محدود می‌باشد (۱۱، ۱۲، ۱۴، ۲۶ و ۳۶). به طور مثال هوانگ و همکاران (۲۰۱۶) به مقایسه مدل‌های عددی مختلف به منظور طبقه‌بندی در تهیه نقشه رقومی خاک پرداختند (۲۲). همچنین، کاوازی و همکاران (۲۰۱۳) با هدف توزیع مکانی ساختارهای زمین‌نما مدل‌های عددی مختلف را مقایسه نمودند (۱۲). در همین راستا پریوری و همکاران (۲۰۱۴) در تخمین بافت خاک به مقایسه دو

۳۵' ۵۸° شرقی و عرض‌های جغرافیایی ۲۳' ۳۷° تا ۲۶' ۳۷° شمالی، در محدوده ارتفاعی ۱۸۰۰ تا ۲۲۰۰ متر از سطح دریا قرار گرفته است (شکل ۱). وسعت منطقه حدود ۱۲۳۵ هکتار می‌باشد. میانگین بارش سالانه ۳۶۵ میلی‌متر با چیرگی بارش در فصل سرد و میانگین دمای سالانه ۱۴/۳ درجه سلسیوس می‌باشد (سازمان هواشناسی کشور). اقلیم منطقه طبق روش آمبرژه نیمه‌خشک سرد و مطابق ضریب خشکی دوماراتن نیمه‌خشک است (۳۳). منطقه مورد مطالعه در ناحیه کپه‌داغ واقع می‌باشد. این ناحیه شمال خراسان و بخشی از ترکمنستان را در بر می‌گیرد. این ناحیه پس از کوهزایی سیمیرین پیشین به صورت حوضه رسوبی در می‌آید که در آن از دوره ژوراسیک تا میوسن، رسوباتی با ضخامت حدود ۸۰۰۰ متر بدون چینه‌شناسی مهمی روی هم انباشته گردیده است. تمام این رسوبات به‌طور دگرشیب در زیر کنگلومرای پلیوسن قرار دارد و بر روی این کنگلومرا، تراس‌های قدیمی آبرفتی قرار گرفته است (۳۳).

روش شبکه عصبی مصنوعی و مدل SVM پرداختند (۳۷). از سوی دیگر، مقایسه و کارایی مدل‌های ژنتیکی در مقایسه با روش‌های پرکاربرد، کم‌تر مورد توجه قرار گرفته است. بنابراین، این پژوهش با هدف ارزیابی عملکرد مدل‌های عددی مختلف شامل روش‌های رگرسیون چندگانه خطی و شبکه عصبی مصنوعی به‌عنوان روش‌های متداول و روش برنامه‌ریزی بیان ژن، که اخیراً مورد توجه پژوهشگران قرار گرفته، به‌منظور پیش‌بینی خصوصیات فیزیکی و شیمیایی خاک در ناحیه مرتعی از طریق معادله اسکورپن و با استفاده از داده‌های کمکی به‌دست آمده از نقشه رقومی ارتفاع، خصوصیات پوشش گیاهی و تصاویر ماهواره‌ای انجام شد.

مواد و روش‌ها

منطقه مورد مطالعه، مطالعات میدانی و نمونه‌برداری: منطقه مورد مطالعه در این پژوهش مراتع بخش باجگیران از توابع شهرستان قوچان در استان خراسان رضوی و در محدوده طول‌های جغرافیایی ۳۲' ۵۸° تا

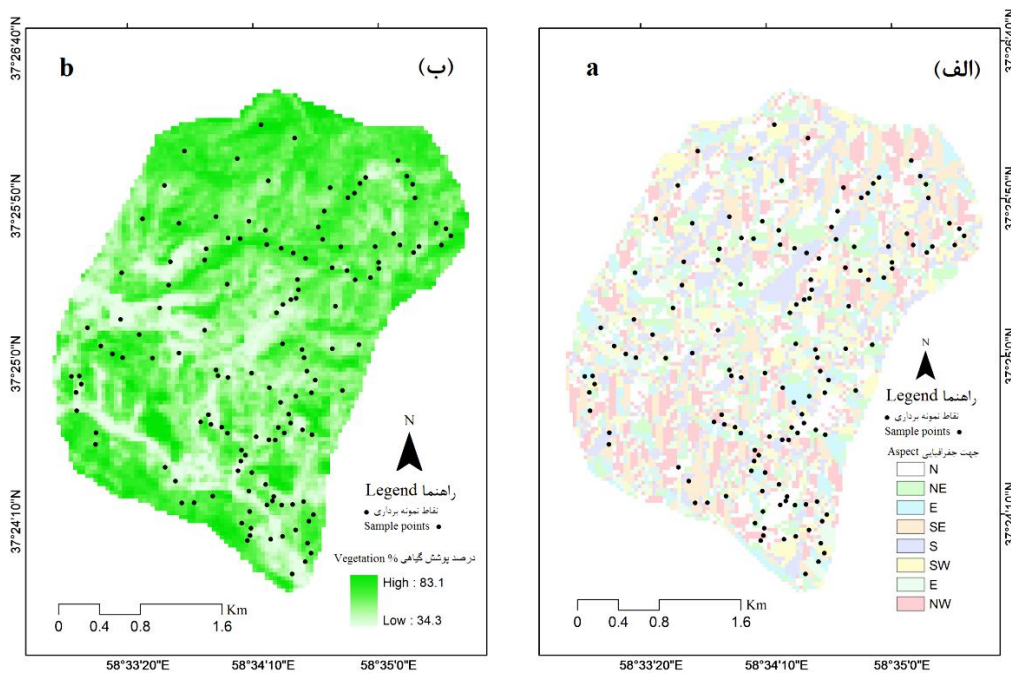


شکل ۱- نقشه موقعیت منطقه به همراه نقشه کاذب رنگی محدوده مورد مطالعه.

Figure 1. Location of the study area as well as false color composite images.

(شکل ۲). در محل‌های انتخاب شده، به‌منظور بررسی جامعه گیاهی از روش ترانسکت استفاده شد بدین‌ترتیب که در هر واحد کاری یک برش ۳۰ تا ۵۰ متری به‌صورت تصادفی انتخاب شد و سپس در امتداد هر برش پنج پلات ۳ تا ۵ پلات یک مترمربعی به‌صورت تصادفی انتخاب شد و نوع و تعداد گونه گیاهی در هر یک از پلات‌ها ثبت شد. سپس یک نمونه خاکی در هر برش به‌صورت تصادفی و از عمق ۰-۱۰ سانتی‌متر برداشته شد. طول و عرض جغرافیایی این نقاط به دستگاه GPS معرفی شدند. در مجموع تعداد ۱۳۷ برش و در نتیجه ۱۳۷ نمونه خاک در سطح منطقه مورد بررسی قرار گرفت.

به‌منظور نمونه‌برداری خاک و گیاه از سطح منطقه، ابتدا با استفاده از نقشه توپوگرافی منطقه، طبقه‌های ارتفاعی، راستای شیب و درجه شیب در محیط نرم‌افزار GIS جدا شدند. به این طریق که با استفاده از نقشه DEM ابتدا نقشه جهت شیب (شمال، شمال‌شرق، شرق، جنوب، جنوب‌غرب، غرب و شمال‌غربی) به‌دست آمد. سپس نقشه درجه شیب را با سه گروه مختلف درجه شیب تهیه شد. همچنین نقشه ارتفاع نیز با دو زیرمجموعه تهیه گردید. سپس با توجه به نقشه‌های خصوصیات توپوگرافی به‌دست آمده (شکل ۲ الف) و همچنین تغییرات پوشش گیاهی در سطح منطقه در پیمایش صحرائی، محل‌های مورد نمونه‌برداری تعیین گردید



شکل ۲- محل نقاط نمونه‌برداری بر روی الف) نقشه جهت شیب؛ ب) نقشه درصد پوشش گیاهی.

Figure 2. Sampling pattern on a) aspects and b) vegetation cover maps.

می‌باشند. با توجه به این‌که در این مطالعه تعداد ۸ ویژگی خاک مورد بررسی قرار می‌گیرد و همچنین هر کدام از ویژگی‌های مختلف خاک انعکاس (طول موج) ویژه خود را دارند بنابراین استفاده از شاخص‌های طیفی متفاوت می‌تواند احتمال پیدا کردن شاخص طیفی مناسب که در تخمین ویژگی خاک نقش مؤثرتری داشته باشد را افزایش می‌دهد. جدول ۱ شاخص‌های طیفی خاک و گیاهی استفاده شده در این مطالعه را نشان می‌دهد.

مشخصات زمین^۳: در ابتدا، مدل رقومی ارتفاع (DEM^۴) با قدرت وضوح افقی ۳۰ در ۳۰ متر تهیه شده از نقشه خطوط تراز با وضوح عمودی ۱۰ متر از منطقه مورد مطالعه از سازمان نقشه‌برداری کشور تهیه گردید. در مرحله بعد، مطابق روش هنگل و همکاران (۲۰۰۹) نقشه‌های ارتفاع و شیب، جهت شیب، طول شیب، انحنای شیب، همگرایی، شاخص خیسی، انحنای افقی، انحنای عمودی از مدل رقومی ارتفاعی توسط نرم‌افزار SAGA محاسبه گردید. مبنای محاسبه شاخص‌های مذکور ایجاد یک شبکه سلولی ۳×۳ بر روی تصویر بود (۲۱).

پوشش گیاهی: بررسی پوشش گیاهی در درون پلات‌ها انجام گرفت بدین‌صورت که در داخل هر پلات درصد پوشش گیاهی، نوع و تعداد گونه گیاهی ثبت شد. با توجه به برهمکنش خصوصیات خاک و پوشش گیاهی، درصد پوشش گیاهی به‌عنوان شاخص وضعیت پوشش گیاهی در مجموعه متغیرهای کمکی در مدل‌های تخمین خصوصیات خاک به‌کار گرفته شد.

آنالیز خاک: پس از این‌که نمونه‌های خاک در هوای آزاد خشک و از الک دو میلی‌متری عبور داده شدند، ویژگی‌های فیزیکی و شیمیایی شامل کربن آلی خاک به روش والکی-بلاک، ظرفیت رطوبت اشباع به روش وزنی، نیتروژن کل خاک با استفاده از روش کدال، بافت خاک به روش هیدرومتری، چگالی خاک به روش کلوخه و کربنات کلسیم معادل به روش تیتراسیون اندازه‌گیری شدند (۳۴).

داده‌های کمکی: داده‌های کمکی شامل سه دسته داده‌های دورسنجی، مشخصات زمین و وضعیت پوشش گیاهی بودند. که در ادامه روش محاسبه آن‌ها شرح داده شده است.

پردازش تصاویر ماهواره‌ای و محاسبه شاخص‌های باندی پوشش گیاهی و خاک: از داده‌های سنجنده ETM⁺ ماهواره لندست ۷ به شماره گذر ۱۶۰ و ردیف ۳۴ مربوط به تاریخ ۲۷ ژوئن ۲۰۱۴ (LE71600342014178SG100) و از باندهای ۱، ۲، ۳، ۴، ۵ و ۷ با قدرت تفکیک مکانی ۳۰ متر با توجه به هم‌زمانی فصلی تصاویر با زمان عملیات صحرائی، قدرت تفکیک طیفی^۱ و قدرت تفکیک زمینی^۲ مناسب، استفاده شد. به‌منظور تصحیح هندسی تصاویر از نقشه DEM منطقه و همچنین استفاده از نقاط کنترل زمینی با استفاده از GPS استفاده گردید.

تعدادی از مهم‌ترین شاخص‌های طیفی خاک و پوشش گیاهی، با توجه به درجه اهمیت و کاربرد در سایر مطالعه‌های خاکشناسی و مطالعات پوشش گیاهی انتخاب شدند. که شامل شاخص‌های NDVI، RDVI، SAVI، DVI، MSAVI، Brightness index، Clay index، COSRI، Carbonate index بودند. این شاخص شامل ۴ شاخص طیفی خاک و ۶ شاخص طیفی گیاهی

3- Terrain attributes
4- Digital elevation model

1- Spectral resolution
2- Spatial resolution

جدول ۱- معادلات محاسبه شاخص های طیفی.

Table 1. Calculating equation of spectral indices.

شماره (No)	نام شاخص (Indices)	رابطه ریاضی (Equation)	توضیحات (Description)	منبع (References)
1	Normalized Difference Vegetation Index (NDVI)	$NDVI = NIR - RED / NIR + RED$	نمایش تغییرات زیستوده سبز	(۷)
2	Soil adjusted vegetation index (SAVI)	$SAVI = \{(NIR - R) / (NIR + R + 0.5)\} \cdot 1.5$	کاهش نوزهای خاک در بررسی پوشش گیاهی	(۷)
3	Modified soil adjusted vegetation index (MSAVI)	$MSAVI = 0.5 \left[2NIR + 1 - \sqrt{((2NIR + 1)^2 - 8(NIR - RED))} \right]$	اصلاح شده SAVI	(۷)
4	Renormalized Difference Vegetation Index (RDVI)	$RDVI = NIR - Red / \sqrt{NIR + Red}$	خطی سازی ارتباط پارامترهای سطحی	(۷)
5	Differenced Vegetation Index (DVI)	$DVI = NIR - R$	نماینده خاک و گیاه	(۷)
6	Soil and Atmospherically Resistant Vegetation Index (SARVI)	$SARVI = (1 + L) \times (P_{NIR} - P_{60}) / (P_{NIR} - P_{60} - L)$	حذف نوزهای خاکی و اتمسفری	(۷)
7	Combined Spectral Response Index	$COSRI = (ETM1 + ETM2) / (ETM3 + ETM4) * NDVI$	ترکیب پاسخ طیفی خاک و گیاه	(۷)
8	Brightness Index (BI ²)	$BI^2 = ((B3)^2 + (B4)^2)^{0.5}$	شاخص روشنایی سطح	(۳۰)
9	Clay index	$B5/B7$	شاخص کانی های رسی	(۱۰)
10	Carbonate index	$B3/B2$	شاخص کربنات های خاک	(۱۰)

متغیرهای مستقل (X ها) نمی‌باشد. مدل‌های زیادی وجود دارند که (Y) را به صورت غیرخطی به (X ها) ارتباط می‌دهند و می‌توان آن‌ها را تا زمانی که معادله از نظر β ها خطی باشد، با مدل‌های رگرسیون خطی بیان نمود. این مدل‌ها اگر تنها دارای یک متغیر مستقل باشند، ساده (رابطه ۳) و اگر دارای بیش از یک متغیر مستقل باشند (رابطه ۴)، چندمتغیره نامیده می‌شوند.

$$Y = \beta_0 + \beta_1 X_1 + e \quad (3)$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + e \quad (4)$$

که در آن‌ها، (Y) متغیر وابسته، X_1, \dots, X_n متغیرهای مستقل و β_0, \dots, β_n پارامترهای مدل هستند (۳۲).

در این مطالعه تمامی داده‌های محیطی شامل داده‌ها و شاخص‌های گیاهی، مشخصات پستی و بلندی و داده‌های توپوگرافیکی و همچنین مقادیر انعکاسی باندها و شاخص‌های طیفی مستخرج از تصاویر ماهواره‌ای، به‌عنوان متغیر مستقل و خصوصیات خاک شامل کربن آلی، رس، شن، سیلت، نیتروژن، کربنات کلسیم معادل، ظرفیت نگه‌داری رطوبت اشباع و چگالی خاک به‌عنوان فاکتورهای وابسته در نظر گرفته شدند. آنالیز رگرسیون چندمتغیره خطی بین متغیرهای وابسته (خصوصیات خاک) و متغیرهای مستقل (داده‌های محیطی) به روش رگرسیون گام به گام^۴ و در محیط نرم‌افزاری SPSS 19 صورت گرفت.

شبکه‌های عصبی مصنوعی^۵: تئوری شبکه عصبی مصنوعی برگرفته از مدل سلول‌های عصبی مغز انسان است. مغز از تعداد زیادی سلول عصبی یا نرون تشکیل شده است. در این تئوری ورودی نرون‌ها در

کمینه‌سازی داده‌های ورودی^۱ و استانداردسازی: برای گزینش مجموعه حداقلی و تعیین مهم‌ترین ویژگی‌های و شاخص‌های مؤثر بر هر یک از معیارهای خاک برای استفاده در شبکه عصبی مصنوعی و الگوریتم بیان ژن از روش تجزیه مؤلفه‌های اصلی (PCA^۲) استفاده شد (۵). در مرحله اول مؤلفه‌های با مقدار ویژه بالاتر از یک ($\text{eigenvalues} \geq 1$) به همراه توضیح حداقل ۵ درصدی از واریانس کل ($\text{PCs variance} > 5\%$) مورد انتخاب قرار گرفتند. سپس در مرحله بعد در داخل هر یک از مؤلفه‌های اصلی، بردارهای ویژه^۳ با مقدار بیش‌تر از ۶۰ درصد به‌عنوان متغیرهای مؤثر در خروجی، وارد پایگاه داده ورودی گردید (۴۸). به‌منظور استاندارد سازی داده‌های ورودی و خروجی برای استفاده در مدل‌های شبکه عصبی مصنوعی و الگوریتم بیان ژن، همه داده‌های متغیرها با استفاده از رابطه ۲ استاندارد شده و بین یک دامنه عددی ۰/۱ تا ۰/۹ قرار گرفتند. در این معادله X_{\max} و X_{\min} به‌ترتیب کوچک‌ترین و بزرگ‌ترین داده در هر سری داده ورودی می‌باشد.

$$y = 0.8 \left(\frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \right) + 0.1 \quad (2)$$

مدل‌سازی

تجزیه رگرسیونی: تجزیه رگرسیون روشی آماری برای بررسی و مدل‌سازی رابطه میان متغیرهاست (۳۲). مدل‌های رگرسیونی گوناگونی، مانند مدل رگرسیونی ساده، چندمتغیره و غیرخطی وجود دارند. در مدل‌های رگرسیونی خطی، پارامترهای مدل خطی هستند و الزاماً متغیر وابسته (Y) تابعی خطی از

- 1- Data reduction
- 2- Principle component analysis
- 3- Loading factor

- 4- Stepwise
- 5- Artificial neural network

خروجی خصوصیات خاک بودند، استفاده گردید. توابع انتقال و تعداد نرون در لایه پنهان مورد آزمایش قرار گرفت و بهترین ساختار شبکه برای هر یک از متغیرها به دست آمد. انتخاب بهترین شبکه بر اساس کمترین میزان RMSE و بالاترین R^2 بود. همه مراحل در نرم افزار MATLAB انجام گرفت.

برنامه ریزی بیان ژن: این روش جزو روش های الگوریتم گردشی محسوب می شود که مبنای تمامی آنها بر اساس نظریه تکامل داروین استوار است. الگوریتم بیان ژن یکی از جدیدترین شیوه ها از بین الگوریتم های گردشی است که به تعریف یک تابع هدف می پردازد و در گام بعد از این تابع برای اندازه گیری و مقایسه روش های مختلف حل، در یک فرآیند گام به گام، تصحیح ساختار داده ها را به کار می گیرد تا در نهایت روش حل مناسب را ارائه نماید. زمینه اصلی برنامه ریزی بیان ژن همان الگوریتم ژنتیک است با این تفاوت که در این روش از شاخه های مجزا به جای نوارهای بیت استفاده می شود. در واقع روش برنامه ریزی بیان ژن ترکیبی از روش های برنامه ریزی ژنتیک و الگوریتم ژنتیک بوده که در آن کروموزوم های خطی و ساده با طول ثابت مشابه آنچه در الگوریتم ژنتیک استفاده می شود و ساختارهای شاخه ای با اندازه ها و اشکال متفاوت مشابه با برنامه ریزی ژنتیک ترکیب می شوند. در GEP بهسازی ها در یک ساختار خطی اتفاق افتاده و سپس به صورت ساختار درختی بیان می شود و این موجب می شود تنها ژنوم اصلاح شده به نسل بعد منتقل شده و نیازی به ساختارهای سنگین برای تکثیر و جهش وجود نداشته باشد. در این روش، پدیده های مختلف با استفاده از مجموعه ای از توابع و مجموعه ای از

وزن های ثابت ضرب و سپس با یک عدد ثابت به نام بایاس جمع می شود و نتیجه از یک تابع خط عبور می کند تا خروجی حاصل شود. یکی از معروف ترین انواع شبکه عصبی مصنوعی شبکه پرسپترون چندلایه (MLP^1) می باشد. این شبکه شامل یک لایه ورودی، یک یا چند لایه پنهانی و یک لایه خروجی است. در این شبکه تعداد لایه پنهان و تعداد نرون های هر لایه ثابت نمی باشد. هر لایه ممکن است از نرون های مختلفی تشکیل شده باشد که این موضوع به نوع کار آنها بستگی دارد. کاربر باید تعیین کند که برای کاربرد مورد نظرش، چند لایه و چند نرون در هر لایه باید انتخاب گردد. اساس آموزش شبکه های عصبی، بر مبنای روش آزمون و خطا می باشد تا بهترین آرایش لایه پنهان با تغییر تعداد لایه پنهان و نرون های آنها، نوع تابع محرک، الگوریتم آموزشی و تعداد تکرار مرحله آموزش جهت برآورد پارامتر خروجی مورد نظر، ارائه شود. معیار مناسب بودن آرایش لایه پنهان را می توان با اعمال رگرسیون خطی میان مقادیر اندازه گیری شده و مقادیر برآورد شده و محاسبه ضریب تبیین بیان کرد. ضریب تبیین (R^2) بالا نشان دهنده دست یافتن به آرایشی مناسب جهت برآورد پارامتر خروجی است. هر نرون در شبکه MLP دو نوع محاسبه انجام می دهد. در نوع اول محاسبات، سیگنال تابعی انجام می شود و در نوع دوم محاسبات، تخمین لحظه ای از گرادینان منحنی خطا انجام می پذیرد. این گرادینانها جهت پخش سیگنال های خطا در شبکه مورد نیاز می باشند (۲۵). در این پژوهش از شبکه پرسپترون چندلایه که لایه ورودی را متغیرهای انتخاب شده از طریق آنالیز مؤلفه های اصلی از بین متغیرهای محیطی بود و لایه

و نزدیک‌تر بودن آن به یک نشان‌دهنده عملکرد بهتر مدل می‌باشد.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{Z}(x_i) - Z(x_i))^2} \quad (5)$$

$$MBE = \frac{1}{n} \sum_{i=1}^n (\hat{Z}(x_i) - Z(x_i)) \quad (6)$$

$$R^2 = \left[\frac{\sum_{i=1}^n (Z(x_i) - \bar{Z}(x_i))(\hat{Z}(x_i) - \bar{Z}(x_i))}{(\sum_{i=1}^n (Z(x_i) - \bar{Z}(x_i))^2)(\sum_{i=1}^n (\hat{Z}(x_i) - \bar{Z}(x_i))^2)} \right]^2 \quad (7)$$

که در آن‌ها، n تعداد نقاط مشاهده‌ای، $\hat{Z}(x_i)$ مقدار برآوردی بر نقطه i ام، $Z(x_i)$ مقدار مشاهده‌ای برای نقطه i ام، $\bar{Z}(x_i)$ میانگین مقادیر مشاهده‌ای می‌باشند.

نتایج

توصیف آماری داده‌ها: نتایج آمار توصیفی خصوصیات خاک در منطقه مورد مطالعه در جدول ۲ ارائه شده است. نتایج این جدول نشان می‌دهد که کربن آلی خاک، نیتروژن و کربنات کلسیم تغییرپذیری نسبتاً بالایی از خود نشان دادند. آزمون نرمالیته کولموگروف-اسمیرونوف نشان داد که به جز متغیرهای شن و نیتروژن، سایر متغیرها از توزیع نرمال پیروی کردند. که نتایج ضریب چولگی نیز بیانگر این مطلب می‌باشد. برای نرمال‌سازی داده‌های شن و نیتروژن از روش لگاریتمی و مجذور مربعات داده‌ها استفاده گردید.

ترمینال‌ها، مدل‌سازی می‌شوند. مجموعه ترمینال‌ها، از مقادیر ثابت و متغیرهای مستقل مسأله تشکیل شده‌اند (۱۸). به‌منظور استفاده از برنامه‌ریزی بیان ژن از نرم‌افزار GeneXprotools 4.0 استفاده گردید. تعداد ۱۱۷ نمونه آموزشی به همراه ۲۰ نمونه به‌عنوان داده تست به نرم‌افزار معرفی گردید. در این مطالعه حداکثر تعداد نسل بر روی حداکثر برازش^۱ یا بهینه‌ترین حالت ممکن تنظیم گردید. از مجموعه توابع شامل توابع اصلی حسابی $\{+, -, \times, /\}$ ، توابع مثلثاتی و ریاضی $\{\sqrt{\quad}, x^2, \exp, \log, \sin, \cos, \arctan, \dots\}$ برای تفسیر مدل‌ها استفاده گردید.

ارزیابی مدل‌ها: جهت اعتبارسنجی مدل‌ها پس از تصادفی نمودن مجموعه داده تعداد ۲۰ عدد از نقاط به‌صورت تصادفی به‌منظور ارزیابی مدل‌ها کنار گذاشته شد. این داده‌ها در مدل‌ها به‌کار گرفته نشدند و از بقیه داده‌ها به‌منظور آموزش و از این ۲۰ نمونه در فاز آزمون مدل استفاده گردید. مقایسه و ارزیابی ۳ روش به‌کار گرفته شده با استفاده از داده‌های آزمون، معیارهای مجذور میانگین مربعات خطا^۲، میانگین اربیبی^۳ خطا و ضریب تبیین مدل‌ها به‌ترتیب توسط رابطه‌های ۵، ۶ و ۷ محاسبه گردید. آماره RMSE دقت مدل را براساس تفاضل بین مقادیر واقعی و برآوردی و آماره MBE انحراف مقادیر برآوردی از مقادیر واقعی را ارزیابی می‌کند معیارهای MBE و RMSE هر چقدر به صفر نزدیک‌تر باشد نشان‌دهنده دقت بیشتر و خطای کم‌تر روش می‌باشد. همچنین از شاخص ضریب تبیین (رابطه ۹) که تعیین‌کننده تناسب برازش مدل رگرسیون خطی به جفت داده‌های تخمینی و مشاهداتی است، نیز استفاده گردید. مقدار این ضریب همواره بین صفر و یک بوده

- 1- Maximum fitness
- 2- Root mean square error (RMSE)
- 3- Mean bias error (MBE)

جدول ۲- آمار توصیفی خصوصیات خاک مورد بررسی (تعداد نمونه: ۱۳۷).

Table 2. Descriptive statistics parameters of soil properties for 137 sample points.

متغیر (Variable)	حداقل (Min)	حداکثر (Max)	میانگین (Mean)	انحراف معیار (SD)	چولگی (Skewness)	کشیدگی (Kurtosis)	ضریب تغییرات (%) (CV%)
کربن آلی (%) (OC)	0.49	3.67	1.60	0.69	1.10	0.90	43.10
شن (%) (Sand)	14.10	51.40	27.10	7.30	1.00	1.20	26.80
سیلت (%) (Silt)	27.00	57.00	43.60	6.40	-0.20	-0.60	14.60
رس (%) (Clay)	16.60	43.60	29.30	4.90	0.40	0.40	16.60
نیتروژن (%) (N)	0.05	0.38	0.16	0.06	1.20	1.80	40.00
کربنات کلسیم معادل (%) (CaCO ₃)	1.00	57.50	23.60	13.30	0.40	-0.60	56.50
ظرفیت رطوبت اشباع (%) (SMC)	26.40	67.43	43.20	7.60	0.40	0.50	17.50
چگالی ظاهری (g cm ⁻³) (D _b)	1.12	1.89	1.48	0.15	0.40	-0.30	10.20

نتایج مدل‌سازی

مدل رگرسیون چندگانه: در جدول ۳ مدل رگرسیونی و ضریب تبیین به دست آمده برای هر یک از خصوصیات خاک آورده شده است. مدل‌های خطی نشان می‌دهند که شاخص‌های پوشش گیاهی نقش کلیدی در تخمین بسیاری از خصوصیات خاک بازی می‌کند به طوری که در تخمین مقدار کربن آلی، سیلت، نیتروژن، ظرفیت نگه‌داری آب، چگالی و مقدار هدایت الکتریکی خاک نقشی اساسی ایفا می‌کرد. نتایج اعتبارسنجی مدل‌ها (جدول ۶) نشان داد که معادلات رگرسیون حاصله، در بهترین حالت ۶۱

درصد از کل تغییرات کربن آلی خاک را توجیه می‌کنند. در همین ارتباط سومفلس و داتمن (۲۰۰۸) از داده‌های عوارض زمین برای پیش‌بینی توزیع خصوصیات خاک از جمله کربن آلی خاک استفاده کردند که نتایج آن‌ها نشان داد که عامل شیب و شبکه آبراهه‌ای در مدل رگرسیون خطی وارد شده است و این مدل می‌تواند فقط ۳۵ درصد تغییرات کربن آلی را توجیه کند (۴۲). تامسون و کولکا (۲۰۰۵) نیز با مدل رگرسیون- زمین‌نما توانستند ۷۰ درصد از تغییرات کربن آلی را پیش‌بینی کنند (۴۷).

جدول ۳- مدل رگرسیونی چندمتغیره بین خصوصیات خاک و داده‌های کمکی.

Table 3. MLR models between soil parameters and covariate.

R	مدل (Model)	متغیر (Variable)
0.73	$0.026(\%Vegetation\ cover)-0.087(Wetness)+0.045(ETM\ B1)+10.56(NDVI)-2.58$	کربن آلی (%) (OC)
0.54	$1.64(B2)-0.5(B7)-0.18(Convergence)-0.88(ETM\ B1)+13.2$	شن (%) (Sand)
0.60	$0.15(\%Vegetation\ cover)-0.5(B2)+0.27(ETM\ B7)+50$	سیلت (%) (Silt)
0.40	$-35.65\ (Clay\ index)+0.024(Elevation)+27.3$	رس (%) (Clay)
0.71	$0.003(\%Vegetation\ cover)+0.13(Slope)+0.002(ETM\ B1)-0.217$	نیتروژن (%) (Total N)
0.73	$0.765(ETM\ B1)-0.265(\%Vegetation\ cover)-0.049(Elevation)+73.43$	کربنات کلسیم (%) (CaCO ₃)
0.61	$-0.004(\%Vegetation\ cover)-2.248\ (NDVI)+1.59$	چگالی ظاهری (g cm ⁻³) (Db)
0.72	$0.24(\%Vegetation\ cover)+0.22(Convergence)+136(NDVI)+0.28(ETM\ B1)+13.6$	ظرفیت رطوبت اشباع (%) (SMC)

MLP به کمک متغیرهای توپوگرافی شامل شاخص خیزی، شاخص موقعیت نسبی، طول شیب و ارتفاع برآورد کردند (۱۹). ایوبی و علیزاده (۱۳۸۵) با استفاده از روش رگرسیونی و داده‌های توپوگرافی به مدل‌سازی مقادیر رس، شن، درصد سنگریزه، درصد ماده آلی، رطوبت وزنی در حالت ظرفیت مزرعه و چگالی ظاهری خاک پرداختند (۶). مدل‌های رگرسیونی به‌دست آمده نتوانستند بخشی از تغییرپذیری خصوصیات خاک را به‌وسیله عوامل کمکی به‌کار گرفته شده توجیه کنند. که این موضوع به‌دلیل وجود روابط غیرخطی بین پدیده‌ها می‌باشد که در مدل رگرسیونی در نظر گرفته نمی‌شود. همچنین به‌دلیل محدودیت‌های موجود نمی‌توان تمامی متغیرهای

معادلات رگرسیونی به‌دست آمده توانستند بین ۴۰ تا ۶۰ درصد تغییرات شن، سیلت، ازت، ظرفیت نگهداری رطوبت، چگالی و هدایت الکتریکی خاک را توجیه کنند ضریب تبیین معادله رگرسیونی پیش‌بینی رس و کربنات کلسیم، کربن و CEC خاک با استفاده از باندهای TM در مطالعه بن- دور و همکاران (۱۹۹۵) به‌ترتیب $R^2=0.76$ ، $R^2=0.7$ ، $R^2=0.69$ و $R^2=0.84$ بود (۹). معادله رگرسیونی در مطالعه دمانته و همکاران (۲۰۰۹) مقدار $R^2=0.61$ را برای پیش‌بینی مقدار رس از طریق تصاویر ماهواره‌ای TM نشان داد (۱۵). گیو و همکاران (۲۰۱۳) در منطقه جیانگجین چین ماده آلی خاک را با استفاده از رگرسیون خطی چندگانه و شبکه عصبی مصنوعی نوع

زمین‌نما و در نهایت وضعیت پوشش گیاهی استفاده گردید. نتایج آنالیز حساسیت نشان داد که اهمیت پارامتر درصد پوشش گیاهی در تخمین تمامی فاکتورهای خاکی بسیار بارز و ملموس بود؛ به طوری که در تمامی مدل‌های تخمین متغیرهای خاک به عنوان یکی از متغیرهای کمکی اصلی بود. در میان اطلاعات سنجش از دوری و شاخص‌های طیفی، شاخص‌های طیفی گیاهی از جمله NDVI و RDVI نقش پررنگ‌تری در تخمین خصوصیات مورد بررسی داشتند. شاخص‌های توپوگرافی در تخمین مقدار رس خاک بیش‌ترین تأثیرگذاری را داشتند. این در حالی بود که نقش فاکتورهای زمین‌نما در تخمین سایر خصوصیات خاکی از اهمیت کم‌تری برخوردار بودند.

برنامه‌ریزی بیان ژن: در جدول ۵ مشخصات به دست آمده برای مدل‌های بیان ژن آورده شده است. در این مرحله به مانند روش شبکه عصبی مصنوعی از نتایج PCR برای انتخاب ورودی‌ها استفاده گردید. همان‌طور که در جدول ۵ دیده می‌شود الگوریتم مهم‌ترین متغیرهای مؤثر را برای مدل‌سازی را انتخاب نموده است به طوری که در بسیاری از مدل‌ها تعداد متغیر به کار برده شده به عنوان ورودی را بیش از ۵۰ درصد کاسته است.

احتمالی مؤثر در تغییرپذیری خصوصیات را وارد مدل‌ها نمود. هر چند که می‌توان برای بهبود مدل‌ها که تغییرات بیش‌تری از خصوصیات خاک در سطح حوضه را پیش‌بینی کند، از فواصل کم‌تر نمونه‌برداری و ابعاد کوچک‌تر پیکسل‌های DEM و یا تصاویر ماهواره‌ای با تفکیک مکانی^۱ بالا استفاده کرد. به منظور در اختیار داشتن حجم بیش‌تری از اطلاعات نیاز به تراکم‌تر شدن شبکه نمونه‌برداری است (۱۷). آن‌ها نشان دادند که با کاهش تراکم شبکه نمونه به میزان ۵۰ درصد R^2 مدل‌های تخمین رس و چگالی ظاهری به ترتیب به میزان ۲۰ و ۲۵ درصد کاسته شد.

مدل شبکه عصبی مصنوعی: ساختار به دست آمده به وسیله شبکه عصبی مصنوعی برای خصوصیات مختلف خاک در جدول ۴ آورده شده است. ساختار شبکه به دست آمده برای خصوصیات مختلف دارای تعداد گره متفاوت ۶ برای رس، ۱۷ (کربنات کلسیم، چگالی، نیتروژن، شن و ظرفیت رطوبتی)، ۱۸ (سیلت) و ۱۹ (کربن آلی) در لایه ورودی بودند و تعداد گره لایه پنهان برای هر کدام از خصوصیات متفاوت و عددی بین ۳ تا ۱۱ نرون بود. در لایه خروجی برای تمام متغیرها یک گره تعریف گردید. پرویزی و گرجی (۲۰۱۰) نیز به منظور پیش‌بینی کربن آلی خاک به مدلی از شبکه عصبی (ساختار پرسپترون چندلایه) با ۳۱ متغیر ورودی، ۲ نرون در لایه پنهان و یک متغیر خروجی دست یافتند (۳۵). اینگل‌بای و گروو (۲۰۰۱) به منظور پیش‌بینی کربن آلی خاک به مدلی از شبکه عصبی (ساختار پرسپترون چندلایه) با ۵ متغیر ورودی، ۱۰ نرون در لایه پنهان و یک متغیر خروجی دست یافتند (۲۳). در این مطالعه از سه دسته داده کمکی شامل داده‌های سنجش از دوری، مشخصات

جدول ۴- خصوصیات بهترین ساختارهای شبکه عصبی مصنوعی برای ویژگی‌های خاک.

Table 4. Characteristics of the best obtained structures for ANN model for soil properties.

تعداد نرون لایه پنهان (Hidden layer No)	تکرار (Repeat)	متغیر کمکی ورودی* (Input auxiliary data)	تابع انتقال (TF)	ساختار شبکه (Net Structure)	مدل (Model)
11	1000	B1- B7, Elev, Curve, NDVI, BI, DVI, MSAVI, RDVI, SARVI, SAVI, COSRI, CVP	Tansig	17-11-1	کربنات کلسیم (CaCO ₃) ^(/)
3	1000	Elev, TWI, Curv, STI, Clay index, CVP	Tansig	6-3-1	رس (Clay) ^(/)
6	1000	B1- B7, Converge, Curve, NDVI, BI, DVI, MSAVI, RDVI, SARVI, SAVI, COSRI, CVP	Tansig	17-6-1	چگالی ظاهری (g cm ⁻³) (Db) ^(/)
4	1000	B1- B7, TWI, Curve, NDVI, BI, DVI, MSAVI, RDVI, SARVI, SAVI, COSRI, CVP	Tansig	17-4-1	نیترژن (Total N) ^(/)
9	1000	B1- B7, TWI, Converge, Curve, Clay index NDVI, BI, DVI, MSAVI, RDVI, SARVI, SAVI, COSRI, CVP	Tansig	19-9-1	کربن آلی (OC) ^(/)
4	1000	B1- B7, Elev, STI, NDVI, BI, DVI, MSAVI, RDVI, SARVI, SAVI, COSRI, CVP	Tansig	17-4-1	شن (Sand) ^(/)
3	1000	B1- B7, Elev, Converge, Clay index, NDVI, BI, DVI, MSAVI, RDVI, SARVI, SAVI, COSRI, CVP	Tansig	18-3-1	سیلت (Silt) ^(/)
8	1000	B1- B7, Elev, TWI, NDVI, BI, DVI, MSAVI, RDVI, SARVI, SAVI, COSRI, CVP	Tansig	17-8-1	ظرفیت رطوبت اشباع (SMC) ^(/)

متغیرهای ورودی شامل باند یک تا پنج و هفت ماهواره لندست (B1, B2, B3, B4, B5 and B7). متغیرهای SAVI, SARVI, NDVI, MSAVI, BI, DVI, MSAVI, RDVI, SARVI, SAVI, COSRI, CVP مشخصات زمین‌نما به ترتیب همگرایی، شاخص حمل جریان، شکستگی، شاخص خیس و ارتفاع.

جدول ۵- مشخصات برنامه‌ریزی بیان ژن به‌دست آمده برای ویژگی‌های خاک.
Table 5. Characteristics of the best obtained structures for GEP model for soil properties.

R ²	معادله به‌دست آمده ^۹ (Obtained equation)	متغیر استفاده شده (Used variables)	نسل (Generation)	تعداد ورودی (Input No)	مدل (Model)
0.56	$CaCO_3 = (BI - SAVI) + \{SARVI - \sin[\text{atan}(CVP \times SARVI)] \times \sin(\sin(NDVI) + \text{atan}(B2 - B3) + \{\sin(SARVI) \times \cos((B4 + CVP) \times B7) \times \exp(B1) - BI\})\}$	10	5721	17	کربنات کلسیم (CaCO ₃) (%)
0.32	$Clay = \text{Curve} + 18.66 - TWI + (\sin(TWI)) \times CVP - (\text{Clay index} \times (\text{Curve} \times CVP) + \text{atan}(\cos(\text{Curve})) - (STI)^3)$	5	8001	6	رس (Clay) (%)
0.36	$Density = B3 + \cos(\exp(CVP))^{1/3} + \sin((B2)3 \times (B3)3) - B2 - 8.2 - BI + \cos(SARVI) \times \cos\{[(4.3 + NDVI) - 4.3] / SARVI - CVP\}$	6	3006	17	چگالی ظاهری (g cm ⁻³) (D _b)
0.5	$TN = SAVI \times (\cos(SAVI) - B7) \times (NDVI + B3) \times (BI + 0.34) + CVP + \cos(0.34) + 0.34(B1) + [\exp(-9.7 \times SARVI)]$	8	2140	17	نیترژن (Total N) (%)
0.56	$OC = SARVI \times (B1)^2 + \sin(B4) \times (CVP - \sin(DVI)) + RDVI + \{[CVP - \exp(\cos(MSAVI \times TWI))] \times (RDVI)^3\}$	8	4345	19	کربن آلی (OC) (%)
0.36	$Sand = \sin[(\text{Converge}) \times B4 + CVP \times 9.4] + \sin[B2 - \exp(B1) \times \sin(MSAVI)] + B1 \times (SARVI)2 \times SAVI - ((B4)3 - BI) + DVI + B2$	10	1168	17	شن (Sand) (%)
0.49	$Silt = \{\sin(6.25) \times [\sin(SAVI) - (2.5 - B2)] + (CVP)\} + [(MSAVI) \times (B7)^2 - SAVI - \cos[\cos(CVP)]]^2 + \sin(0.15 + RDVI) - 0.15 - [\sin(B2) / 4.3]^2 + \text{atan}[\text{atan}(CVP)]^2 - (SARVI)^3$	7	6785	18	سیلت (Silt) (%)
0.47	$SMC = (CVP \times SAVI) + 0.16 - (RDVI) - (NDVI) \times 0.41 + ((NDVI) \times 0.16) + [(B4)3 \times (SAVI)] + [(CVP) \times (SARVI)] + (SARVI) + \sin(DVI)$	7	9427	17	ظرفیت رطوبت اشباع (SMC) (%)

متغیرهای به‌کار رفته شده در معادلات شامل باتد یک تا پنج و هفت ماهراره اندست (B1, B2, B3, B4, B5 and B7). مشخصات زمین‌نما به‌ترتیب همگرایی، شاخص حمل جریان، شکستگی، شاخص خمسی و ارتفاع و در نهایت درصد پوشش گیاهی CVP. Elev, TWI, Curv, STI, Converge

شبکه عصبی مصنوعی برای خصوصیات کربنات کلسیم معادل، چگالی، نیتروژن، شن و سیلت بین مقادیر حدود ۰/۶ تا ۰/۷ بود (جدول ۶). بهترین ضریب تبیین برای مقدار کربن آلی با تعداد لایه ورودی ۱۹ و تعداد نرون پنهان ۹ و ظرفیت رطوبت اشباع با ۱۷ نرون ورودی و ۸ لایه پنهان به ترتیب با مقادیر ۰/۷۷ و ۰/۸۵ بود.

مقایسه روش‌ها: در جدول ۶ معیارهای خطا و مقدار ضریب تبیین (R^2) برای داده‌های ارزیابی بر اساس مقادیر اصلی و بدون استانداردسازی و نرمال‌سازی ارائه گردید. بر اساس این جدول، شبکه عصبی مصنوعی با بیش‌ترین ضریب تبیین، کم‌ترین خطای RMSE و MBE برای برآورد خصوصیات مختلف خاک، مناسب‌ترین مدل ارزیابی شد. ضریب تبیین

جدول ۶- پارامترهای آماری مقایسه مدل‌های رگرسیونی و شبکه عصبی مصنوعی و الگوریتم بیان ژن.

Table 6. Comparing of MLR, ANN and GEP models statistical parameters.

برنامه‌ریزی بیان ژن (GEP)			شبکه عصبی مصنوعی (ANN)			مدل رگرسیون خطی (MLR)			
MBE	RMSE	R ²	MBE	RMSE	R ²	MBE	RMSE	R ²	
0.75	8.20	0.60	0.45	7.46	0.72	-0.06	9.44	0.51	کربنات کلسیم معادل (%) (CaCO ₃)
-0.46	3.60	0.37	0.15	4.46	0.46	10.90	11.7	0.30	رس (%) (Clay)
0.22	0.30	0.46	0.01	0.08	0.69	0.05	0.15	0.57	چگالی ظاهری (g cm ⁻³) (D _b)
-0.01	0.05	0.49	0.03	0.03	0.67	-0.03	0.06	0.41	نیتروژن (%) (Total N)
-0.22	0.49	0.59	-0.04	0.27	0.77	0.04	0.49	0.61	کربن آلی (%) (OC)
0.41	4.90	0.44	-0.60	5.60	0.62	-0.45	4.66	0.40	شن (%) (Sand)
0.01	3.80	0.53	-0.30	3.55	0.70	0.28	4.40	0.45	سیلت (%) (Silt)
-1.20	5.10	0.52	-0.47	3.40	0.85	1.19	5.90	0.50	ظرفیت رطوبت اشباع (%) (SMC)

خطی، شبکه عصبی مصنوعی پرسپترون و الگوریتم بیان ژن صورت گرفت که بهترین نتایج به دست آمده را مدل شبکه عصبی مصنوعی پرسپترون نشان داد. نتایج اعتبارسنجی مدل شبکه عصبی مصنوعی نشان داد که مقدار MBE مدل برای متغیرها نزدیک به صفر

بحث و نتیجه‌گیری

نتایج کلی این پژوهش نشان داد، پس از انتخاب ورودی‌ها از طریق آنالیز PCA و سپس ساخت مدل‌های تشریح‌شده در بخش مواد و روش، تخمین خصوصیات خاک از طریق روش‌های رگرسیون

بوده و این امر بیانگر این مطلب است که برازش، توسط مدل ایجاد شده ناریب بوده است. مقدار RMSE پایین مدل نیز نشان‌دهنده دقت مناسب و قابل قبول برآورد برای متغیرهای خاک می‌باشد.

دو مدل شبکه عصبی مصنوعی پرسپترون و بیان ژن نسبت به معادلات رگرسیونی کارایی مناسب‌تری را نشان دادند. در واقع الگوریتم‌های شبکه عصبی مصنوعی و بیان ژن از نوع ویژه‌ای از معادلات استفاده نکرده و شبکه با ایجاد رابطه مناسب بین داده‌های ورودی و خروجی می‌تواند به نتایج رضایت‌بخشی دست یابد (۴۰). در صورت وجود رابطه غیرخطی بین متغیرهای وابسته و مستقل شبکه عصبی مصنوعی با توجه به بررسی غیرخطی بودن داده‌ها می‌تواند به نتایج بهتری نسبت به روش‌های رگرسیونی که تنها به بررسی وجود رابطه خطی می‌پردازند، دست یابد. شایان ذکر است که در مدل‌های رگرسیونی، همبستگی ضعیف بین دو متغیر وابسته به هم، همیشه نشان‌دهنده عدم ارتباط این دو متغیر با همدیگر نیست و ممکن است همبستگی غیرخطی بین دو متغیر وجود داشته باشد که به وسیله ضریب همبستگی خطی قابل اندازه‌گیری نیست (۲۴). از سوی دیگر، مقیمی و همکاران (۲۰۱۵) اشاره داشتند که روش‌های رگرسیون خطی برای تخمین مناسب به حجم بالایی از داده‌ها نیاز دارند، در صورتی که شبکه عصبی مصنوعی با داده‌های کم‌تر، قادر به برآورد دقیق‌تر است (۳۱). که این امر در راستای نتایج پیلهور و همکاران (۲۰۱۱) و همچنین امینی و همکاران (۲۰۰۵) مبنی بر برتری روش شبکه عصبی مصنوعی بر روش رگرسیون چندگانه می‌باشد (۴ و ۳۶). در مقایسه روش بیان ژن و شبکه عصبی مصنوعی پرسپترون، با توجه به مقادیر RMSE و ضریب تبیین بالاتر روش شبکه عصبی مصنوعی، این روش از دقت بالاتری برخوردار بود.

در مقابل، تقی‌زاده و همکاران (۲۰۱۶) برنامه‌ریزی ژنتیک را برای استفاده در مناطق خشک ایران به منظور تخمین شوری خاک پیشنهاد دادند (۴۵) همچنین در مطالعه‌ای دیگر، در تهیه نقشه ظرفیت تبادل خاک، تقی‌زاده و همکاران (۲۰۱۵) برنامه‌ریزی ژنتیک را برای مدل‌سازی تخمین مقدار ظرفیت تبادل خاک مورد ارزیابی نمودند که نتایج حاصل از دقت مناسب روش حکایت داشت (۴۴). اما در این مطالعه روش شبکه عصبی مصنوعی برای تمامی خصوصیات مورد بررسی نتایج قابل قبول‌تری ارائه نمود. نتایج تابدوردا و همکاران (۲۰۱۳) نیز از کاربرد رضایت‌بخش روش شبکه عصبی مصنوعی پرسپترون به منظور تخمین و تهیه نقشه خاک حکایت دارد. آن‌ها روش شبکه عصبی مصنوعی را روشی بالقوه در کلاس‌بندی واحدهای خاک عنوان داشتند (۴۳). پریوری و همکاران (۲۰۱۴) در مقایسه روش SVM^۱ و شبکه عصبی مصنوعی در تخمین مقادیر رس و شن خاک، هر دو روش را مناسب ارزیابی نمودند و اختلاف چندانی در صحت تخمین این دو روش مشاهده نمودند (۳۷). در واقع تفاوت روش شبکه عصبی مصنوعی و روش بیان ژن در این است که در برنامه‌ریزی بیان ژن ابتدا ساختار بلوک‌ها شامل متغیرهای ورودی، هدف و مجموعه توابع مشخص می‌شود و در مرحله‌های بعدی ساختار بهینه و ضرایب مدل در طول فرآیند آموزش تعیین می‌گردد. در صورتی که در مدل شبکه‌های عصبی، ابتدا ساختار تعیین می‌شود و ضرایب مدل، طی فرآیند آموزش مشخص می‌شود (۳). در استفاده از روش ژنتیک همواره باید به دو مسأله مهم انتخاب صحیح ورودی و جلوگیری از پف کردن توجه داشت. از آنجایی که در روش‌های ژنتیکی متغیرهای ورودی کاهش می‌یابند، باید در انتخاب ورودی‌های مناسب دقت

استفاده در لایه‌های پنهان و خروجی، تعداد لایه‌های پنهان و تعداد نرون در لایه پنهان به روش سعی و خطا همراه می‌باشد (۳۶).

نتیجه‌گیری کلی

در این پژوهش سه مدل عددی رگرسیون خطی چندگانه، شبکه عصبی مصنوعی و برنامه‌ریزی بیان ژن به منظور استفاده در معادله اسکورپن برای تخمین ویژگی‌های فیزیکی و شیمیایی مورد مقایسه قرار گرفتند. نتایج مطالعه حاضر نشان داد که براساس مقدار آماره‌های ارزیابی در فاز آزمون، روش شبکه عصبی مصنوعی دارای دقت بیش‌تری در برآورد مقادیر رس، شن، سیلت، کربن آلی، نیتروژن، کربنات کلسیم، ظرفیت آب اشباع و چگالی خاک در مقایسه با دیگر مدل‌های بررسی شده برخوردار بود. در حالی که مدل رگرسیونی به‌عنوان ضعیف‌ترین مدل در برآورد ویژگی‌های مورد بررسی بود که این امر نشان از رفتار غیرخطی متغیرهای مورد مطالعه بود که موجب برتری شبکه عصبی مصنوعی بر مدل رگرسیون چندمتغیره خطی شد. نتایج روش شبکه عصبی مصنوعی برای تخمین ویژگی‌های خاک رضایت‌بخش بود به‌گونه‌ای که این مدل توانست ۶۰ تا ۸۵ درصد تغییرپذیری خصوصیات مورد بررسی را نشان دهد. با توجه به نتایج به‌دست آمده می‌توان مدل شبکه عصبی مصنوعی را به‌عنوان مدلی مناسب و دقیق برای استفاده در مدل‌های تخمین خصوصیات خاک مانند معادله اسکورپن به‌کار برد.

شود، در غیر این صورت، با ارائه داده‌های ورودی غیرمرتبط دقت برنامه‌نویسی ژنتیک کاهش می‌یابد، همچنین مدل‌های پیچیده‌تری ایجاد می‌گردد. یکی از مشکلات احتمالی که کاربر در اجرای برنامه ژنتیک با آن روبرو شود، رشد بیش از حد برنامه است که به این پدیده در اصطلاح پف کردن^۱ برنامه گفته می‌شود. روش‌های مختلفی برای کنترل رشد بیش از حد برنامه از جمله فاکتور دقت و پیچیدگی مدل به‌دست آمده به‌زای تعداد نسل ثابت (برای مثال هر ۱۰۰ نسل)، در این برنامه‌ها وجود دارد (۳۹). بر اساس نتایج شعبانی و همکاران (۲۰۱۱) حساسیت کم‌تر شبکه عصبی مصنوعی نسبت به وجود خطا در داده‌های ورودی را می‌توان یکی دیگر از عوامل نتایج رضایت‌بخش‌تر این روش نسبت به برنامه‌ریزی بیان ژن در این مطالعه دانست (۴۱).

اگرچه روش شبکه عصبی مصنوعی توانست ۶۰ تا ۸۵ درصد تغییر پذیری اکثریت خصوصیات مورد بررسی را نشان دهد ولی در نهایت این روش از تبیین ۱۵ تا ۴۰ درصد از تغییرپذیری متغیرها در منطقه عاجز بود که این مسأله به اندازه زیادی به‌دلیل در نظر نگرفتن پارامترهای دیگر محیطی مانند اقلیم و یا مواد مادری که در معادله اسکورپن آمده‌اند و بر مقادیر خصوصیات خاک تأثیرگذار بوده نسبت داد. روش‌های موجود کم‌ترین پوشش تغییرپذیری را برای مقدار رس خاک نشان دادند. که این مسأله را می‌تواند به رابطه کم‌تر داده محیطی به‌کار برده شده به‌عنوان داده کمکی با مقدار رس نسبت به سایر ویژگی‌های خاک در منطقه نسبت داد که نتایج همبستگی پیرسون نیز این امر را نشان دادند. از سوی دیگر مدل شبکه عصبی مصنوعی نیز با مشکلاتی مانند آزمون اجزای مختلف شبکه مانند قانون یادگیری، توابع انتقال مورد

1- Bloating

منابع

1. Aitkenhead, M.J., and Coull, M.C. 2016. Mapping soil carbon stocks across Scotland using a neural network model. *Geoderma*. 262: 187-198.
2. Aitkenhead, M.J., Coull, M., Towers, W., Hudson, G., and Black, H.J. 2013. Prediction of soil characteristics and colour using data from the National Soil Inventory of Scotland. *Geoderma*. 200/201: 99-107.
3. Alami, M., Sadegh Fam, S., and Fazelifard, M. 2012. *Data series Modeling*, 3rd edition, Tabriz University Press, Tabriz, Iran, 622p. (In Persian)
4. Amini, M., Abbaspour, K.C., Khademi, H., Fathianpour, N., Afyuni, M., and Schulin, R. 2005. Neural network models to predict cation exchange capacity in arid regions of Iran. *Eur. J. Soil Sci.* 56: 551-559.
5. Andrews, S.S., Mitchell, J.P., Mancinelli, R., Karlen, D.L., Hartz, T.K., Horwath, W.R., Pettygrove, G.S., Scow, K.M., and Munk, D.S. 2002. On-farm assessment of soil quality in California's central valley. *Agron. J.* 94: 1. 12-23.
6. Ayoubi, S., and Alizadeh, M.H. 2006. Soil Surface properties prediction using digital elevation model. *J. Agric. Sci. Natur. Resour.* 10: 2. 85-96. (In Persian)
7. Bannari, A., Morin, D., Bonn, F., and Huete, A.R. 1995. A review of vegetation indices. *Rem. Sens. Rev.* 13: 2. 95-120.
8. Bagheri, M.B., and Mart, A. 2015. Digital soil mapping using artificial neural networks and terrain-related attributes. *Pedosphere*. 25: 4. 580-591.
9. Ben-Dor, E., and Banin, A. 1995. Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties. *Soil Sci. Soc. Am. J.* 59: 364-372.
10. Boettinger, J.L., Ramsey, R.D., Bodily, J.M., Cole, N.J., Kienast-Brown, S., Nield, S.J., Saunders, A.M., and Stum, A.K. 2008. Landsat Spectral Data for Digital Soil Mapping, P 193-202. In: Hartemink, A.E., McBratney, A., and de Lourdes Mendonça-Santos, M. (eds.). *Digital Soil Mapping with Limited Data*. Springer.
11. Bourennane, H., Couturier, A., Pasquier, C., Chartin, C., Hirschberger, F., Macaire, J., and Salvador-blanc, S. 2014. Comparative performance of classification algorithms for the development of models of spatial distribution of landscape structures. *Geoderma*. 219: 136-44.
12. Cavazzi, S., Corstanje, R., Mayr, T., Hannam, J., and Fealy, R. 2013. Are fine resolution digital elevation models always the best choice in digital soil mapping? *Geoderma*. 195/196: 111-121.
13. Coleman, T.L., Agbu, P.A., and Montgomery, O.L. 1993. Spectral differentiation of surface soils and soil properties: is it possible from space platforms? *Soil Sci.* 155: 283-293.
14. Collard, F., Kempen, B., Heuvelink, G.B.M., Saby, N.P.A., Richer, A.C., Forges, D., Lehmann, S., Nehlig, P., and Arrouays, D. 2014. Regional refining a reconnaissance soil map by calibrating regression models with data from the same map (Normandy, France). *Geoderma Reg.* 1: 21-30.
15. Demattê, J.A.M., Fiorio, P.R., Ben-dor, E., Fioriob, P.R., and Ben-Dorc, E. 2009. Estimation of soil properties by orbital and laboratory reflectance means and its relation with soil classification. *Open Remote Sens. J.* 2: 12-23.
16. Emamgolizadeh, S., Bateni, S.M., Shamsavani, D., Ashrafi, T., and Ghorbani, H. 2015. Estimation of soil cation exchange capacity using genetic expression programming (GEP) and multivariate adaptive regression splines. *J. Hydrol.* 529: 1590-1600.
17. Fazeli Sangani, M., Sarmadian, F., and Shorafa, M. 2010. Surveying and mapping of soil physical properties using Geostatistic. M.Sc Thesis, Soil Science Department, Faculty of Agriculture, University of Tehran, Tehran. (In Persian)
18. Ferreira, C. 2002. Gene Expression Programming in Problem Solving, Pp: 635-653. In: Roy, R., Koeppen, M., Ovaska, S., Furuhashi, T., and Hoffmann, F. (eds.) *Soft Computing and Industry*. Springer London.
19. Guo, P.T., Wu, W., Sheng, Q.K., Li, M.F., Liu, H.B., and Wang, Z.Y. 2013. Prediction of soil organic matter using artificial neural network and topographic indicators in hilly areas. *Nutr. Cycl. Agroecosys.* 95: 3. 333-344.

20. Han, J., Pei, J., and Kamber, M. 2011. Data mining: concepts and techniques. Elsevier, Waltham, USA, 673p.
21. Hengl, T., and Reuter, H.I. 2009. Geomorphometry: Concepts, Software, Applications. AE Amsterdam, Netherlands. 775p.
22. Heung, B., Chak, H., 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.
23. Ingleby, H.R., and Crowe, T.G. 2001. Neural network models for predicting organic matter content in Saskatchewan soils. *Canad. Bisys. Engin.* 43: 7. 1-7.5.
24. Karamouz, M., and Araghinejad, S. 2014. Advanced Hydrology. 3rd edition. Amirkabir University of Technology Press, 464p. (In Persian)
25. Kia, M. 2009. Neural Networks in Matlab. Kian Rayan Sabz Publication, Tehran, 408p. (In Persian)
26. Ließ, M., Glaser, B., and Huwe, B. 2012. Uncertainty in the spatial prediction of soil texture: Comparison of regression tree and random forest models. *Geoderma*. 170: 70-79.
27. Luo, Z., Yaolin, L., Jian, W., and Jing, W. 2008. Quantitative mapping of soil organic material using field spectrometer and hyperspectral remote sensing. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* 37: 901-906.
28. Mahmoudabadi, E., and Karimi, A. 2015. Mapping of calcium carbonate equivalent and clay content of surface soil using geostatistical methods (Case study: Chitgar park, Tehran). *RS GIS Tech. Nat. Resour.* 6: 3. 73-85. (In Persian)
29. McBratney, A.B., Santos, M.L.M., and Minasny, B. 2003. On digital soil mapping. *Geoderma*. 117: 3-52.
30. Metternicht, G.I., and Zinck, J.A. 2003. Remote sensing of soil salinity: potentials and constraints. *Rem. Sens. Environ.* 85: 1-20.
31. Moghimi, S., and Parvizi, Y. 2015. Comparison of applying multi linear regression analysis and artificial neural network methods for simulating topographic factors effect on soil organic carbon. *Water. Eng. Manage.* 6: 312-322.
32. Montgomery, D.C., Peck, E.A., and Vining, G.G. 2015. Introduction to Linear Regression Analysis. John Wiley & Sons.
33. Naseri, K. 2008. Calibration and application of rangeland health assessment method in the range ecosystem of Khorasan province (Case study: Tandoureh area). Ph.D. Thesis, Gorgan University of Agriculture Sciences and Natural Resources. (In Persian)
34. Pansu, M., and Gautheyrou, J. 2007. Handbook of Soil Analysis: Mineralogical, Organic and Inorganic Methods. Springer Science & Business Media. 987p.
35. Parvizi, Y., Gorji, M., Omid, M., Mahdian, M.H., and Amini, M. 2010. Determination of soil organic carbon variability of rainfed crop land in semi-arid region (Neural Network Approach). *Mod. Appl. Sci.* 4: 7. 25-33.
36. Pilevar, A.R., Ayoubi, S., and Khademi, H. 2011. Comparison of artificial neural network (ANN) and multivariate linear regression (MLR) models to predict soil organic carbon using digital terrain analysis (Case study: Zargham Abad Semirom, Isfahan proviance). *J. Water Soil.* 24: 1151-1163. (In Persian)
37. Priori, S., Bianconi, N., and Costantini, E.A.C. 2014. Can γ -radiometrics predict soil textural data and stoniness in different parent materials? A comparison of two machine-learning methods. *Geoderma*. 226: 354-364.
38. Ratnayake, R.R., Karunaratne, S.B., Lessels, J.S., Yogenthiran, N., Rajapaksha, R.K., and Gnanavelrajah, N. 2016. Regional digital soil mapping of organic carbon concentration in paddy growing soils of Northern Sri Lanka. *Geodрма Reg.* 7: 2. 167-176.
39. Rouhnavaz, M., and Htamloo, A. 2014. Modeling of fluent- participation using genetic algorithm programming, in: 1st National Industrial Mathematics Conference (NIMC 2014) 28 May 2014. Tabriz. (In Persian)

40. Schaap, M.G., Leij, F.J., and van Genuchten, M.T. 1998. Neural network analysis for hierarchical prediction of soil hydraulic properties. *Soil Sci. Soc. Am. J.* 62: 847-855.
41. Shabani, A. 2011. Topographic and soil attributes effects on rainfed wheat yield in Sisab region, Northeastern Iran. M.Sc. Thesis, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran. (In Persian)
42. Sumfleth, K., and Duttman, R. 2008. Prediction of soil property distribution in paddy soil landscapes using terrain data and satellite information as indicators. *Ecol. Ind.* 8: 485-501.
43. Taborda, C., Oka-fiori, C., José, L., Santos, C., Evaristo, A., Ribeiro, C., and Faria, M., 2013. Soil prediction using artificial neural networks and topographic attributes. *Geoderma*. 195: 165-172.
44. Taghizadeh-mehrjardi, R. 2015. Digital mapping of cation exchange capacity using genetic programming and soil depth functions in Baneh region, Iran. *Arch. Agrono. Soil Sci.* 62: 1. 37-41.
45. Taghizadeh-mehrjardi, R., Ayoubi, S., Namazi, Z., and Malone, B.P. 2016. Prediction of soil surface salinity in arid region of central Iran using auxiliary variables and genetic programming. *Arid Land Res. Manage.* 30: 1. 49-64.
46. Thomas, M., Clifford, D., Bartley, R., Philip, S., Brough, D., Gregory, L., Willis, R., and Glover, M. 2015. Putting regional digital soil mapping into practice in tropical Northern Australia. *Geoderma*. 241: 145-157.
47. Thompson, J.A., and Kolka, R.K. 2005. Soil carbon storage estimation in a forested watershed using quantitative soil-landscape modeling. *Soil Sci. Soc. Am. J.* 69: 1086-1093.
48. Wuttichaikitcharoen, P., and Babel, M.S. 2014. Principal component and multiple regression analyses for the estimation of suspended sediment yield in Ungauged Basins of Northern Thailand. *Water*. 6: 8. 2412-2435.
49. Zhou, P., Zhao, Y., Zhao, Z., and Chai, T. 2015. Source mapping and determining of soil contamination by heavy metals using statistical analysis, artificial neural network and adaptive genetic algorithm. *J. Environ. Chem. Engin.* 3: 4. 2569-79.



Gorgan University of Agricultural
Sciences and Natural Resources

J. of Water and Soil Conservation, Vol. 24(2), 2017
<http://jwsc.gau.ac.ir>

Assessing performance of multivariate linear regression (MLR), artificial neural network (ANN) and gene expression programming (GEP) in estimating soil properties

E. Mahmoudabadi¹, *A.R. Karimi², Gh.H. Haghnia³ and A. Sepehr⁴

¹Ph.D. Student, Dept. of Soil Science, Faculty of Agriculture, Ferdowsi University of Mashhad,

²Associate Prof., Dept. of Soil Science, Faculty of Agriculture, Ferdowsi University of Mashhad,

³Professor, Dept. of Soil Science, Faculty of Agriculture, Ferdowsi University of Mashhad,

⁴Assistant Prof., Dept. of **Drylands and Desert Management**, ~~Faculty of Natural Resources~~,
Ferdowsi University of Mashhad

Received: 09/17/2016; Accepted: 06/19/2017

Abstract

Background and Objectives: With the emergence of computers and geographic information system (GIS), as well as access to spatial digital data, different methods of data mining, modeling and estimation of soil properties found their place in soil sciences. Data mining of soil properties using computer-based statistical methods uncovers hidden patterns in the database which ultimately leads to the fitness of a model for estimation of soil properties. These methods can be used in the scorpan equation. Two main components of scorpan model include environmental variables and learning program. In the present study, three different methods including multiple linear regression (MLR), artificial neural network (ANN) and Gene Expression Programming (GEP) as “f” function in scorpan model were evaluated and compared in estimating of soil properties using auxiliary data such as vegetation data, topography and remote sensing data.

Material and Methods: The study area with an area of 1225 ha was located in Bajgiran rangelands, Khorasan Razavi province, Iran. In order to investigate vegetation cover and soil 137 units were investigated in which 3-5 plots were selected with a distance of 10 meters apart along an accidental transect and plant species names and numbers besides vegetation percentage were recorded. Next, one soil sample was taken from each transect (Totally 137 soil sample). Train attributes derived from digital elevation model; different bands derived from the ETM and used for computing spectral indices; and plant diversity indices were calculated using Simpson and Shannon-Wiener. These obtained parameters were used as covariate in estimating calcium carbonate equivalent, clay, density, nitrogen, carbon, sand, silt and saturated moisture capacity. Data deduction was done by PCA analysis to deduct the number of input data for ANN and GEP models and finally, Normalization and standardization were carried out on the data.

Results: The results obtained from the evaluation of three numerical methods based on root mean square error (RMSE), mean bias error (MBE) and coefficient of determination (R^2) showed that ANN model had the highest accuracy in estimating soil properties, given the higher coefficients of determination for calcium carbonate equivalent, clay, density, nitrogen, carbon, sand, silt and saturated moisture capacity with the values of 0.72, 0.46, 0.69, 0.67, 0.77, 0.62, 0.7 and 0.85, respectively, moreover, lower RMSE with the values of 7.46, 4.46, 0.08, 0.03, 0.27, 5.6, 3.5 and 3.4, respectively. ANN could explain 60-85 percent of variability of soil properties, among which the best estimates were for saturated moisture capacity and soil organic carbon with $R^2 = 0.85$ and $R^2 = 0.77$, respectively.

Conclusion: Evaluating the estimation of soil properties through three numerical models introduced ANN as the most accurate model. ANN validation results showed that mean bias error (MBE) for estimated soil properties were close to zero and this confirms that the fitting has been created unbiased by model. Furthermore, the low RMSE of model verified accurate estimation of soil variables. The results also indicate that GEP had higher accuracy than the linear regression method for most soil properties.

Keywords: Artificial neural network, Remote sensing, Scorpan model, Gene expression programming, Soil modeling

* Corresponding Author; Email: karimi-a@um.ac.ir