نشریه آب و خاک (علوم و صنایع کشاورزی)
جلد ۳۲، شماره ۲، خرداد – تیر ۱۳۹۷، ص. ۳۴۲-۳۲۷

Journal of Water and Soil
Vol. 32, No. 2, May.-Jan. 2018, p. 327-342

انتخاب ویژگی‌های مؤثر بر برخی شاخص‌های کیفیت فیزیکی خاک و پیش‌بینی آن‌ها با درخت تصمیم و رگرسیون چند متغیره خطی

حسین شکفته[۱*]- مریم دوستکی[۲]- اعظم مسعودی[۳]

تاریخ دریافت: ۱۳۹۶/۰۸/۱۷
تاریخ پذیرش: ۱۳۹۷/۰۱/۲۰

چکیده

اثر راهکارهای مدیریتی و استفاده از اراضی روی کیفیت خاک با شاخص‌های کیفیت فیزیکی خاک تعیین می‌شود که با ویژگی‌های فیزیکی- شیمیایی تخمین زده می‌شوند. جهت پیش‌بینی این شاخص‌ها، انتخاب ویژگی‌های مؤثر بر آن‌ها امری ضروری می‌باشد. مسأله انتخاب زیرمجموعه ویژگی‌ها، به مفهوم شناسایی و انتخاب یک زیرمجموعه مفید از ویژگی‌ها از میان مجموعه داده اولیه می‌باشد. در این مطالعه از ضریب همبستگی پیرسون برای انتخاب ویژگی‌های مؤثر بر شاخص‌های کیفیت فیزیکی خاک (از جمله؛ ظرفیت هوایی (AC)، ظرفیت زراعی نسبی (RFC) و آب قابل دسترس گیاه (PAWC)) استفاده شد و در ادامه پیش‌بینی این شاخص‌های کیفیت با الگوریتم درخت تصمیم رگرسیونی و رگرسیون چند متغیره خطی انجام شد. بدین منظور از ۱۰۴ نقطه از چهار اراضی باغ، جنگل، مرتع و زراعی شهرستان رابر واقع در استان کرمان نمونه‌برداری شد و پارامترهای از جمله؛ بافت، تخلخل، چگالی ظاهری و حقیقی، هدایت هیدرولیکی، اسیدیته کربنات کلسیم، ظرفیت هوایی، ظرفیت زراعی نسبی و آب قابل دسترس اندازه‌گیری شدند. نتایج نشان داد برای ظرفیت هوایی، ویژگی‌های تخلخل، چگالی ظاهری، رس و شن، برای ظرفیت زراعی نسبی، تخلخل، شن، رس و سیلت و برای آب قابل دسترس گیاه، چگالی ظاهری، رس، قابلیت هدایت الکتریکی، تخلخل، شن و سیلت به عنوان پارامترهای ورودی مهم انتخاب شدند. همچنین مقدار R^2 به‌دست آمده برای مدل درخت تصمیم برای ظرفیت هوایی، ظرفیت زراعی نسبی و آب قابل دسترس به ترتیب ۰/۹۵، ۰/۸۴ و ۰/۸۵ بود در حالی‌که در مدل رگرسیون چند متغیره خطی این شاخص‌ها به ترتیب ۰/۶۳، ۰/۶۲ و ۰/۶۱ مشاهده شدند. ویژگی‌های تخلخل و چگالی ظاهری بر ظرفیت هوایی، تخلخل بر ظرفیت زراعی نسبی و چگالی ظاهری بر روی آب قابل دسترس گیاه به عنوان مهم‌ترین پارامترهای مؤثر شناخته شدند. این پژوهش یک اساس برای پیش‌بینی و شناسایی پارامترهای مهم بر روی این سه ویژگی فیزیکی یا هیدرولیکی در خاک‌های کشاورزی، در منطقه نیمه خشک را فراهم کرد که می‌توان به مناطق دیگر نیز تعمیم داده شود.

واژه‌های کلیدی: آب قابل‌دسترس، ظرفیت زراعی نسبی، ظرفیت هوایی، مدل‌سازی، مدیریت خاک

مقدمه

شاخص‌های کیفیت خاک ویژگی‌های قابل اندازه‌گیری خاک هستند که ظرفیت خاک برای تولید محصول یا عملکرد زیست محیطی را تحت تأثیر قرار می‌دهند و به تغییر کاربری اراضی، مدیریت یا عملیات حفاظتی حساس می‌باشند (۲۶). از آن‌جا که کیفیت خاک به‌طور مستقیم قابل اندازه‌گیری نیست، معمولاً از طریق اندازه‌گیری شاخص‌های مختلف فیزیکی، شیمیایی و زیستی خاک، بیان می‌گردد (۱۶). برخی از مهم‌ترین شاخص‌های کیفیت فیزیکی خاک ظرفیت زراعی نسبی (Relative field capacity)، آب قابل دسترس گیاه (Plant available water capacity) و ظرفیت هوایی (Air capacity) می‌باشند (۲۳). به علت اهمیت زیاد کیفیت فیزیکی خاک، ارزیابی کمی آن برای تعیین پایداری سیستم‌های مدیریت اراضی و کمک به سازمان‌های دولتی برای ایجاد کشاورزی پایدار و کاربری صحیح اراضی ضروری است. از این رو تعیین شاخص‌های پایداری و کیفیت خاک ضروری می‌باشد (۱۲). پژوهشگران مختلف بر اهمیت کیفیت فیزیکی خاک در رابطه با رشد گیاه و وضعیت شیمیایی و بیولوژیکی خاک تأکید کرده‌اند (۲، ۱۱ و ۲۷). تعیین شاخص‌های بالا درک ما را از روابط بین کیفیت فیزیکی خاک، عملکرد گیاه، اثرات محیط زیست و حرکت آب و

۱- استادیار گروه علوم خاک، دانشگاه جیرفت
(*- نویسنده مسئول: Email: h.shekofteh@vru.ac.ir)
۲ و ۳- دانشجوی دکتری و دانشجوی کارشناسی ارشد گروه علوم خاک، دانشگاه ولیعصر رفسنجان

DOI: 10.22067/jsw.v32i2.68121

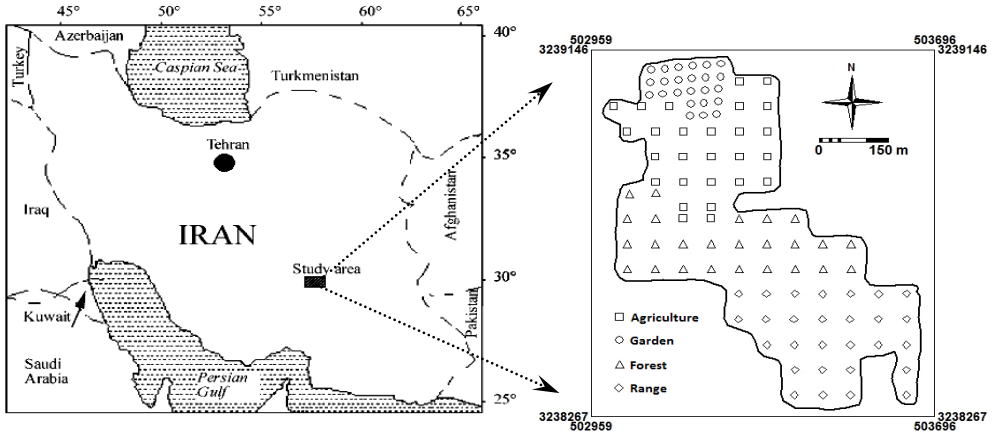
املاح در پروفیل خاک بهبود می‌بخشد. یکی از ویژگی‌های مشترک همه این ویژگی‌ها این است که آنها بطور مستقیم یا غیرمستقیم حجم و اندازه منافذ خاک را بیان می‌کنند. به عنوان مثال RFC، AC و PAWC ارتباط مستقیم با تخلخل و ویژگی‌های انتشار آب در خاک را دارد (۱۰، ۲۲ و ۲۶). امروزه از روش‌های غیرمستقیم مانند توابع انتقالی خاک و توابع تخمین مکانی خاک به‌منظور برآورد. شاخص‌های ذکر شده بالا استفاده می‌گردد. در این توابع با استفاده از خصوصیات زودیافت، خصوصیات دیریافت مانند ویژگی‌های فیزیکی، مکانیکی و هیدرولیکی خاک تخمین زده می‌شوند. راه‌حل اصولی برای تعیین پارامترهای مؤثر بر یک پارامتر روش‌های انتخاب ویژگی می‌باشند. روش‌های انتخاب ویژگی در خاکشناسی تعدادی محدود هستند (۲۵ و ۲۶).

روش‌های انتخاب ویژگی باعث کاهش ابعاد داده‌های ورودی می‌شوند و این کاهش ابعاد داده‌های ورودی منجر به افزایش دقت مدل‌های پیش‌بینی کننده می‌گردد. یکی از مشکلات داده‌های با ابعاد زیاد این است که در بیشتر مواقع تمام ویژگی‌های داده‌ها برای یافتن دانشی که در داده‌ها نهفته است مهم و حیاتی نیستند. بنابراین، انتخاب ویژگی‌های مرتبط و ضروری در مرحله پیش‌پردازش از اهمیتی بنیادین برخوردار است. انتخاب ویژگی به‌عنوان گام اولیه در تشخیص الگو به شمار می‌آید (۳۲). یکی از روش‌های انتخاب ویژگی ضریب همبستگی پیرسون است که رابطه خطی بین دو متغیره کمی را نشان می‌دهد. این شاخص مقادیر بین یک (همبستگی کامل) و صفر (عدم همبستگی) را نشان می‌دهد (۴). بعد از انتخاب ویژگی‌ها با کمک مدل‌سازی می‌توان دانش یا قانون نهفته‌ای را در ورای داده‌ها کشف کرد. از جمله روش‌های مدل‌سازی استفاده از الگوریتم درخت تصمیم می‌باشد. درخت تصمیم یکی از پرکاربردترین روش‌های یادگیری ماشین (machine learning) می‌باشد که در سال ۱۹۶۰ ایجاد شد که به‌صورت ساختار درخت می‌باشد که می‌توانند با توجه به قوانینی داده هدف را پیش‌بینی کند. درخت تصمیم قادر به تولید توصیفات قابل درک برای انسان می‌باشد و مجموعه زیادی از داده‌ها در مدت زمان کوتاهی آنالیز می‌شوند (۲۶). مدل درختی طبقه‌بندی و رگرسیون، یک روش ناپارامتری الگوریتمی است که قادر به پیش‌بینی متغیرهای کمی یا متغیرهای طبقه‌بندی شده بر اساس مجموعه‌ای از متغیرهای پیش‌بینی کننده کمی و کیفی است (۳۱). مدل درخت تصمیم بر خلاف مدل شبکه عصبی مصنوعی به تولید قانون می پردازد. در ساختار درخت تصمیم، پیش‌بینی به دست آمده در قالب یک سری قواعد توضیح داده خواهد شد. همچنین در درخت تصمیم بر خلاف شبکه عصبی مصنوعی، ضرورتی وجود ندارد که داده‌ها لزوماً به صورت عددی باشند (۹). درختان تصمیم ابزاری هستند که قابلیت پاسخگویی به مسائل پیچیده و غیرخطی را دارند و از آنجایی که در علوم دیگر از جمله برق و الکترونیک که اساسا دارای مسائل غیرخطی هستند به خوبی از عهده حل مسائل بر آمده‌اند، لذا کم کم پای این ابزار جدید به علوم دیگر که جزء پدیده‌های غیرخطی و پیچیده هستند نیز باز شده است (۲۰). پژوهش‌های اندکی در حوزه‌های مختلف خاکشناسی نیز از درخت تصمیم برای پیش‌بینی و مدل‌سازی استفاده کرده‌اند. شیرانی و همکاران (۲۶) طی مقایسه‌ای که بین الگوریتم ترکیبی PSO-DT و رگرسیون چند متغیر در تعیین ویژگی‌های مؤثر بر شاخص فیزیکی خاک‌های آهکی شهرستان بافت داشتند، پی بردند که کارایی درخت تصمیم و PSO خیلی بهتر از رگرسیون چند متغیره خطی بود. همین‌طور نتایج پژوهش‌های نبی‌الهی و همکاران (۱۸) نشان داد که مدل‌های درختی دارای دقت بالاتری در پیش‌بینی تهیه نقشه رقومی خاک نسبت به روش شبکه عصبی مصنوعی بوده و همچنین تفسیر نتایج مدل درختی بسیار راحتتر می‌باشد. مطالعات تقی زاده مهرجردی و همکاران (۳۱) که به منظور پیش بینی مکانی گروه‌های بزرگ خاک با استفاده از انواع روش‌های داده‌کاوی در منطقه اردکان یزد انجام شد، نشان داد که مدل درخت تصمیم نسبت به سایر روش‌های داده‌کاوی (شبکه عصبی، ترکیب شبکه عصبی مصنوعی- الگوریتم ژنتیک، رگرسیون لاجیستیک چندجمله‌ای و آنالیز تشخیصی) دارای عملکرد بهتری می‌باشد. پژوهش حاضر به‌منظور تعیین ویژگی‌های مؤثر بر شاخص‌های کیفیت فیزیکی خاک با استفاده شاخص پیرسون انجام شد و در ادامه پیش‌بینی مقادیر آن‌ها با الگوریتم درخت تصمیم برای این شاخص‌ها مدل‌سازی صورت گرفت و با مدل رگرسیون چند متغیره خطی مقایسه شد.

مواد و روش‌ها

منطقه‌ی مورد مطالعه شهرستان رابر، واقع در جنوب غربی استان کرمان برای این پژوهش انتخاب شد که در حد فاصل عرض‌های جغرافیایی "۲۱ '۱۶ °۲۹ تا "۵۶ '۱۶ °۲۹ شمالی و طول‌های جغرافیایی"۴۶ '۱ °۵۷ تا "۲۰ '۲ °۵۷ شرقی قرار گرفته است. شهرستان رابر، شهر مرتفعی است که ارتفاعی معادل ۲۳۴۳ متر از سطح دریا دارد. منطقه دارای اقلیم نیمه‌خشک، معتدل تا سرد با میانگین دمای سالیانه‌ی ۱۵ درجه‌ی سلسیوس و میانگین بارندگی سالیانه‌ی ۲۵۰ میلی‌متر با رژیم رطوبتی اریدیک ضعیف می‌باشد. در منطقه‌ی مطالعاتی، ابتدا نقشه‌ی توپوگرافی منطقه با مقیاس ۱:۲۵۰۰۰ تهیه شد. پس از پیاده کردن محدوده‌ی مطالعاتی بر روی این نقشه و اسکن نمودن آن، با استفاده از نرم‌افزار ایلویس[1]، منطقه‌ی مزبور زمین مرجع[2] گردید.

1- ILWIS
2- Georeference



شکل ۱- شمایی از منطقه‌ی مورد مطالعه به همراه نقاط نمونه‌برداری
Figure 1- The general status of the region and sampled sites

و بدین ترتیب مختصات جغرافیایی ۱۰۴ نقطه‌ی مشاهداتی به دست آمد. پس از آن، ابتدا موقعیت هر یک از نقاط مشاهداتی با استفاده از سامانه‌ی موقعیت‌یاب جهانی[1] مشخص شد (شکل ۱). سپس ۱۰۴ نمونه از اراضی کشاورزی از عمق ۰ تا ۱۵ سانتی‌متری از سطح خاک با استفاده از مته، نمونه‌برداری انجام گرفت و نمونه‌های برداشت‌شده به آزمایشگاه منتقل شدند. هم‌چنین نمونه‌های دست نخورده برای اندازه‌گیری چگالی ظاهری خاک و تعیین منحنی رطوبتی با استفاده از دستگاه صفحه‌ی فشار توسط استوانه‌ای به قطر داخلی ۵۲ میلی‌متر و ارتفاع ۴۴ میلی‌متر برداشت شدند.

تجزیه‌های آزمایشگاهی

پس از هوا خشک نمودن نمونه‌های برداشت‌شده و عبور آن‌ها از الک دو میلی‌متری، تجزیه‌های آزمایشگاهی لازم بر روی آن‌ها انجام شد. تعیین بافت خاک به روش هیدرومتری (۵)، کربنات کلسیم معادل به روش خنثی‌سازی با اسیدکلریدریک (۱)، واکنش خاک (pH) در گل اشباع با دستگاه پ‌هاش‌متر (۲۴)، چگالی حقیقی به روش پیکنومتر (۶)، چگالی ظاهری به روش سیلندر (۵)، هدایت الکتریکی (ECe) عصاره اشباع خاک با استفاده از دستگاه هدایت سنج (۲۴)، کربن آلی نیز به روش والکلی و بلک (۳۵) انجام شد. مقدار تخلخل نیز از رابطه زیر محاسبه شد:

$$f = 1 - \rho b/\rho s \qquad (۱)$$

که ρb چگالی ظاهری و ρs چگالی حقیقی می‌باشد.

شاخص‌های کیفیت فیزیکی خاک

شاخص ظرفیت هوایی (AC)

ظرفیت هوایی، AC (m^3m^{-3}) به‌صورت زیر تعریف می‌شوند.

1- Global Positioning System; GPS

(۳۵)

$$AC = \theta_S(\Psi = 0) - \theta_{FC}(\Psi = -1m); 0 \le AC \le \theta_S \qquad (۲)$$

که θS ($m^3\ m^{-3}$)، رطوبت حجمی خاک اشباع، θFC ($m^3\ m^{-3}$)، رطوبت حجمی گنجایش زراعی و Ψ(m)، مکش یا پتانسیل ماتریک منافذ می‌باشد. اگر $AC \ge 0.14\ m^3\ m^{-3}$ به‌عنوان حداقل حساسیت برای آسیب به گیاه یا کاهش عملکرد تعریف می‌شود و نشان‌دهنده کمبود هوادهی در منطقه ریشه در خاک‌های لومی شنی تا لومی رسی می‌باشد (۱۷ و ۲۳). محدوده شاخص ظرفیت هوایی در جدول ۱ آورده شده است.

شاخص ظرفیت زراعی نسبی (RFC)

ظرفیت زراعی نسبی به‌صورت زیر تعریف می‌شود (۲۳):

$$RFC = \left(\frac{\theta_{FC}}{\theta_S}\right) = \left[1 - \left(\frac{AC}{\theta_S}\right)\right]; 0 \le RFC \le 1 \qquad (۳)$$

ظرفیت زراعی نسبی، قابلیت خاک برای ذخیره‌سازی آب و هوا نسبت به حجم کل منافذ خاک (که با θS نشان داده می‌شود) را بیان می‌کند. محدوده ظرفیت زراعی نسبی در جدول ۱ بیان شده است.

شاخص آب قابل‌دسترس گیاه (PAWC)

آب قابل‌دسترس گیاه، توانایی خاک برای ذخیره و فراهمی آب قابل‌دسترس برای ریشه گیاه نشان می‌دهد که به‌صورت زیر تعریف می‌شود (۳۵):

$$PAWC = \theta_{FC}(\Psi = -1m) - \theta_{PWP}(\Psi = -150m) \qquad (۴)$$

که θ_{PWP}، مقدار رطوبت در نقطه پژمردگی دائم است. محدوده شاخص آب قابل‌دسترس گیاه در جدول ۱ آورده شده است.

شاخص‌های کیفیت فیزیکی خاک و رنج دامنه بحرانی آن‌ها در

جدول ۱ تعریف شده است (۱۲، ۲۲ و ۱۴) که این حدود بر اساس محدودیت در تولید محصول بیان شده‌اند. در اینجا به‌طور خلاصه این شاخص‌ها توضیح داده می‌شوند.

جدول ۱- حد بحرانی شاخص‌های فیزیکی خاک

Table 1- Critical range of physical indicators of soil

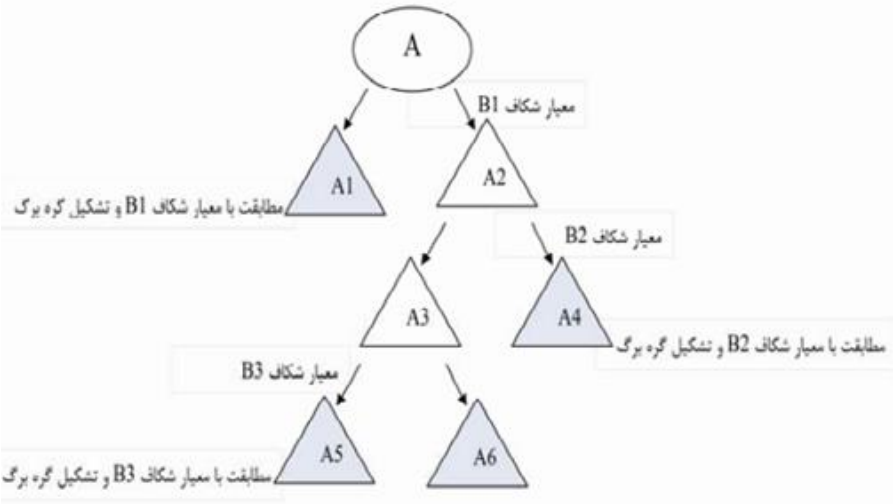
PAWC	RFC	AC	
>0.3	0.6-0.7	>0.2	No limit بدون محدودیت
0.2-0.3	0.5-0.6 و 0.7-0.75	0.18-0.2	Low limit محدودیت کم
0.08-0.2	0.75-0.8 و 0.4-0.5	0.15-0.18	Mid limit محدودیت متوسط
0.02-0.08	0.8-0.9 و 0.35-0.4	0.1-0.15	High limit محدودیت زیاد
<0.02	0.9< و 0.35>	<0.1	Very high limitمحدودیت شدید

انتخاب ویژگی‌های مؤثر بر شاخص‌های کیفیت خاک

هنگامی که تعداد ویژگی‌ها مورد استفاده برای فرآیند مدل‌سازی زیاد باشند، پیچیدگی مدل‌ها زیاد خواهد و صحت کلی آن‌ها کاهش خواهد یافت (۱۳) بنابراین به منظور کاهش تعداد متغیرهای ورودی و انتخاب مؤثرترین آن‌ها بر شاخص‌های ظرفیت هوایی، ظرفیت زراعی نسبی و آب قابل‌دسترس گیاه از روش انتخاب ویژگی (Feature selection) بر اساس ضریب همبستگی پیرسون توسط نرم‌افزار Clementine نسخه ۱۲ استفاده شد. ضریب همبستگی پیرسون میزان ارتباط خطی دو متغیر کمی را می‌دهد این ضریب همبستگی بین متغیرهای ورودی و متغیر هدف را نشان می‌دهد که هرچقدر این ضریب نزدیک به یک باشد یعنی ارتباط و همبستگی بین متغیر ورودی و هدف وجود دارد. از بین ویژگی‌های اندازه گرفته شده، ویژگی‌هایی که ضریب همبستگی آن‌ها بیشتر یا مساوی ۰/۹ باشد را به‌عنوان مهم و کمتر از آن را به‌عنوان غیرمهم انتخاب می‌شوند. خصوصیات مؤثر خاک بر شاخص کیفیت فیزیکی خاک انتخاب شدند و ۹ ویژگی واکنش خاک، قابلیت هدایت الکتریکی، شن، سیلت، رس، ماده آلی، چگالی ظاهری، کربنات کلسیم و تخلخل به‌عنوان ورودی مدل و شاخص‌های ظرفیت هوایی، ظرفیت زراعی نسبی و آب قابل‌دسترس گیاه به‌عنوان خروجی مدل در نظر گرفته شدند.

پیش‌بینی شاخص‌های کیفیت فیزیکی خاک با الگوریتم درخت تصمیم

در این پژوهش پس از انتخاب ویژگی‌های مؤثر بر شاخص‌های کیفیت فیزیکی، مدل‌سازی این شاخص‌ها با درخت تصمیم انجام شد و مقادیر پیش‌بینی شده ظرفیت هوایی، ظرفیت زراعی نسبی و آب قابل‌دسترس گیاه به‌عنوان متغیر هدف تخمین زده شدند. درخت تصمیم یکی از روش‌های داده‌کاوی و از ابزارهای قوی و متداول برای دسته‌بندی و پیش‌بینی می‌باشد. درخت تصمیم پیش‌بینی خود را در قالب یک سری قوانین توضیح می‌دهد (شکل ۲).



شکل ۲- یک نمونه ساده از درخت تصمیم

Figure 2- A sample of simple decision tree.

در این پژوهش از الگوریتم داده‌کاوی درخت تصمیم رگرسیونی (tree regression and classification) به عنوان یکی از انواع درختان تصمیم جهت پیش‌بینی شاخص‌های کیفیت فیزیکی خاک استفاده شده است. ساختار این درختان بر سه اصل استوار است:

۱- مجموعه‌ای از سؤالات به شکل $x \leq d?$ که در آن x یک متغیر مستقل و d یک مقدار ثابت است و جواب هر سؤال بله/خیر است.

۲- بهترین معیار شاخه زدن جهت انتخاب بهترین متغیرمستقل برای ایجاد شاخه جدید.

۳- ایجاد آمار خلاصه برای گره انتهایی (۸).

معیارهای مختلفی جهت ایجاد شاخه و تولید درخت تصمیم وجود دارد، ولی از آنجا که تحقیق حاضر به استفاده از درخت تصمیم رگرسیونی پرداخته، معیار مورد استفاده در این مدل که حداقل مربعات خطا $(LSD)^2$ نام دارد تشریح می‌گردد. این معیار به صورت زیر تعریف می‌شود:

$$SS(t) = \sum_{i=1}^{N_{tt}} (y_i(t) - \bar{y}(t))^2 \quad (۵)$$

N_t: تعداد رکوردها (داده‌ها) در گره برگ t .

$y_i(t)$: مقدار خروجی (متغیر هدف در گره برگ).

$\bar{y}(t)$: میانگین مقادیر متغیر هدف برای همه گره‌ها.

حال متغیر ورودی S زمانی بهترین متغیر برای ایجاد شاخه در گره t می‌باشد که مقدار $Q(s,t)$ را بیشینه نماید.

$$Q(s,t) = SS(t) - SS(t_R) - SS(t_l) \quad (۶)$$

که در آن $SS(t_R)$ و $SS(t_l)$ به ترتیب میزان $SS(t)$ در شاخه سمت راست و سمت چپ گره t می‌باشد.

در این فرآیند ۷۰ درصد داده‌ها به‌عنوان داده آموزشی و ۳۰ درصد به‌عنوان داده آزمون در نظر گرفته شد. هنگام اجرای الگوریتم، ابتدا مدل با استفاده از داده‌های آموزش ایجاد شد و در گام بعدی با داده‌های آزمایش مورد آزمون قرار گرفت. در آخر نمودار رگرسیونی برای داده هدف واقعی و پیش‌بینی شده مدل رسم شد که تمامی این مراحل در نرم‌افزار MATLAB انجام شد.

رگرسیون چند متغیره خطی

رگرسیون خطی یکی از قدیمی‌ترین تکنیک‌های آماری است. رگرسیون چند متغیره روشی برای مشارکت جمعی و فردی دو یا چند متغیر مستقل در تغییرات یک متغیر وابسته است. برای مدل‌سازی توسط رگرسیون چند متغیره‌ی خطی ۷۰ درصد داده‌ها به‌عنوان آموزشی و ۳۰ درصد به‌عنوان داده‌های آزمون در نظر گرفته شدند. در نهایت بین متغیرهای انتخاب شده و متغیر هدف (ظرفیت هوایی، ظرفیت زراعی نسبی و آب قابل‌دسترس گیاه)، رگرسیون گرفته شد. در نهایت مدل رگرسیونی برای تخمین شاخص‌های فیزیکی ذکر شده به دست آمد. در اینجا از نرم‌افزار SPSS برای انجام محاسبات استفاده شد.

شاخص‌های بررسی اعتبار مدل

در این پژوهش، به‌منظور بررسی عملکرد و دقت مدل‌ها از شاخص‌های ضریب تبیین (R2)، جذر میانگین مربعات خطا (RMSE)، میانگین مطلق خطا (MAE) استفاده شد.

$$R^2 = \sum_{i=1}^{n}(Pi - \overline{O})^2 / \sum_{i=1}^{n}(Oi - \overline{O})^2 \quad (۷)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{[p_i - o_i]}{Oi}\right| \quad (۸)$$

$$RMSE = [\frac{1}{n}\sum_{i=1}^{n}(Pi - Oi)^2]^{1/2} \quad (۹)$$

که در روابط بالا Oi: مقدار مشاهده‌شده، Pi: مقدار پیش‌بینی شده، n: تعداد مشاهدات و $\overline{O}$: میانگین مقادیر مشاهداتی می‌باشند.

آنالیز حساسیت

فرآیند آنالیز حساسیت، میزان حساسیت مدل را نسبت به متغیرهای ورودی آن نشان می‌دهد که برای انجام آنالیز حساسیت مدل از روش استات سافت استفاده گردید (۳۰). در این روش مقادیر ضریب حساسیت متغیرهای ورودی از تقسیم نمودن خطای کل شبکه در غیاب یک متغیر بر خطای کل شبکه در حضور تمامی متغیرهای ورودی، به دست می‌آید. بر این اساس اگر مقدار ضریب حساسیت یک متغیر بیشتر از یک باشد، آن متغیر سهم زیادی در توضیح تغییرپذیری مؤلفه‌های عملکرد دارد (۱۹).

نتایج و بحث

توصیف آماری خصوصیات خاک

توصیف آماری خصوصیات خاک انتخاب شده برای بررسی کیفیت خاک برای ۱۰۴ نقطه و کاربری‌های مختلف در جدول ۲ ارائه شده است. همان‌گونه که جدول ۲ نشان می‌دهد منطقه‌ی مورد مطالعه دارای مقادیر شن بالا می‌باشد، به‌طوری که بافت غالب منطقه لوم شنی است. طبق جدول ۲ کم‌ترین درصد ضریب تغییرات مربوط به pH (۲/۷۸) می‌باشد که دلیل آن را می‌توان خاصیت بافری بالا و آهکی بودن خاک منطقه مورد مطالعه عنوان کرد. همچنین نتایج به‌دست آمده، نشان می‌دهد که واکنش خاک منطقه‌ی مورد مطالعه، در محدوده‌ی قلیایی می‌باشد. بیشترین درصد ضریب تغییرات مربوط به ماده آلی (۸۲/۴) می‌باشد که زیاد بودن ضریب تغییرپذیری برای ماده‌ی آلی، احتمالاً به دلیل اضافه کردن کودهای آلی در باغات نسبت به سایر کاربری‌ها می‌باشد که این باعث عدم یکنواختی در میزان ماده آلی در کاربری‌های مختلف شده است.

جدول ۲- خلاصه آماری خصوصیات فیزیکی و شیمیایی خاک منطقه مورد مطالعه

Table 2- Traditional statistics of physicochemical properties of soils under studied area

خصوصیات Parameter	میانگین Mean	میانه Median	کمینه Minimum	بیشینه Maximum	ضریب تغییرات Coefficient of variation (%)
قابلیت هدایت الکتریکی EC(dS m^{-1})	0.69	0.6	0.325	2.21	44.94
شن Sand (%)	55.9	56.5	36.5	85	13.5
رس Clay (%)	12.6	12	5.5	23.5	31.5
سیلت Silt (%)	31.5	31.5	9.5	43.5	15.8
ماده آلی Organic matter (%)	2.2	1.39	0.24	7.93	82.4
چگالی ظاهری Bulk density (g cm^{-3})	1.27	1.27	0.96	1.66	10.5
چگالی حقیقی density Particle (g cm^{-3})	2.37	2.38	1.98	2.74	5.77
کربنات کلسیم معادل Lime (%)	12.1	10.4	1	4.73	54.5
واکنش خاک (pH)	7.75	7.8	6.74	7.99	2.78
تخلخل Porosity	0.46	0.46	0.32	0.57	10.7

انتخاب ویژگی‌های مؤثر بر شاخص‌های کیفیت خاک با ضریب همبستگی پیرسون

جدول ۳ ویژگی‌های مهم انتخاب شده بر روی شاخص‌های کیفیت فیزیکی خاک را نشان می‌دهد. طبق ضریب همبستگی پیرسون برای ظرفیت هوایی ۶ ویژگی شامل؛ تخلخل، چگالی ظاهری، رس و شن که ضریب همبستگی پیرسون برای تخلخل و چگالی ظاهری یک و برای بقیه ۰/۹ بود. ظرفیت هوایی به‌طور مستقیم تابعی از تخلخل و منافذ خاک می‌باشد و همین طور جرم مخصوص ظاهری خاک نیز به‌طور واضح تابعی از میزان تخلخل و بر توزیع منافذ خاک اثر دارد (۲۳) درنتیجه تغییر در جرم مخصوص ظاهری خیلی از خصوصیات ساختمانی خاک را تغییر می‌دهد. به‌عنوان مثال با تغییر در جرم مخصوص ظاهری، تخلخل، توزیع اندازه منافذ، ظرفیت نگهداری آب و هوای خاک تغییر می‌کند. رابطه بین پارامتر AC و مقدار رس احتمالاً مربوط به تأثیر مقدار رس روی رطوبت خاک، شرایط تهویه‌ای یا اثر مثبت رس روی تخلخل و ساختمان خاک است. ظرفیت هوایی در مکش‌های خیلی کم محاسبه می‌شود. به‌طور کلی شاخص ظرفیت هوایی به‌طور مستقیم یا غیرمستقیم بیانی از منافذ خاک می‌باشد پس توزیع اندازه منافذ تأثیر زیادی روی این شاخص دارد (۱۰). خاک‌های منطقه‌ی مورد مطالعه دارای درصد بالایی از مقادیر شن بودند و در واقع اثری که ذرات شن بر روی شاخص ظرفیت هوایی دارد به علت منافذ درشت آن‌ها می‌باشد که می‌توانند بر نگهداری و انتقال آب در خاک مؤثر باشند.

از بین ۹ خصوصیت خاکی که به‌عنوان ورودی برای تعیین عوامل مؤثر بر شاخص ظرفیت هوایی وجود داشت، ۵ خصوصیت شامل، قابلیت هدایت الکتریکی، سیلت، کربنات کلسیم، واکنش خاک و ماده آلی تأثیر قابل ملاحظه‌ای بر شاخص AC خاک نداشتند و به‌عنوان پارامترهای اضافی در نظر گرفته شدند. هدایت الکتریکی نشان‌دهنده‌ی غلظت املاح محلول می‌باشد، به‌گونه‌ای که املاح محلول از طریق کاهش ضخامت لایه‌ی دو گانه پخشیده می‌توانند باعث هماوری ذرات رس گردند. ولی از آنجایی که مقادیر هدایت الکتریکی اندازه‌گیری شده در نمونه‌های خاک مورد مطالعه پایین بود (جدول ۲) اثر آن بر بهبود ساختمان و منافذ خاک و شاخص AC مشهود نبود. در رابطه با اثرات ذرات سیلت می‌توان گفت، از آنجایی که عمده ذرات خاک‌های منطقه مورد مطالعه شنی بود، ذرات شن به‌عنوان نماینده ذرات عمل نموده و اثر ذرات سیلت بر روی شاخص AC قابل ملاحظه نبود. در برخی موارد آهک باعث کاهش شدت نفوذ می‌گردد که دلیل آن قرار گرفتن ذرات کربنات کلسیم در اندازه رس و سیلت و بسته شدن منافذ می‌شود (۱۵). همان طور که گفته شد، ظرفیت هوایی تابعی از تخلخل و حجم منافذ خاک است پس واکنش خاک اثری بر روی این شاخص ندارد.

همچنین برای ظرفیت زراعی نسبی، ۴ ویژگی شامل؛ تخلخل، شن، رس و سیلت با ضریب همبستگی ۰/۹ ویژگی‌های مهم انتخاب

شدند. ظرفیت زراعی نسبی رابطه مستقیم با ظرفیت هوایی دارد و همانند آن تحت تأثیر حجم منافذ خاک می‌باشد. از آنجائی که تخلخل و بافت خاک بر ظرفیت هوایی مؤثر می‌باشند بر RFC نیز تأثیرگذار هستند. و سایر ویژگی ها به دلایلی که برای ظرفیت هوایی گفته شد، برای ظرفیت زراعی نسبی نیز انتخاب نشدند.

۷ متغیر، قابلیت هدایت الکتریکی، رس، شن، سیلت، چگالی ظاهری و تخلخل همگی با ضریب یک به‌عنوان مهمترین متغیرهای انتخابی شاخص آب قابل‌دسترس گیاه انتخاب شدند. آب قابل‌دسترس در واقع در محدوده‌ی بین ظرفیت زراعی نسبی و نقطه پژمردگی قرار می‌گیرد و درنتیجه عوامل زیادی می‌توانند در این محدوده بر آب فراهمی برای گیاه مؤثر باشد. ساختمان خاک، تخلخل، نوع بافت خاک، حجم و توزیع منافذ ازجمله عوامل تأثیرگذار بر آب قابل‌دسترس گیاه می‌باشند. ویژگی‌های انتخاب شده برای آب قابل‌دسترس گیاه در اینجا بافت خاک، تخلخل، چگالی ظاهری و قابلیت هدایت الکتریکی بودند که همگی بر حجم منافذ و توزیع پذیری آن‌ها اثر دارند و درنتیجه می‌توانند بر آب قابل‌دسترس نیز اثر بگذارند. دو ویژگی کربنات کلسیم و واکنش خاک به عنوان ویژگی‌های غیرمهم شناخته شدند. همان طور که گفته شد آهک باعث کاهش شدت نفوذ و بسته شدن منافذ می‌شود.

جدول ۳- ویژگی‌های انتخاب شده مربوط به شاخص‌های کیفیت فیزیکی خاک

Table 3- Selected features of soil physical quality indicators

AC		RFC		PAW	
ویژگی انتخاب شده Selected feature	ضریب پیرسون Pearson coefficient	ویژگی انتخاب شده Selected feature	ضریب پیرسون Pearson coefficient	ویژگی انتخاب شده Selected feature	ضریب پیرسون Pearson coefficient
تخلخل Porosity	1	تخلخل Porosity	0.996	چگالی ظاهری Bulk density	1
چگالی ظاهری Bulk density	1	شن Sand	0.993	رس Clay	1
رس Clay	0.999	رس Clay	0.98	قابلیت هدایت الکتریکی EC	1
شن Sand	0.995	سیلت Silt	0.97	تخلخل Porosity	1
				شنSand	1
				سیلت Silt	1

پیش‌بینی شاخص‌های کیفیت فیزیکی خاک

پس از انتخاب ویژگی‌های مؤثر بر شاخص ظرفیت هوایی، ظرفیت زراعی نسبی و آب قابل دسترس خاک، پیش‌بینی آن با استفاده از مدل درخت تصمیم و رگرسیون چند متغیره خطی صورت گرفت. همچنین آماره‌های اعتبار سنجی مربوط به درخت تصمیم در جدول ۴ آورده شده‌اند که با استفاده از داده‌های آموزشی و آزمون محاسبه شده‌اند.

نتایج داده های اعتبار سنجی مربوط به جدول ۴ برای داده‌های آموزشی و آزمون نشان می‌دهد که درخت تصمیم با دقت قابل قبولی توانسته است رابطه بین پارامترهای ورودی با شاخص‌های ظرفیت هوایی، ظرفیت زراعی نسبی و آب قابل دسترس را به‌دست آورد.

ظرفیت هوایی

شکل ۳ درخت تصمیم رسم شده برای شاخص ظرفیت هوایی را نشان می‌دهد. همان طور که دیده می‌شود در گره ریشه متغیر تخلخل قرار گرفته است که اگر تخلخل کمتر از ۰/۴۳۴ در یک انشعاب و تخلخل بیشتر و مساوی ۰/۴۳۴ در انشعاب دیگر قرار می‌گیرد و همین طور انشعابات ادامه پیدا می‌کند تا به گره برگ برسد. فقط ۱۰ درصد خاک‌های مورد مطالعه از نظر تخلخل تهویه‌ای، بدون محدودیت برای رشد گیاه بودند و بیشتر این خاک‌ها دارای محدودیت متوسط (۳۰ درصد) و ۲۲ درصد دارای محدودیت کم بودند. همچنین در ۱۷ درصد خاک‌ها محدودیت شدید و ۱۶ درصد محدودیت زیاد وجود داشت که این خاک‌ها از نظر تهویه باید مدیریت شوند.

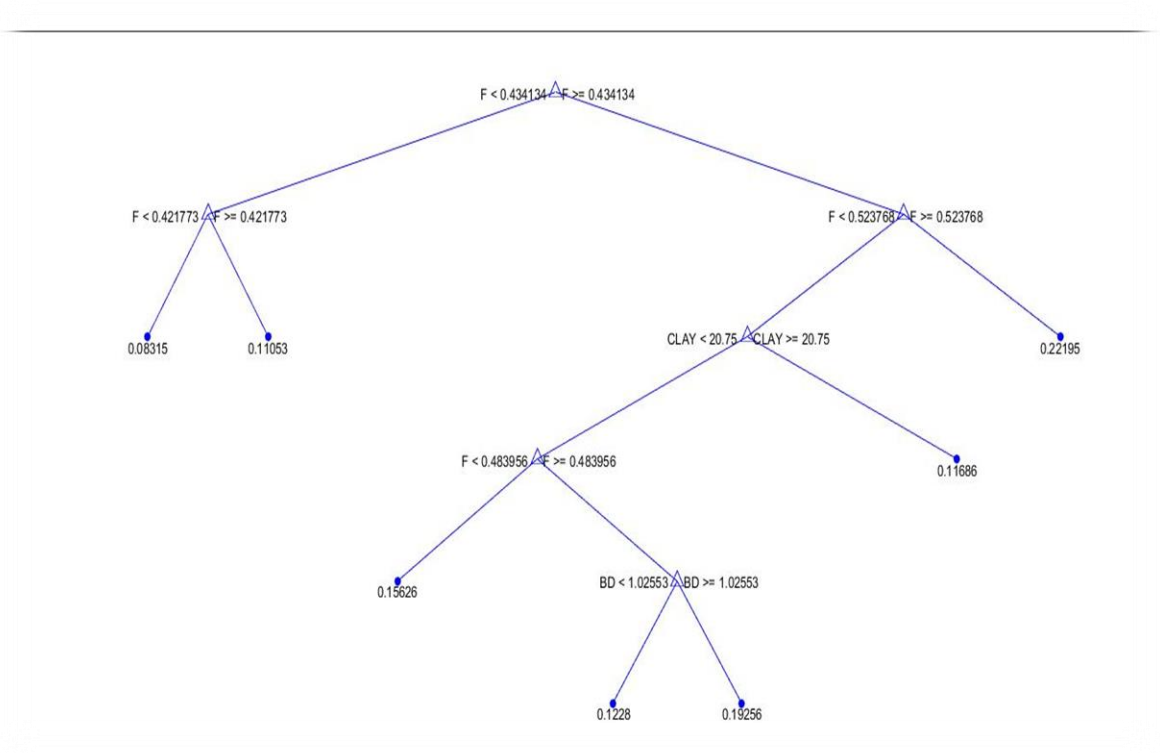
به‌منظور ارزیابی مدل درخت تصمیم، مقادیر برآورده شده برای نمونه‌های آزمونی و مقادیر واقعی متناظر با آن‌ها ترسیم شد. مقایسه مقادیر عملکرد مشاهده شده (Target) و عملکرد پیش‌بینی شده توسط مدل (Output) درخت تصمیم در نرم‌افزارMATLAB برای داده‌های آموزشی و آزمون در ادامه آورده شده است. ضریب تبیین (R^2) بین شاخص AC مشاهداتی و پیش‌بینی شده با استفاده از درخت تصمیم برای داده‌های آموزشی ۰/۹۶ بود (شکل ۴ – A). جذر

میانگین مربعات خطا و میانگین مطلق خطا بین شاخص AC مشاهداتی و پیش‌بینی شده با استفاده از درخت تصمیم برای داده‌های آموزشی به ترتیب ۰/۰۱۳۸ و ۰/۰۰۹۸ بود (جدول ۴).

جدول ٤- معیارهای ارزیابی در برآورد شاخص‌های کیفیت فیزیکی خاک برای مدل درخت تصمیم

Table 4- Values of evaluation criteria in predicting soil physical quality indicators for Decision tree model

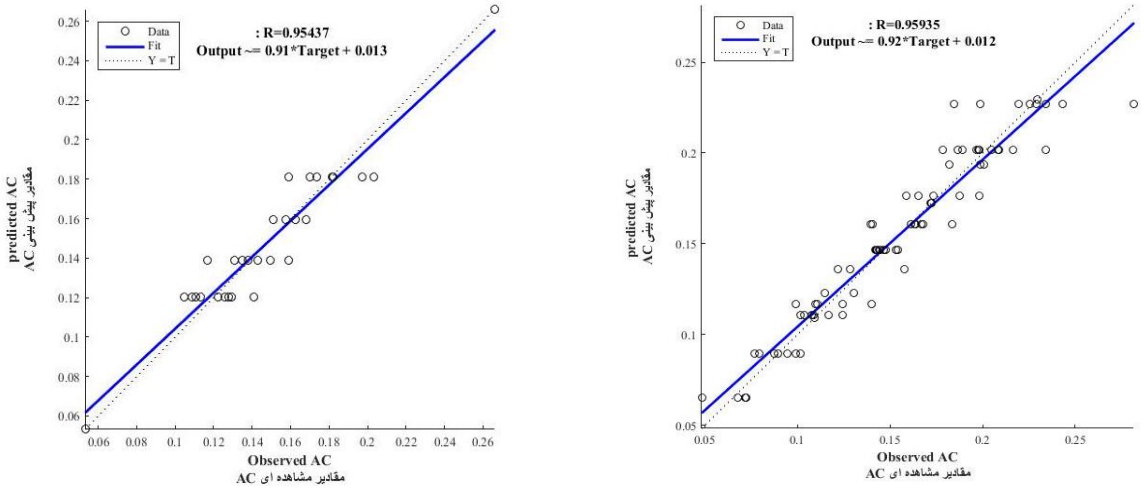
شاخص‌های فیزیکی Soil physical quality indicators	نوع داده Data Type	معیارهای ارزیابی Evaluation criteria		
		R^2	RMSE	MAE
ظرفیت هوایی	داده‌های آموزشی Train data	0.96	0.0138	0.0098
Air capacity	داده‌های آزمون Test data	0.95	0.0104	0.0084
ظرفیت نسبی زراعی	داده‌های آموزشی Train data	0.86	0.0461	0.0356
Relative field capacity	داده‌های آزمون Test data	0.84	0.0411	0.0332
آب قابل‌دسترس گیاه	داده‌های آموزشی Train data	0.92	0.0134	0.0085
Plant available water capacity	داده‌های آزمون Test data	0.85	0.0089	0.0068



شکل ۳- مدل درخت تصمیم برای شاخص ظرفیت هوایی

Figure 3- Decision tree model for Air Capacity Index

ضریب تبیین (R^2) بین شاخص AC مشاهداتی و پیش‌بینی شده با استفاده از درخت تصمیم برای داده‌های آزمون ۰/۹۶۱ بود (شکل ۴ – B). هم‌چنین جذر میانگین مربعات خطا و میانگین مطلق خطا برای داده‌های آزمون به ترتیب ۰/۰۱۰۴ و ۰/۰۰۸۴ درصد بود (جدول ۴).



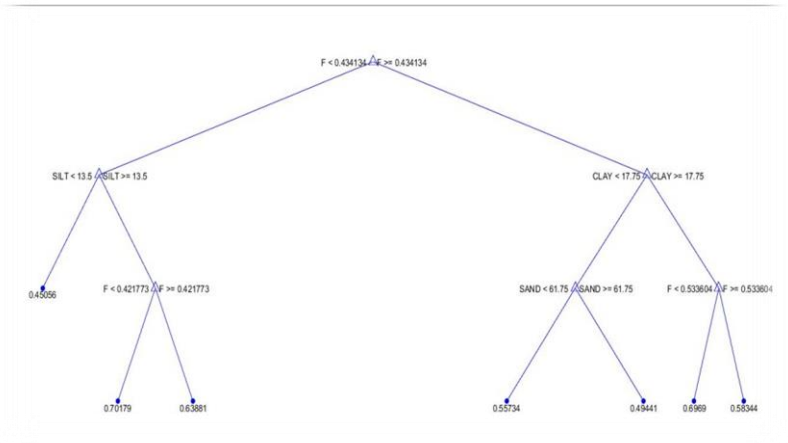
A B

شکل ٤- همبستگی بین مقادیر مشاهده‌ای و پیش‌بینی شده AC با استفاده از الگوریتم درخت تصمیم برای داده‌های آموزشی (A) و داده‌های آزمون (B)

Figure 4- Correlation between the observed and predicted AC index values using Decision -Tree Algorithm for the train data (A) and test data (B)

ظرفیت زراعی نسبی

شکل ۵ درخت تصمیم رسم شده برای داده‌های متغیر ظرفیت زراعی نسبی می‌باشد. بیشتر خاک مورد مطالعه (۳۸ درصد) از نظر شاخص ظرفیت زراعی نسبی بدون محدودیت و ۳۲ درصد محدودیت کم و ۱۶ درصد با محدودیت متوسط مشاهده شد.



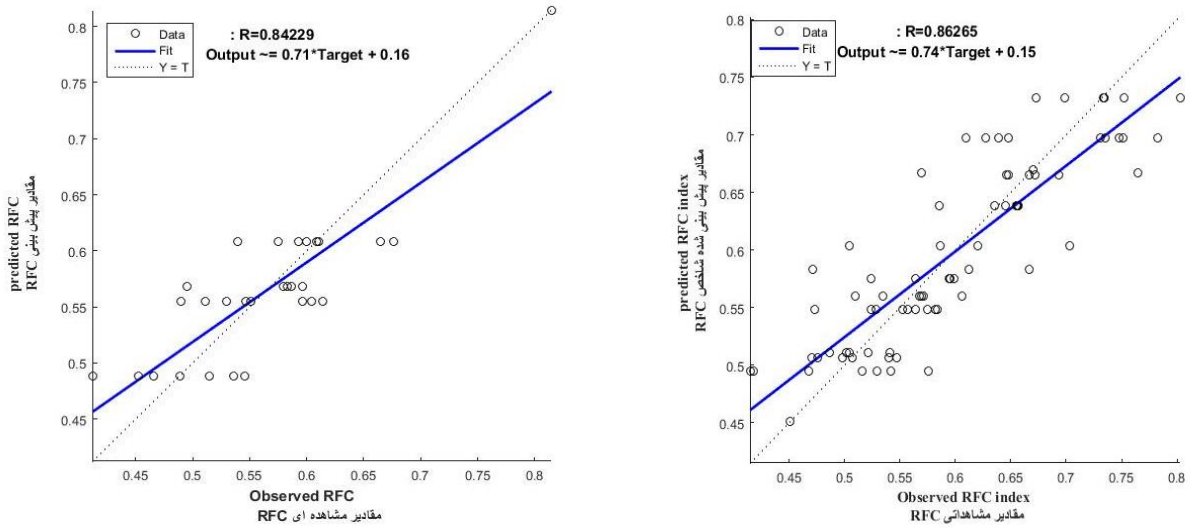
شکل ۵- مدل درخت تصمیم برای ظرفیت زراعی نسبی

Figure 5- Decision tree model for relative field capacity

ضریب تبیین (R^2) بین شاخص ظرفیت زراعی نسبی مشاهداتی و پیش‌بینی شده با استفاده از درخت تصمیم برای داده‌های آموزشی ۰/۸۶۲ بود (شکل ۶- A). جذر میانگین مربعات خطا و میانگین مطلق خطا بین شاخص ظرفیت زراعی نسبی مشاهداتی و پیش‌بینی شده با

استفاده از درخت تصمیم برای داده‌های آموزشی به ترتیب ۰/۰۴۶۱ و ۰/۰۳۵۶ بود (جدول ۴).

ضریب تبیین (R^2) بین شاخص ظرفیت زراعی نسبی مشاهداتی و پیش‌بینی شده با استفاده از درخت تصمیم برای داده‌های آزمون ۰/۸۴۲ بود (شکل ۶ – B). هم‌چنین جذر میانگین مربعات خطا و میانگین مطلق خطا برای داده‌های آزمون به ترتیب ۰/۰۴۱۱ و ۰/۰۳۳۲ درصد بود (جدول ۴).



شکل ٦- همبستگی بین مقادیر مشاهده‌ای و پیش‌بینی شده RFC با استفاده از الگوریتم درخت تصمیم برای داده‌های آموزشی (A) و داده‌های آزمون (B)

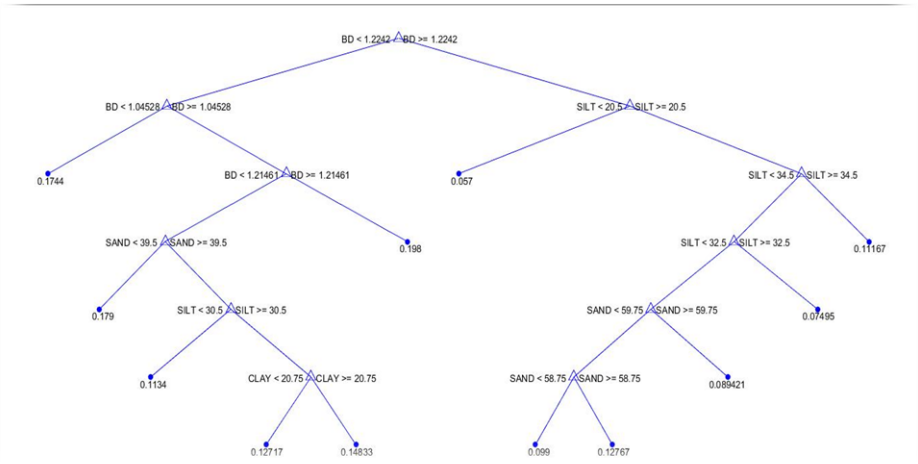
Figure 6- Correlation between the observed and predicted RFC index values using Decision -Tree Algorithm for the train data (A) and test data (B)

آب قابل‌دسترس گیاه

قوانین استخراج شده برای مدل درخت تصمیم برای آب قابل‌دسترس گیاه با توجه به شکل ۷ و جدول ۱ به‌صورت محدودیت متوسط تا زیاد می‌باشد. با توجه به داده‌های اندازه‌گیری شده، محدوده آب قابل‌دسترس برای تمامی داده بین ۰/۱۸ تا ۰/۰۱ می‌باشد که در این محدوده آب قابل‌دسترس گیاه دارای محدودیت متوسط تا زیاد می‌باشد. پس تمامی خاک مورد مطالعه از نظر شاخص PAWC در این محدوده بحرانی قرار می‌گیرند. با توجه به محدودیت آب قابل‌استفاده گیاه در این خاک‌ها اعمال روش‌های مناسب آبیاری و وجود آب کافی برای رشد گیاهان اجتناب‌ناپذیر است.

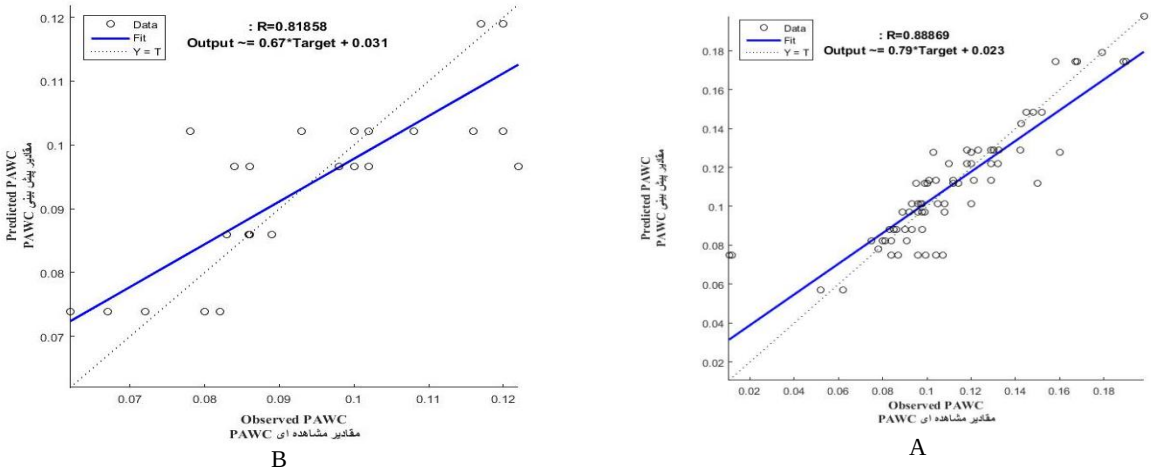
ضریب تبیین (R^2) بین شاخص آب قابل‌دسترس گیاه مشاهداتی و پیش‌بینی شده با استفاده از درخت تصمیم برای داده‌های آموزشی ۰/۹۱۹ بود (شکل ۸ – A). جذر میانگین مربعات خطا و میانگین مطلق درصد خطا بین شاخص آب قابل‌دسترس گیاه مشاهداتی و پیش‌بینی شده با استفاده از درخت تصمیم برای داده‌های آموزشی به ترتیب ۰/۰۱۳۴ و ۰/۰۰۸۵ بود (جدول ۴). ضریب تبیین (R^2) بین شاخص آب قابل‌دسترس گیاه مشاهداتی و پیش‌بینی شده با استفاده از درخت تصمیم برای داده‌های آزمون ۰/۸۵۰ بود (شکل ۸ – B). هم‌چنین جذر میانگین مربعات خطا و میانگین مطلق درصد خطا برای داده‌های آزمون به ترتیب ۰/۰۰۸۹ و ۰/۰۰۶۸ درصد بود (جدول ۴).

مقادیر معیارهای ارزیابی برای داده‌های آموزش نشان می‌دهد که درخت تصمیم با دقت خوبی توانسته است رابطه بین ظرفیت هوایی، ظرفیت زراعی نسبی و آب قابل‌دسترس گیاه و پارامترهای ورودی را به دست آورد. همچنین مقادیر معیاری ارزیابی داده‌های آزمون، نشان‌دهنده توانایی مدل درخت تصمیم در تخمین میزان ظرفیت هوایی، ظرفیت زراعی نسبی و شاخص آب قابل‌دسترس گیاه با ویژگی‌های انتخاب شده توسط نرم‌افزار Celementain با دقت بالایی می‌باشد.



شکل ۷- مدل درخت تصمیم برای آب قابل‌استفاده گیاه

Figure 7- Decision tree model for plant available water



شکل ۸- همبستگی بین مقادیر مشاهده‌ای و پیش‌بینی شده PAWC با استفاده از الگوریتم درخت تصمیم برای داده‌های آموزشی (A) و داده‌های آزمون (B)

Figure 8- Correlation between the observed and predicted PAWC index values using Decision -Tree Algorithm for the train data (A) and test data (B)

آماره‌های اعتبارسنجی (ضریب تبیین، جذر میانگین مربعات خطا و میانگین مطلق خطا) و معادله رگرسیونی مربوط به هر کدام از شاخص‌های فیزیکی برای داده‌های آموزشی و داده‌های آزمون در جدول ۵ آورده شده‌اند. ضریب تبیین بین شاخص‌های ظرفیت هوایی، ظرفیت زراعی نسبی و آب قابل‌دسترس مشاهداتی و پیش‌بینی شده برای داده‌های آموزشی رگرسیون چند متغیره‌ی خطی به ترتیب ۰/۷۲، ۰/۷۲ و ۰/۶۷ بود (جدول ۵). جذر میانگین مربعات خطا و میانگین مطلق خطا بین شاخص‌های فیزیکی مشاهداتی و پیش‌بینی شده برای داده‌های آموزشی و آزمون در جدول ۵ نشان داده شده است. مقادیر معیارهای ارزیابی برای داده‌های آموزش و آزمون نشان می‌دهد که مدل رگرسیون چند متغیره قادر به برقراری ارتباط مؤثر بین ورودی و خروجی مدل نبوده و قادر به پیش‌بینی مناسب شاخص‌های کیفیت فیزیکی خاک هم نبوده است. درواقع رگرسیون خطی تنها قادر به تشخیص روابط خطی بین متغیرهای مستقل و وابسته هستند و چنانچه روابط غیرخطی بین این متغیرها وجود داشته باشد، کارایی این مدل‌ها به حد قابل ملاحظه‌ای کاهش می‌یابد (۲۸). نتایج مدل‌سازی رگرسیون چند متغیره نشان داد که مدل‌های رگرسیون برای داده‌های خطی توانایی بالایی جهت پیش‌بینی دارند. در حالی که داده‌ها در طبیعت ممکن است به‌طور خطی نباشند و یا تغییرات زمانی و مکانی تغییر کنند. مدل‌های خطی مناسب توابع خطی با جفت داده‌های ورودی و خروجی مدل هستند. اما مدل‌های غیرخطی قادرند با تغییرات زمانی و مکانی تغییر کنند.

جدول ۵- مقادیر معیارهای ارزیابی در برآورد شاخص‌های کیفیت فیزیکی خاک برای مدل رگرسیون خطی چند متغیره

Table 5- Values of evaluation criteria in predicting soil physical quality indicators for multiple linear regression model.

شاخص‌های فیزیکی Soil physical quality indicators	نوع داده Data Type	معیارهای ارزیابی Evaluation criteria		
		R^2	RMSE	MAE
ظرفیت هوایی Air capacity	داده‌های آموزشی Train data	0.72	0.894	0.892
	داده‌های آزمون Test data	0.63	0.877	0.875
معادله رگرسیون AC= -0.324+(0.003*sand)+(-0.22*EC)+(-0.001*clay)+(-0.67*BD)+(0.965*F)				
ظرفیت نسبی زراعی Relative field capacity	داده‌های آموزشی Train data	0.72	0.065	0.052
	داده‌های آزمون Test data	0.62	0.066	0.0502
معادله رگرسیونی RFC= 0.917+(0.006*silt)+(0.014*clay)+(-1.51*F)				
آب قابل‌دسترس گیاه Plant available water capacity	داده‌های آموزشی Train data	0.67	0.59	0.574
	داده‌های آزمون Test data	0.61	0.64	0.644
معادله رگرسیونی PAWC= 0.143+(0.12*EC)+(0.006*silt)+(0.001*clay)+(-0.76*BD)+(0.12*F)				

مقایسه نتایج مدل‌های درخت تصمیم و رگرسیون چند متغیره‌ی خطی

با توجه به مباحث گفته شده نتایج حاکی از آن است که مقدار ضریب تبین در مدل درخت تصمیم نسبت به رگرسیون چند متغیره‌ی خطی از دقت بالاتری برخوردار می‌باشد. همین طور کمترین مقادیر شاخص‌های خطای محاسبه شده مربوط به مدل درخت تصمیم هستند پس با توجه به معیارهای ارزیابی برای مدل درخت تصمیم و مدل رگرسیون چند متغیره‌ی خطی مشخص شد که دقت مدل درخت تصمیم نسبت به رگرسیون چند متغیره‌ی خطی بیشتر است.

برخی محققین نشان دادند که دقت درخت تصمیم نسبت به رگرسیون چند متغیره‌ی خطی در پیش‌بینی‌های بعضی از خصوصیات خاک معتبرتر است. به عنوان مثال پهلوان و همکاران (۲۱) دریافتند که روش درخت تصمیم گیری نسبت به رگرسیون چند متغیره کارایی بهتری برای تهیه نقشه خاک در سطح پایین‌تر دارد. شیرانی و همکاران (۲۶) طی مقایسه‌ای که بین الگوریتم ترکیبی PSO-DT و رگرسیون چند متغیر در تعیین ویژگی‌های مؤثر بر شاخص فیزیکی خاک‌های آهکی شهرستان بافت داشتند، پی بردند که کارایی درخت تصمیم و PSO خیلی بهتر از رگرسیون چند متغیره خطی بود.

بررسی اهمیت متغیرهای ورودی بر میزان شاخص‌های کیفیت خاک

با توجه به این‌که فقط دقت الگوریتم درخت تصمیم در پیش‌بینی شاخص‌های ظرفیت هوایی، ظرفیت زراعی نسبی و آب قابل‌دسترس گیاه قابل قبول بود آنالیز حساسیت با درخت تصمیم صورت گرفت. نتایج آنالیز حساسیت برای مؤلفه‌های شاخص‌های فیزیکی به‌صورت نمودار آورده شده است (شکل ۹-۱۱).

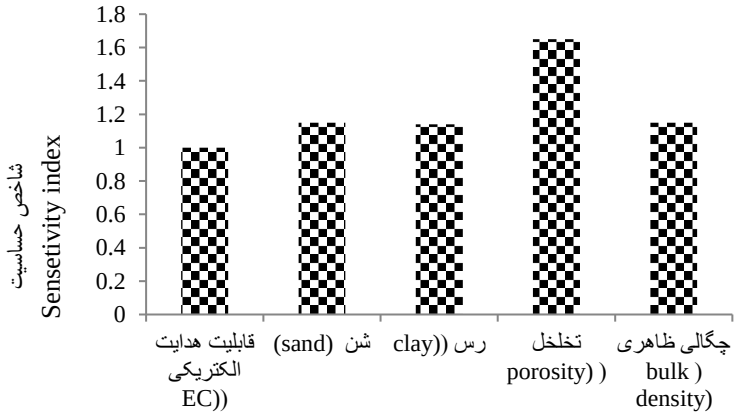
نتایج نشان داد که شاخص ظرفیت هوایی بیشترین حساسیت را به ترتیب به تخلخل (شاخص حساسیت ۱/۶۵)، چگالی ظاهری داشت (شاخص حساسیت ۱/۱۵) (شکل ۹). تخلخل تهویه‌ای با بافت خاک و تراکم خاک در ارتباط است و همان طور که قبلاً گفته شد ظرفیت هوایی رابطه مستقیم با تخلخل دارد و این عامل تأثیر مهمی بر ظرفیت هوایی دارد. از طرفی ظرفیت زراعی در مکش‌های خیلی کم محاسبه می‌شود (بین رطوبت اشباع و مکش ظرفیت زراعی). لذا توزیع اندازه منافذ تأثیر بسزایی بر روی این شاخص دارد. چگالی ظاهری نیز بر توزیع منافذ خاک اثر دارند.

همین طور شاخص ظرفیت زراعی نسبی بیشترین حساسیت را به

ترتیب به تخلخل (شاخص حساسیت ۱/۹۵) دارا بود (شکل ۱۰). ظرفیت زراعی نسبی نیز تابعی از توزیع منافذ درشت در خاک است درنتیجه تخلخل نیز بر توزیع و حجم منافذ تأثیرگذار هستند و چگالی ظاهری (شاخص حساسیت ۱/۲۵) نیز به‌عنوان پارامتر مهم و مؤثر آب قابل‌دسترس گیاه مشاهده شد (شکل ۱۱). چگالی ظاهری و ارتباط آن با تخلخل و نگهداشت آب در خاک می‌تواند از مهمترین عوامل بر PAWC باشد.

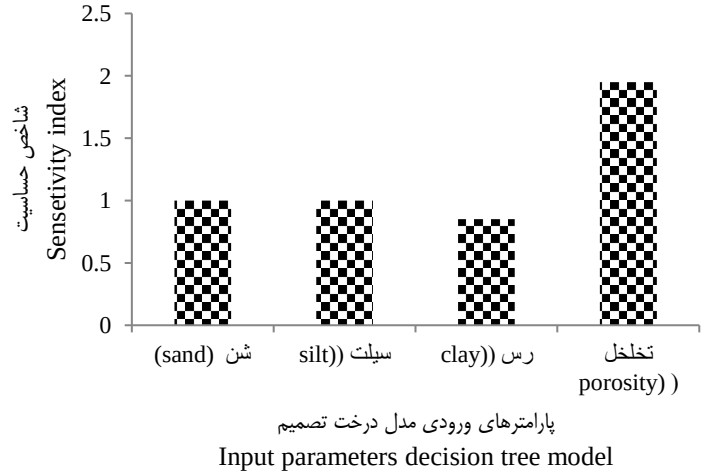
عوامل متعددی مانند کمبود میزان اکسیژن خاک، توانایی ریشه گیاه برای فروروی در خاک، کاهش هدایت هیدرولیکی غیراشباع با خشک شدن خاک، بر میزان آب قابل‌استفاده خاک برای گیاه مؤثر است. در مطالعات مشابه آنلاف و ررمان (۳) دریافتند که شاخص‌های کیفیت فیزیکی خاک به بافت خاک، چگالی ظاهری و مواد آلی بستگی دارد. رطوبت در مکش‌های پایین متأثر از منافذ درشتی است که در اثر ساختمان خاک ایجاد می‌شود و متغیری که بیشتر از همه توانسته تغییرات ساختمان را به‌صورت کمی نشان دهد جرم مخصوص ظاهری است.

جرم مخصوص ظاهری که مهمترین پارامتر کمی بیان کننده تغییرات ساختمان خاک و تابعی از خلل و فرج است که محل نگهداری رطوبت هستند درنتیجه با کمترین تغییر در آن بر آب قابل‌دسترس گیاه در خاک اثر می‌گذارد (۲۳).



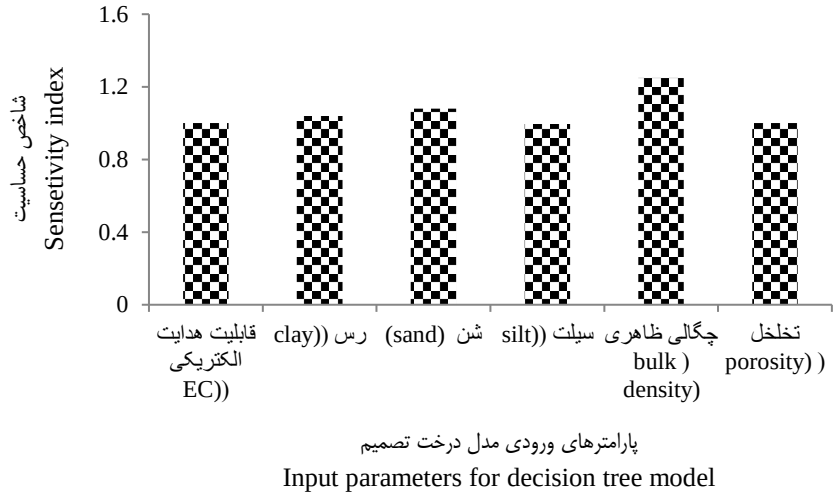
شکل ۹- نتایج آنالیز حساسیت مدل درخت تصمیم برای AC

Figure 9– Results of sensitivity analysis user decision tree model for AC



شکل ۱۰- نتایج آنالیز حساسیت مدل درخت تصمیم برای RFC

Figure 10– Results of sensitivity analysis user decision tree model for RFC



شکل ۱۱- نتایج آنالیز حساسیت مدل درخت تصمیم برای شاخص PAWC

Figure 11– Results of sensitivity analysis user decision tree model for PAWC

نتیجه‌گیری

نتایج این پژوهش نشان داد که الگوریتم درخت تصمیم توانسته است چهارچوب قوی و قابل قبولی برای انتخاب ویژگی‌های مؤثر بر شاخص‌های کیفیت فیزیکی خاک از میان ویژگی‌های زود یافت خاک فراهم آورد که برای مناطق و مسائل مشابه نیز قابل کاربرد می‌باشد. هم‌چنین نتایج نشان داد که دقت مدل درخت تصمیم در پیش‌بینی شاخص ظرفیت هوایی، ظرفیت زراعی نسبی و آب قابل‌دسترس نسبت به مدل رگرسیون خطی بهتر می‌باشد و می‌توان از این مدل در پیش‌بینی این شاخص‌ها با دقت قابل قبولی استفاده نمود.

منابع

1- Alison L. E., and Modie C. D. 1965. Carbonate. P. 1379-1396. In: C. A. Black et al. (Ed.), Methods of soil analysis. Part II, Am. Soc. Agron., Madison, WI.
2- Allmaras R.R., Fritz V.A., PFLEGER F.L., and Copeland S.M. 2002. Impaired internal drainage and Aphanomyces euteiches root rot of pea caused by soil compaction in a finetextured soil. Soil and Tillage Research, 1740:1 -12.
3- Anlauf R., and Rehrmann P. 2012. Effect of compaction on soil hydraulic parameters of vegetative landfill covers. Geomaterials, 2: 29–36.
4- Asghari Jafarabadi M., Soltani A., and Mohamadi S. M. 2013. Series Statistics; Correlation and regression. Iranian Journal of Diabetes and Lipid, 12(6): 479-506.
5- Blake G. R. 1965. Bulk density. Methods of Soil Analysis. Part 1 Society of Agronomy, Madison, Wisconsin, USA.
6- Blake G. R., and Hartage K. H. 1986. Particle density, In: Klute, A. (Ed.), Methods of Soil Analysis, Part 1, # 9(2nd ed.), Agronomy Monograph. American Society of Agronomy, Madison, 377-381.
7- Bouyoucos G. J. 1951. A recalibration of hydrometer method for making mechanical analysis of soil. Agronomy Journal, 43: 434-438.
8- Breiman L., Friedman J., Olshen R., and Stone C. 1984 Classification and Regression Trees, Chapman & Hall/CRC Press, Boca Raton, FL. Development of a decision tree modeling approach .Geoderma, 139:277-287.
9- Dastourani M. T ., Habibipoor A., Ekhtesasi M. R., Talebi A., and Mahjoobi J. 2013. Evaluation of the Decision Tree Model in Precipitation Prediction (Case study: Yazd Synoptic Station). Iran-Water Resources Research, 8(3): 14-27. (In Persian with English abstract).
10- Dexter A.R. 2004. Soil physical quality. Part I: Theory, effects of soil texture, density, and organic matter, and effects on root growth. Geoderma, 120:201-214.
11- Drury, C.F., T.Q. Zhang, and B.D. Kay. 2003. The non-limiting and least limiting water range for soil nitrogen mineralization. Soil Science Society of America Journal, 67:1388-1404.
12- Emami H. 2012. Investigation the stability of agricultural soils in field Karaj. Journal of Soil Research (Soil and

Water Sciences), 26:245-254.
13- Hughes G. F. 1968. On the Mean Accuracy of Statistical Pattern Recognizers. IEEE Transactions on Information Theory, 14, pp. 55-63.
14- Lal R. 1994. Soil Methods and guidelines for Sustainable use of soil and water resources in the tropics. Soil Management Support System, USDA,-NRCS. Washington, DC.
15- Mahmodabadi M., and Mazaheri M. 2012. Effect of some soil physical and chemical properties on permeability in field conditions. Jurnal of Engineering Water and Soil, 8: 14-26.
16- Moebius B.N., Van Es H.M., Schindelbeck R.R., Idowu O.J., Clune D.J., and Thies J.E. 2007. Evaluation of laboratory-measured soil physical properties and indicators of soil physical quality. Soil Science, 172: 895-912.
17- Mueller L., Kay B.D., Deen B., Hu C., Zhang Y., Wolff M., Eulenstein F., and Schindler U. 2009. Visual assessment of soil structure: Part II. Implications of tillage, rotation and traffic on sites in Canada, China and Germany. Soil and Tillage Research 103, 188-196.
18- Nabiollahi K., Haidari A., and Taghizadeh- Mehrjerdi R. 2014. Digital Mapping of Soil Texture Using Regression Tree and Artificial Neural Network in Bijar, Kurdistan. Journal of Water and Soil, 28(5):1025-1036. (In Persian with English abstract).
19- Norouzi M. 2009. Prediction of rainfed wheat yield using artificial neural network in Ardal district of Chaharmahal and Bakhtiari province. M.Sc. Thesis, Collage of Agriculture, Isfahan University of Technology, Isfahan, Iran, 112 p. (In Persian).
20- Omidvar K., Shafie SH., Taghizade Z., and Alipoor M. 2014. Evaluation of the Decision Tree Model in Precipitation Prediction Kermanshah Synoptic Station. Journal of Operational Research of Geographical Science, 14(34): 89-110. (In Persian).
21- Pahlavan Rad M., Khormali F., Tomanian N., Kiani F., and Komaki B. 2015. Predict soil classes using decision tree and logistic regression multivariate random on Golestan province. 14th Iranian Soil Science Congress, Vali-e-Asr University of Rafsanjan. 168-172.
22- Reynolds W.D., Bowman B.T., Drury C.F., Tan C.S., and X. Lu. 2002. Indicators of good soil physical quality: density and storage parameters. Geoderma, 110:131-146.
23- Reynolds W., Drury C., Yang X., and Tan C. 2008. Optimal soil physical quality inferred through structural regression and parameter interactions. Geoderma, 146, 466-474.
24- Richards L. A. 1954. Diagnosis and Improvement of Saline-Alkali Soils. U.S.D.A. Hand book, 60. Washington, D.C., U.S.A.
25- Shekofte H., Ramezani F., and Shirani H. 2017. Optimal feature selection for predicting soil CEC: Comparing the hybrid of ant colony organization algorithm and adaptive network-based fuzzy system with multiple linear regression. Geoderma, 298: 27 –34.
26- Shirani H., Habibi M., Besalatpour A.A., and Esfandiarpour I. 2015. Determining the features influencing physical quality of calcareous soils in a semiarid region of Iran using a hybrid PSO-DT algorithm, 259-260: 1-11.
27- Silva A.P., and Kay B.D. 1996. The sensitivity of shoot growth of corn to the least limiting water range of soils. Plant and Soil, 184: 323-329.
28- Singh M.J., and Khera K.L. 2009. Physical indicators of soil quality in relation to soil erodibility under different land uses. Arid Land Research and Management, 23:152-167.
29- Sobhani J., Najimi M., Pourkhorshidi A.R., and Parhizkar T. 2010. Prediction of the compressive strength of noslump concrete: A comparative study of regression, neural network and ANFIS models. Journal of Construction and Building Materials, 24: 709-718.
30- StatSoft Inc. 2004. Electronic Statistics Textbook (Tulsa, OK). http://www.statsoft.com/ textbook/stathome.html.
31- Taghizadeh-Mehrjardi R., Minasny B., Sarmadian F., and Malone P.B. 2013. Digital mapping of soil salinity in Ardakan region, central Iran. Geoderma, 213: 15-28.
32- Tan M., Tsang I. W., and Wang L. 2014. Towards ultrahigh dimensional feature selection for big data.J. Mach. Learn. Research., 15(1): 1371-1429.
33- Topp G.C., Reynolds W.D., Cook F.J., Kirby J.M., and Carter M.R. 1997. Physical attributes of soil quality. In: Gregorich, E.G., Carter, M.R. (Eds.), Soil Quality for Crop Production and Ecosystem Health. Developments in Soil Science, 25: 21 – 58.
34- Walkley A., and Black T. A. 1934. An examination of the Degtjareff method for determinating organic matter and a proposed modification of chromic acid titration method. Soil Sciences, 37: 29-38.
35- White R. 2006. Principles and practice of soil science 4th ed. Blackwell Publishing.

Journal of Water and Soil
Vol. 32, No. 2, May.-Jan. 2018, p. 327-342

نشریه آب و خاک (علوم و صنایع کشاورزی)
جلد ۳۲، شماره ۲، خرداد – تیر ۱۳۹۷، ص. ۳۴۲-۳۲۷

Determining Features Influencing Some Soil Physical Quality Indicators and their Predictions Using Decision Tree and Multiple Linear Regression Models

H. Shekofteh*[1]- M. Doustaky[2]- A. Masoudi [3]

Received: 08-11-2017
Accepted: 09-04-2018

Introduction: Soil quality is defined as the capacity of a soil to function within different land uses and ecosystem boundaries, sustain biological productivity, maintain environmental quality and promote plant, animal, and human health. Soil quality cannot be directly measured but can be evaluated on the basis of several parameters; the type of parameter to be used depends on research scale and goals. Soil quality indicators (SQIs) are used to evaluate the effect of different management and types of land use on soil quality and can be achieved by easily-measured soil physicochemical properties. Soil quality indicators are measurable characteristics of the soil affecting the soil capacity for crop production or environmental performance. Air capacity (AC), relative field capacity (RFC) and plant available water (PAWC) are the most important indicators. Selection of appropriate input parameters is the first and most important step in predicting SQIs. Feature selection can be defined as the identification and selection of a subset of useful features among the primary data collected. One of the methods for choosing the features is the Pearson coefficient, which shows the correlation between the input variables and target variable. When the coefficient is close to one, there is a strong relationship between the input and the target variable. The features having a correlation coefficients of greater than or equal to 0.9 are considered important and less than that are considered non-important. Decision tree algorithm is one of the prediction approaches in statistics and data mining literature. This algorithm can select the property with the highest separation capability. Working with this algorithm and interpret its results is very straightforward. The aims of this study were to select the best set of input properties influencing SQIs using Pearson correlation coefficient and then model the effect of the input properties by decision tree and multiple linear regression.

Materials and Methods: In this study, the Pearson correlation coefficient was used for selecting effective soil properties influencing SQIs and these indices were modeled and predicted by the decision tree algorithm with selected input properties. For this purpose, 104 soil samples were collected from the soil surface (0-15 cm depth) of four land uses including a garden with 20 year-old walnut trees, pasture, agriculture and a mountain almond in a semi-arid area in Iran (Rabor region, 29° 27′ N to 38° 54′ N and 56° 45′ E to 57° 16′ E). A multiple linear regression (MLR) model was constructed as the benchmark for the comparison of performances. Sensitivity analysis of decision tree model was performed with input variables using StatSoft method. The predictive capabilities of the proposed models were evaluated by the mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2) between measured and predicted SQIs values.

Results and Discussion: The soil properties including porosity, bulk density, clay and sand content for air capacity, porosity and sand, clay and silt content for relative field capacity, and bulk density, electrical conductivity, porosity, and sand, clay and silt content for plant available water were selected as important input parameters. In addition, the values of r^2 for the decision tree model for air capacity, relative field capacity and plant available water were 0.95, 0.84 and 0.85, respectively, while the r^2 values for multiple linear regression for AC, RFC and PAWC were 0.63, 0.62 and 0.61, respectively. According to the evaluation indices, it appears that the conventional regression model was poor in predicting SQIs. Therefore, conventional regression techniques (i.e., multiple-linear regression) may not be reliable for predicting the SQIs. The results of sensitivity analysis for decision tree model showed that porosity and bulk density for air capacity, porosity for relative field capacity and bulk density for plant available water had the greatest influence.

Conclusion: This research work provided a basis for predicting soil physical quality indicators and identifying important parameters impacting these indicators in agricultural soils, grassland and forests in semi-arid regions which can be generalized to other areas. Further studies are needed to assess the effects of selected input variables under different conditions.

Keywords: Air capacity, Available water, Modeling, Relative field capacity, Soil management

1- Assistant Professor Department of Soil Sciences, Univesity of Jiroft, Jiroft, Iran
(*- Corresponding Author Email: h.shekofteh@vru.ac.ir)
2 and 3- Ph.D. and M.Sc. Students, Department of Soil Sciences, Vali-e-Asr Univesity of Rafsanjan, Rafsanjan, Iran