

مقایسه روش‌های k نزدیک‌ترین همسایگی و شبکه عصبی مصنوعی برای پهنه‌بندی رقومی شوری خاک در منطقه چاه افضل اردکان

شمس‌الله ایوبی^{۱*}، روح‌اله تقی‌زاده^۲، زینب نمازی^۱، علی ذوالفقاری^۳، فاطمه روستایی صدرآبادی^۲

(تاریخ دریافت: ۱۳۹۳/۹/۴؛ تاریخ پذیرش: ۱۳۹۴/۱۰/۳۰)

چکیده

استفاده از داده‌های کمکی رقومی و ارتباط آنها با داده‌های مشاهداتی صحرایی از طریق روش‌های کامپیوتری که اصطلاحاً نقشه‌برداری رقومی خاک خوانده می‌شود، نسبت به روش‌های سنتی نقشه‌برداری خاک قابل‌اعتمادتر و کم هزینه‌تر است. بنابراین در مطالعه حاضر، از مدل‌های شبکه عصبی مصنوعی و k نزدیک‌ترین همسایگی جهت پیش‌بینی مکانی شوری خاک در منطقه‌ای خشک به وسعت ۷۰۰ کیلومتر مربع در شمال شهرستان اردکان استفاده گردید. در این منطقه براساس روش شبکه‌بندی منظم ۱۸۰ نمونه خاک مشخص شده و سپس نمونه‌برداری و خصوصیات فیزیکی-شیمیایی خاک اندازه‌گیری شدند. متغیرهای محیطی استفاده شده در این مطالعه شامل پارامترهای استخراج شده از مدل رقومی ارتفاع، داده‌های تصویر ETM⁺ ماهواره لندست و هدایت الکتریکی ظاهری (اندازه‌گیری شده توسط دستگاه هدایت‌گر الکترومغناطیس) بودند. نتایج این تحقیق نشان داد که مدل k نزدیک‌ترین همسایگی دارای دقت بیشتری نسبت به شبکه عصبی مصنوعی به‌منظور پیش‌بینی شوری خاک است. به‌طوری‌که این مدل به‌خوبی توانسته ارتباط قوی بین داده‌های شوری خاک و اطلاعات محیطی برقرار کند. مجموع ریشه مربعات خطا و ضریب تبیین مدل k نزدیک‌ترین همسایگی به ترتیب ۱۸/۷۳ و ۰/۹۲ است. نتایج نشان داد که برای پیش‌بینی شوری خاک، هدایت الکتریکی ظاهری، شاخص‌های سنجش از دور و شاخص خیسی مهم‌ترین پارامترها بودند.

واژه‌های کلیدی: شوری خاک، داده‌های کمکی، نقشه‌برداری رقومی خاک

۱. گروه خاک‌شناسی، دانشکده کشاورزی، دانشگاه صنعتی اصفهان

۲. گروه مدیریت مناطق خشک، دانشکده کشاورزی و منابع طبیعی دانشگاه اردکان

۳. گروه بیابان‌زدایی، دانشکده کویرشناسی، دانشگاه سمنان

*: مسئول مکاتبات: پست الکترونیکی: ayoubi@cc.iut.ac.ir

مقدمه

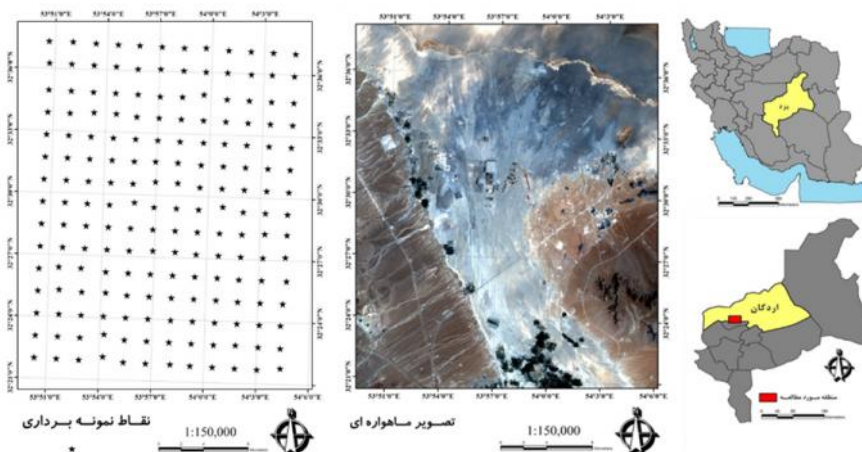
شوری زایی یکی از دلایل اصلی تخریب خاک در مناطق خشک و نیمه خشک در جهان می باشد (۲۰). کاهش رشد گیاهان و تولیدات گیاهی و همچنین افزایش فرسایش خاک از جمله پیامدهای نامطلوب این پدیده به شمار می رود (۹). به منظور مدیریت این پدیده، بررسی و پایش تغییرات زمانی و مکانی آن اجتناب ناپذیر است. روش های مرسوم یا سنتی شناسایی و تهیه نقشه خاک به طور عمده بر پایه اطلاعات خاک رخ شاهد استوار هستند. تعمیم نتایج حاصل از خاک رخی های شاهد به کل واحد نقشه بدون در نظر گرفتن تغییرات زمانی و مکانی خاک، نیازمندی به حفر تعداد زیاد خاک رخ، مطالعات آزمایشگاهی گسترده و زمانبر بودن، شاید از مهم ترین محدودیت های جدی نقشه برداری مرسوم خاک باشند (۱). چالش های موجود در رابطه با درک حجم وسیعی از داده ها، منجر به توسعه روش های جدید در تجزیه و تحلیل این داده ها شده است. مک براتنی و همکاران (۱۳)، با مروری بر مدل های مختلف تخمین و برآورد مکانی خاک و با توجه به نکات کلیدی مذکور اقدام به ارائه یک چارچوب تعمیم یافته برای تخمین و برآورد مکانی خاک نمودند. ایشان مدل پنج عاملی ینی را به یک مدل هفت عاملی از طریق کمی سازی و بهره مندی از مدل های رقومی ارتفاع، سنجش از دور، حس گرهای سنجنده خاک و دیگر تکنیک های پهنه بندی، بسط و تعمیم دادند (۱۶).

انتخاب متغیرهای محیطی وابستگی زیادی به نوع ویژگی خاک دارند (۱۲). به عنوان مثال، برای ارزیابی شوری، مقادیر هدایت الکتریکی ظاهری خاک مفید است (۲۲، ۲۳). ابزارهای مبتنی بر القای الکترومغناطیس که قابلیت هدایت الکتریکی ظاهری یک توده از خاک را اندازه گیری می کنند، در تعیین الگوی تغییرات مکانی درون مزرعه ای برخی خصوصیات فیزیکی- شیمیایی خاک نظیر شوری، درصد رس، درصد رطوبت و عمق خاک لایه سطحی آن هم به صورت همزمان با لحظه پیمایش مفید هستند. به عنوان مثال، تقی زاده و همکاران (۱۹) از خاصیت القای الکترومغناطیس جهت پایش شوری

خاک در منطقه اردکان یزد استفاده کردند. نتایج ایشان نشان داد که مقادیر هدایت گر الکترومغناطیس به خوبی می تواند تغییرات شوری خاک را پیش بینی کند. البته این ارتباط در خاک های سطحی بهتر و در اعماق خاک کم گزارش شده است.

یکی از انواع مدل های مکانی جهت ارتباط دادن داده های خاک و متغیرهای محیطی شبکه های عصبی مصنوعی است. شبکه های عصبی مصنوعی یکی از روش های هوشمند پردازش داده ها هستند که در ابتدا به وسیله پاچپسکی و همکاران (۱۸) و تاماری و همکاران (۲۱) در بحث توابع انتقالی خاک (Apparent Electrical Conductivity) مورد استفاده قرار گرفته اند. در مورد نقشه برداری رقومی خاک می توان به پژوهش های انجام گرفته توسط مک کی و همکاران (۱۳)، زو و همکاران (۲۵) و منزس و همکاران (۱۲) اشاره نمود. یکی دیگر از روش های غیر پارامتریک که اخیراً در بسیاری از علوم از جمله کشاورزی (۶) و هیدرولوژی (۲۴) کاربرد چشمگیری یافته است که مبتنی بر الگوشناسی و استفاده از اصل تشابه و نزدیکی داده ها است. چنین رویکردی در تخمین ویژگی های مورد نظر، هنگامی که نوع رابطه بین داده های ورودی و خروجی روشن نباشد از سودمندی بیشتری برخوردار است (۱۱). روش k نزدیک ترین همسایگی (k-Nearest Neighbor) یکی از تکنیک های مبتنی بر چنین رویکردی است که در تخمین برخی ویژگی های فیزیکی و شیمیایی خاک مورد استفاده قرار گرفته است (۲۴).

منس و همکاران (۱۶) توانایی روش k نزدیک ترین همسایگی در تخمین توابع هیدرولیکی خاک را در مقیاس کل خاک های ایالت متحده بهتر از توابع پارامتریک ارزیابی نمودند (۴). پژوهش حاضر ضمن بررسی سودمندی کاربرد تکنیک k نزدیک ترین همسایگی در برآورد هدایت الکتریکی خاک با استفاده از انواع داده های کمکی (تصاویر ماهواره، مدل رقومی ارتفاع و هدایت الکتریکی ظاهری) در منطقه چاه افضل اردکان، دقت تخمین این کمیت با دو روش k نزدیک ترین همسایگی و روش شبکه عصبی مصنوعی مورد ارزیابی قرار خواهد گرفت.



شکل ۱. موقعیت منطقه مورد مطالعه و محل نمونه‌ها بر روی تصویر ماهواره‌ای ETM+ در استان یزد

مواد و روش ها

منطقه مورد مطالعه

منطقه مورد مطالعه در ایران مرکزی واقع و وسعتی برابر با ۷۰۰۰۰ هکتار از اراضی دشت یزد- اردکان را شامل می‌شود. این منطقه بین طول‌های جغرافیایی $41^{\circ} 53'$ تا $41^{\circ} 57'$ شرقی و عرض‌های جغرافیایی $23^{\circ} 32'$ تا $23^{\circ} 41'$ شمالی قرار دارد و دارای مشکلات شوری منابع آب و خاک است. منطقه مورد مطالعه از جنوب به شهر اردکان و از شرق به رشته کوه هرشت و از شمال به چاه افضل و از غرب به رشته کوه‌های شیرکوه منتهی می‌شود. میانگین تبخیر و تعرق مرجع، دما و بارندگی سالانه منطقه مورد مطالعه براساس ایستگاه سینوپتیک اردکان در بازه زمانی ۱۰ ساله، به ترتیب برابر ۳۴۸۳ میلی‌متر، $18/5$ درجه سلسیوس و ۷۵ میلی‌متر است. در اراضی جنوبی منطقه، کاربری کشاورزی پسته، روناس، گندم و در حاشیه‌ها، اراضی بایر و مرتع ضعیف وجود دارد. پوشش گیاهی طبیعی منطقه گون، آترپلکس، پسته کوهی، ارژن و گز و پوشش گیاهی مصنوعی تاغ است. شکل (۱) موقعیت منطقه مورد مطالعه را بر روی تصویر ماهواره‌ای ETM+ در استان یزد را نشان می‌دهد.

نمونه برداری

در منطقه مورد مطالعه تعداد ۱۸۰ نمونه خاک، از عمق صفر تا

۲۰ سانتی متری و در یک شبکه منظم دو کیلومتری برداشت شد (شکل ۱). جهت حفظ دقت در بزرگ‌نمایی داده‌ها و هم مقیاس بودن و پرهیز از تغییرات مکانی کوچک از هر نقطه به فاصله ۱۰ متر از چهار طرف و یک نقطه در مرکز نمونه برداری انجام و یک نمونه مرکب تهیه شد. از این تعداد ۵۴ عدد خاک (۳۰ درصد نمونه‌ها) به صورت تصادفی انتخاب و به عنوان خاک‌های مورد آزمون (هدف) جهت تخمین هدایت الکتریکی خاک در نظر گرفته شدند؛ و باقی خاک‌ها به عنوان خاک‌های مرجع (آموزش) برای آموزش مدل‌های استفاده شدند. کلیه نمونه‌ها پس از هوا خشک شدن و کوبیده شدن از الک دو میلی‌متری عبور داده شدند و هدایت الکتریکی آنها در عصاره گل اشباع تعیین شد.

متغیرهای کمکی مورد استفاده در برآورد هدایت الکتریکی

در تحقیق حاضر برای برآورد هدایت الکتریکی خاک از برخی متغیرهای کمکی که از تصاویر ماهواره‌ای قابل استخراج بودند، استفاده شد. همچنین از مدل رقومی سازمان زمین‌شناسی آمریکا مطابق با استاندارد پروژه جهانی خاک (قدرت وضوح ۹۰ متر) برای استخراج ویژگی‌های زمین‌نما از قبیل شیب، ارتفاع، ارتفاع بالای شبکه زهکشی، مساحت حوزه‌های اصلاح‌شده، موقعیت میانی شیب، عمق دره، شاخص خیسی، شاخص همواری دره

دره با درجه تفکیک بالا، شاخص خیسی، مساحت حوزه‌های اصلاح شده و ارتفاع بودند. متغیرهای فوق برای تهیه نقشه‌های رقومی استفاده شد. در روش دوم از آنالیز مؤلفه‌های اصلی یا (PCA) *Principale Component Analysis* برای انتخاب متغیرهای ورودی استفاده شد. آنالیز مؤلفه‌های اصلی این امکان را در اختیار قرار می‌دهد که از چندین متغیری که داری همبستگی درونی هستند یک یا چند متغیر را که همبستگی کمتری با یکدیگر دارند، در برآورد متغیر هدف (به‌عنوان مثال هدایت الکتریکی خاک) مورد استفاده قرار داد؛ بنابراین در این مطالعه، با استفاده از این روش تعداد چهار متغیر از ۲۲ متغیر کمکی در برآورد هدایت الکتریکی خاک استفاده شد. متغیرهای مورد استفاده براساس آنالیز مؤلفه‌های اصلی عبارتند از هدایت الکتریکی ظاهری عمودی، مساحت حوزه‌های اصلاح‌شده، ارتفاع و شاخص شوری بودند. همچنین، با استفاده از تمامی ۲۲ متغیر کمکی اقدام به پیش‌بینی متغیر هدف شد. دقت برآورد هدایت الکتریکی خاک با توجه متغیرهای ورودی به‌دست آمده از همبستگی و تجزیه مؤلفه‌های اصلی با زمانی که تمامی متغیرها برای برآورد هدایت الکتریکی خاک استفاده شدند، مقایسه شد.

مدل‌های مورد استفاده در برآورد هدایت هیدرولیکی خاک شامل موارد ذیل می‌باشند:

مدل شبکه عصبی مصنوعی

در این تحقیق از الگوریتم پرسپترون چند لایه با سه لایه جهت برآورد داده‌های هدف (داده‌های آزمون) استفاده شد. این نوع الگوریتم از یک لایه ورودی به‌منظور اعمال ورودی‌های مسئله جهت آموزش شبکه با داده‌های متغیرهای کمکی، یک لایه پنهان با هفت نورون و تابع عملگر سیگموئیدی و یک لایه خروجی با تابع عملگر خطی تشکیل شده است و قادر خواهد بود که برآوردی از هدایت الکتریکی خاک‌های مورد آزمون را باتوجه به بانک داده‌های ورودی (مرجع) ارائه نمایند (۸). در

با درجه تفکیک بالا یا (Multiresolution index of Valley Bottom Flatness) *MrVBF*، شاخص بالای پشته با درجه تفکیک یال یا (Multiresolution ridgetop flatness) *MrRTF* (index ETM+) و شیب حوزه استفاده گردید (۱۰). علاوه بر ویژگی‌های سرزمین، از اطلاعات شش باند تصویر *ETM+* (تابستان ۱۳۹۲) به‌همراه یکسری شاخص‌های ماهواره‌ای از قبیل شاخص گیاهی نرمال شده یا (Normalized difference) *NDVI* (vegetation index) شاخص روشنایی یا (*Brightness*) *BI* (index) و شاخص شوری یا (*Salinity index*) *SI* بهره گرفته شد. علاوه بر این از مقادیر هدایت الکتریکی ظاهری نیز بهره گرفته شد. در مجموعه ۲۲ متغیر کمکی برای برآورد هدایت الکتریکی خاک مورد استفاده قرار گرفت.

$$NDVI = \frac{B_4 - B_3}{B_4 + B_3} \quad [1]$$

$$BI = \sqrt{((B_3)^2 + (B_4)^2)} \quad [2]$$

$$SI = \frac{B_3 - B_4}{B_3 + B_4} \quad [3]$$

انتخاب بهترین متغیرهای کمکی در برآورد هدایت الکتریکی خاک

افزایش تعداد متغیرهای کمکی در برآورد هدایت الکتریکی خاک سبب می‌شود که مدت‌زمان اجرای برنامه‌های کامپیوتری در تهیه نقشه‌های رقومی افزایش یابد. لذا بهتر است که با استفاده از تعداد متغیر کمتر به نقشه‌هایی با دقت مناسب رسید. در این مطالعه به‌منظور کاهش تعداد متغیرهای کمکی مورد استفاده در برآورد هدایت الکتریکی خاک از دو روش (همبستگی و تجزیه به مؤلفه‌های اصلی) استفاده شد. در روش اول متغیرهایی که ارتباط معنی‌داری را با هدایت الکتریکی خاک در سطح احتمال یک درصد آماری داشتند به‌عنوان متغیرهای ورودی در برآورد هدایت الکتریکی خاک استفاده شدند. با توجه به روش مذکور متغیرهای همبستگی معنی‌داری را با هدایت الکتریکی خاک داشتند عبارتند از هدایت الکتریکی ظاهری عمودی و افق، شاخص همواری

جدول ۱. تغییرات هدایت الکتریکی خاک‌های منطقه مورد مطالعه

حداکثر	حداقل	میانگین	انحراف معیار	میان	ضریب تغییرات (%)
۱۳۹	۰/۴۹	۴۹/۰۱	۴۳/۴۹	۳۷/۵۷	۸۹

از معادلات زیر به دست آمدند.

$$r = \frac{n \left(\sum_{i=1}^n (P_i O_i) \right) - \left(\sum_{i=1}^n P_i \right) \left(\sum_{i=1}^n O_i \right)}{\sqrt{\left[n \sum_{i=1}^n (P_i)^2 - \left(\sum_{i=1}^n P_i \right)^2 \right] \left[n \sum_{i=1}^n (O_i)^2 - \left(\sum_{i=1}^n O_i \right)^2 \right]}} \quad [4]$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}} \quad [5]$$

$$MAE = 1/n \sum_{i=1}^n |P_i - O_i| \quad [6]$$

$$ME = \max |P_i - O_i| \quad [7]$$

$$EF = \frac{\sum_{i=1}^n (O_i - \bar{O})^2 - \sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \times 100 \quad [8]$$

P_i و O_i به ترتیب مقادیر هدایت الکتریکی پیش‌بینی شده، \bar{O} میانگین مقادیر مشاهده‌ای هدایت الکتریکی خاک و n تعداد نمونه‌های مورد آزمون است.

نتایج و بحث

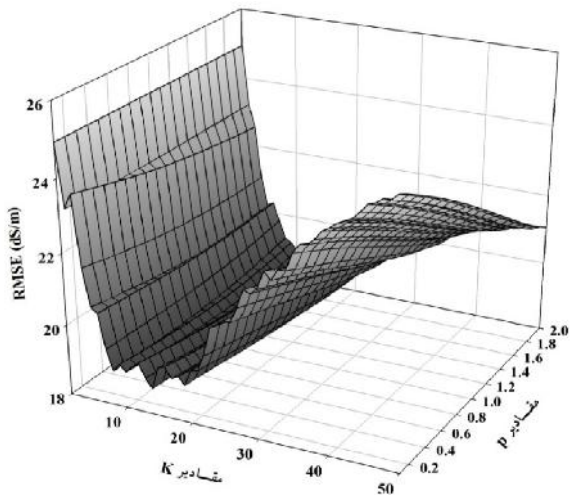
هدایت الکتریکی عصاره اشباع خاک‌های مورد مطالعه دارای حداقل ۰/۴۹ و حداکثر ۱۳۹ با میانگین ۴۹/۰۱ دسی‌زیمنس بر متر بود که بیانگر وضعیت خیلی نامطلوب منطقه مورد مطالعه از نظر توزیع شوری خاک است. انحراف معیار هدایت الکتریکی عصاره اشباع ۴۳/۴۹ دسی‌زیمنس بر متر است. ضریب تغییرات هدایت الکتریکی خاک در منطقه مورد مطالعه برابر با ۸۹ درصد بود که نشان دهنده تغییرات زیاد هدایت الکتریکی خاک در منطقه مورد مطالعه است (جدول ۱).

این مطالعه مدل شبکه عصبی با توجه به تعداد متفاوت متغیرهای کمکی اجرا شد (۶ ورودی، ۴ ورودی، ۲۲ ورودی).

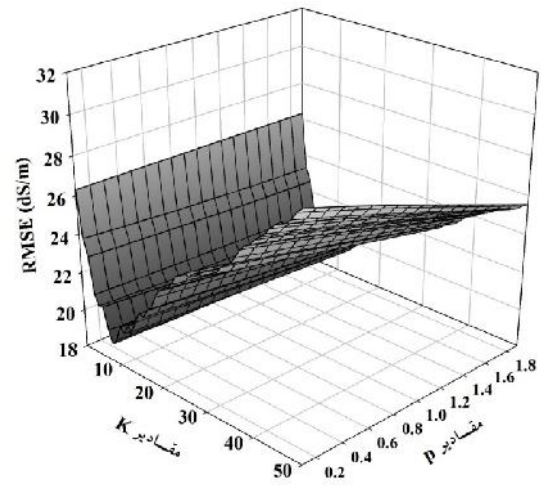
مدل k نزدیک‌ترین همسایگی (k-NN)

روش k نزدیک‌ترین همسایگی دارای یکی از ساده‌ترین و کارآمدترین الگوریتم‌های آموزش جهت پیش‌بینی داده‌ها است (۴). در این تکنیک، تعداد بهینه‌ای از نمونه‌های موجود در یک مجموعه که واجد شبیه‌ترین ویژگی‌ها به نمونه هدف باشند (k عدد) در نظر گرفته می‌شود و سپس با توجه به فواصل اقلیدسی نمونه هدف و هر یک از نمونه‌های بانک مرجع، وزندهی عناصر آموزش‌دهنده انجام می‌پذیرد. در یک بانک داده مرجع خاک، K عدد از نزدیک‌ترین (مشابه‌ترین) خاک‌ها به خاک مورد آزمون انتخاب می‌شوند و این خاک‌ها با توجه به میزان مشابهت به خاک مورد آزمون وزندهی می‌شوند. به این ترتیب که در ابتدا فواصل اقلیدسی بین خاک مورد آزمون و تمامی خاک‌های مرجع محاسبه شده، سپس تعداد k عدد از نزدیک‌ترین خاک‌ها براساس فاصله به دست آمده، وزندهی می‌شوند. در انتها با توجه وزن هر یک از خاک‌ها در مجموعه‌ای با k عدد خاک (نزدیک‌ترین همسایه)، برآوردی از داده‌های هدف که مورد آزمون واقع شده‌اند، صورت می‌پذیرد (۱۵، ۲۶).

به منظور مقایسه عملکرد دو روش مدل شبکه عصبی مصنوعی و روش k نزدیک‌ترین همسایگی آماره‌های ضریب همبستگی r (Correlation coefficient) ریشه میانگین مربعات خطا یا RMSE (Root mean square error)، میانگین قدرمطلق خطا یا MAE (Mean Absolute Error)، بیشترین خطا یا MaxE (Maximum Error)، میانگین باقی‌مانده یا ME (Mean Error) و کارایی مدل یا EF (Efficiency of Model) با استفاده



شکل ۳. رابطه مجذور میانگین مربعات خطا، با تغییرات همزمان K و p در یک نمودار سه بعدی با توجه به انتخاب متغیرها براساس تجزیه مؤلفه‌های اصلی



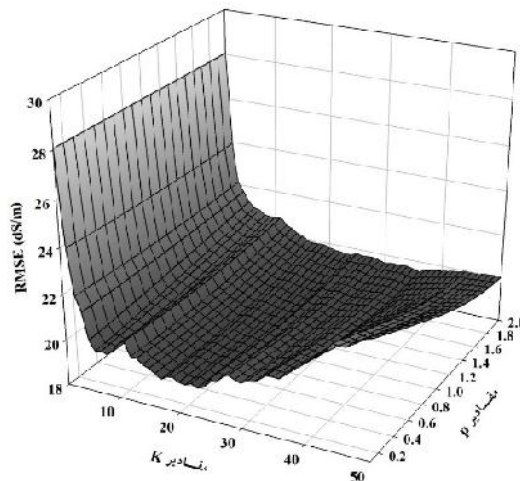
شکل ۲. رابطه مجذور میانگین مربعات خطا، با تغییرات همزمان K و p در یک نمودار سه بعدی در هنگام استفاده از تمامی ۲۲ متغیر وردی در پیش‌بینی هدایت الکتریکی خاک

نتایج نشان می‌دهد که p تأثیر زیادی دقت برآورد هدایت الکتریکی خاک ندارد. نمس و همکاران (۱۵) نتیجه گرفتند که p می‌تواند اثر منفی کوچک بودن بانک داده مرجع را در عملکرد الگوریتم بهبود بخشد. در حقیقت مقادیر غیر از ۱ برای p، با تأثیر بر وزندهی نقاط سبب تعدیل اثرات منفی کوچک بودن بانک داده مرجع می‌شود. به طوری که در مقادیر کوچک P ($P < 1$)، اثر وزن نقاط نزدیک به داده هدف کاهش و با افزایش P به بزرگتر از ۱، این اثر افزایش می‌یابد.

شکل (۳) نمودار سه بعدی تغییرات مجذور میانگین مربعات خطا را با تغییرات همزمان K و p در هنگامی که چهار متغیر وردی حاصل از تجزیه مؤلفه‌های اصلی برای پیش‌بینی هدایت الکتریکی خاک مورد استفاده قرار گرفتند را نشان می‌دهد. نتایج نشان می‌دهند که کمترین خطای برآورد هدایت الکتریکی خاک در $K=10$ و $p=1/2$ به دست آمد. مشابه نمودار (۲) نتایج این نمودار نیز نشان می‌دهد که تغییرات p تأثیر معنی داری را در مقدار خطای برآورد هدایت الکتریکی ندارند. مقایسه خطای برآورد هدایت الکتریکی با استفاده از وردی‌های استخراج شده از تجزیه مؤلفه‌های اصلی و روشی که از تمامی متغیرها برای پیش‌بینی مقادیر هدایت الکتریکی‌ها استفاده شده

بررسی روش K نزدیک‌ترین همسایگی در پیش‌بینی هدایت الکتریکی

بهینه‌سازی توأم K و p با استفاده از روش ارزیابی تقاطعی به گونه‌ای انجام شد که به ازای هر مقدار p از ۰/۱ تا دو (با فواصل ۰/۱) و K های مختلف از یک تا ۵۰ در نظر گرفته شد و در ترکیبات مختلف p و K مقادیر مجذور میانگین مربعات خطا محاسبه شد. در این مطالعه بهینه‌سازی K و P در سه حالت مختلف که متغیرهای وردی متفاوتی برای پیش‌بینی هدایت الکتریکی اشباع خاک مورد استفاده قرار گرفتند، انجام شد. شکل (۲) نمودار سه بعدی تغییرات مجذور میانگین مربعات خطا را با تغییرات همزمان K و p را برای هنگامی که تمامی ۲۲ متغیر وردی برای پیش‌بینی هدایت الکتریکی خاک مورد استفاده قرار گرفتند را نشان می‌دهد. نتایج نشان داد که کمترین مقدار خطای برآورد هدایت الکتریکی خاک در $K=8$ وجود دارد. نتایج این نمودار همچنین نشان داد که تغییرات p سبب تغییرات جزئی در مقدار خطای برآورد هدایت الکتریکی خاک می‌شوند. به عنوان مثال مقدار مجذور میانگین مربعات خطا در $K=8$ و $p=0/1$ برابر با $18/93$ دسی‌زیمنس بر متر و مقدار آن در $p=2$ برابر با $18/83$ دسی‌زیمنس بر متر است به عبارت دیگر این



شکل ۴. رابطه مجذور میانگین مربعات خطا، با تغییرات همزمان K و p در یک نمودار سه بعدی در صورت استفاده از متغیرهایی که دارای همبستگی زیادی با هدایت الکتریکی هستند

یک نرون که همان شوری خاک است در نظر گرفته شد. تعداد نرون ها و لایه های مخفی بستگی به پیچیدگی مسئله مورد نظر دارد. در این جا یک لایه مخفی برای ایجاد شبکه پرسپترون مورد استفاده قرار گرفت. تعداد نرون های لایه مخفی و همچنین تعداد اپوک در فرآیند آموزش به صورت سعی و خطا به دست آمدند. در این مطالعه تعداد نرون لایه مخفی دو تا ۱۰ و تعداد اپوک ۱۰۰ تا ۱۰۰۰ متغیر در نظر گرفته شد. برای تعیین بهترین ترکیب از شاخص میانگین مربعات خطا استفاده گردید. نتایج نشان داد که تعداد پنج نرون در لایه مخفی و تعداد اپوک ۱۰۰۰ دارای بهترین ترکیب در پیش بینی شوری است. نتایج نشان داد که استفاده از تصاویر ماهواره ای کارایی شبکه عصبی را در برآورد شوری خاک بهبود می بخشد. محققین مختلفی نیز کارایی تصاویر ماهواره را در تهیه نقشه شوری سطح خاک تأیید کرده اند (۳، ۵، ۱۷).

ارزیابی کارایی مدل ها در برآورد شوری خاک

جدول (۲) مقادیر آماره های مورد استفاده برای ارزیابی روش های K نزدیک ترین همسایگی و شبکه عصبی مصنوعی را نشان می دهد. برای تعیین آماره ها در ابتدا بهترین مقادیر P و K تعیین شد سپس این مقادیر در برنامه ثابت شدند و سپس مقادیر

بود نشان داد که نتایج این دو روش بسیار به یکدیگر نزدیک می باشند. روند تغییرات ریشه میانگین مربعات خطا این دو روش (شکل های ۲ و ۳) نیز این مطلب را نشان می دهد.

شکل (۴) نمودار سه بعدی تغییرات ریشه میانگین مربعات خطا را با تغییرات همزمان p و K را با توجه به متغیرهای وردی حاصل از بیشترین همبستگی با هدایت الکتریکی را نشان می دهد. روند کلی این نمودار اختلافاتی را با نمودارهای ۲ و ۳ نشان می دهد. نتایج این نمودار نشان داد که کمترین مقدار ریشه میانگین مربعات خطا در $p=1$ و $K=20$ به دست آمد. نتایج نشان می دهد که نسبت تغییرات ریشه میانگین مربعات خطا با افزایش K در مقایسه با روش تجزیه مؤلفه های اصلی و روش استفاده از تمامی متغیرها ملایم تر است. به عنوان مثال افزایش K از ۲۰ به ۵۰ سبب افزایش سه واحدی مجذور میانگین مربعات خطا در این روش شده است در حالی که افزایش K از ۲۰ به ۵۰ در روش استفاده از تمامی متغیرها سبب افزایش پنج واحدی مجذور میانگین مربعات خطا شده است.

بررسی روش مدل شبکه عصبی مصنوعی در پیش بینی EC

لایه های ورودی شبکه با توجه به روش های مختلف بین ۴ تا ۲۲ ورودی متغیر بود؛ اما در تمامی موارد لایه خروجی شامل

جدول ۲. آماره‌های مقایسه عملکرد در روش‌های k نزدیک‌ترین همسایگی و شبکه عصبی مصنوعی

EF	MaxE	MAE	RMSE	r	متغیرهای ورودی	روش
۷۹	۴۰/۵۸	۱۵/۱۸	۱۸/۷۳	۰/۹۲	براساس PCA	k نزدیک‌ترین همسایگی
۷۸	۵۱/۱۵	۱۴/۳۷	۱۹/۳۰	۰/۹۰	براساس ضریب همبستگی	
۷۸	۴۶/۲۵	۱۵/۸۱	۱۹/۳۲	۰/۹۲	براساس تمام متغیرها	
۷۹	۷۷/۶۹	۱۳/۱۷	۱۸/۶۶	۰/۸۹	براساس PCA	شبکه عصبی مصنوعی
۷۹	۵۴/۶۷	۱۴/۷۰	۱۸/۷۴	۰/۹۰	براساس ضریب همبستگی	
۷۸	۷۱/۱۵	۱۵/۴۶	۱۹/۴۳	۰/۸۹	براساس تمام متغیرها	

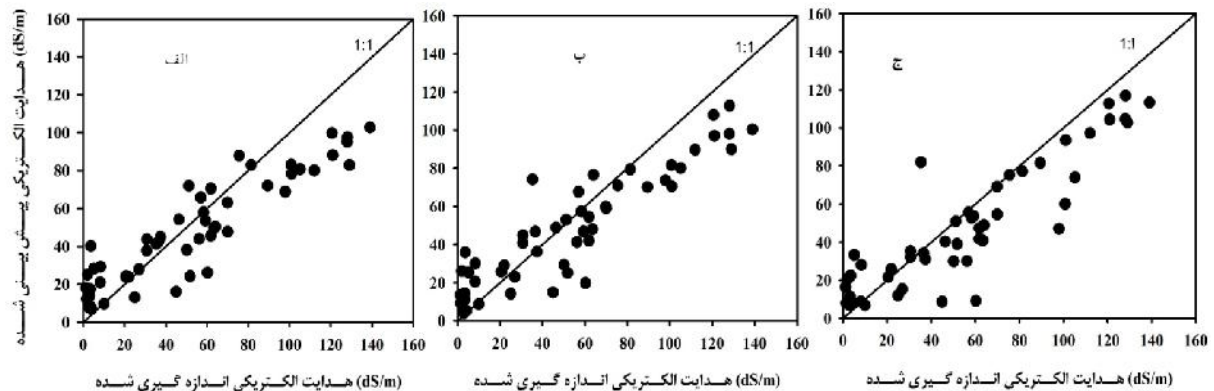
ریشه میانگین مربعات خطا (RMSE)، میانگین قدر مطلق خطا (MAE)، بیشترین خطا (MaxE)، میانگین خطا (ME) بر حسب و کارایی مدل (EF) بر حسب درصد است.

متغیرهای ورودی هزینه بر نخواهد بود لذا بسیاری از محققین ترجیح می‌دهند که از متغیرهای ورودی زیادی برای تعیین این نقشه‌های استفاده نمایند. این عمل باعث می‌شود که مدت‌زمان اجرای برنامه‌های بخصوص برای مناطق مطالعاتی با وسعت زیاد به‌طور تصاعدی افزایش یابد. درحالی‌که نتایج این مطالعه نشان داد که می‌توان تعداد متغیرهای ورودی مورد استفاده در تهیه نقشه‌های رقومی را کاهش داد، بدون این‌که سبب افزایش خطا در تولید نقشه‌ها شود.

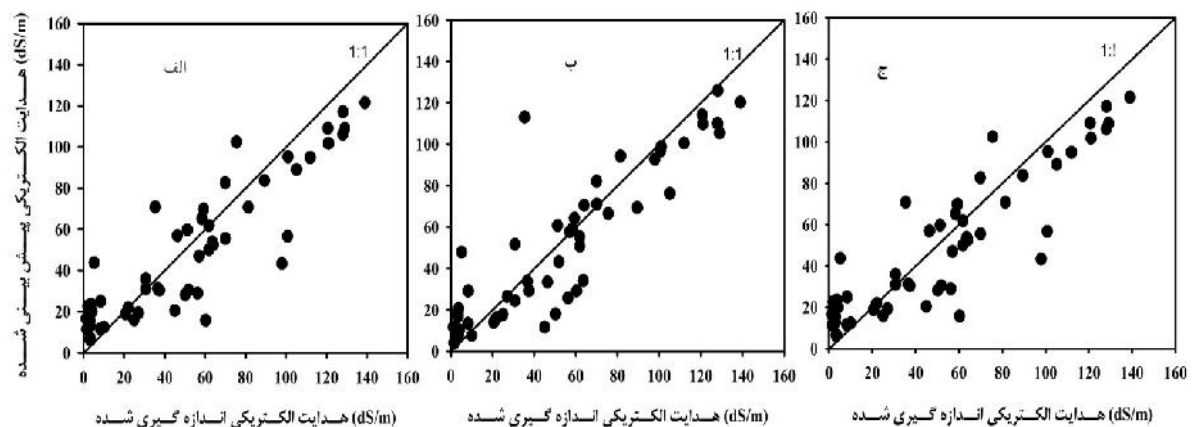
مقایسه مقادیر برآوردی هدایت الکتریکی خاک‌های مورد آزمون در روش k نزدیک‌ترین همسایگی با مقادیر مشاهده‌ای این کمیت نسبت به خط ۱:۱ توانایی و دقت بسیار بالای این روش را در تخمین هدایت الکتریکی خاک از روی ویژگی‌های مذکور را نشان می‌دهد (شکل ۵). در این شکل مقادیر هدایت الکتریکی پیش‌بینی شده در مقابل مقادیر مشاهده‌ای این کمیت در روش k نزدیک‌ترین همسایگی ترسیم شده است. نتایج این نمودار نشان می‌دهد که استفاده از روش تجزیه مؤلفه‌های اصلی در انتخاب پارامترهای ورودی سبب می‌شود که مقادیر پیش‌بینی و اندازه‌گیری شده هدایت الکتریکی به خط ۱:۱ نزدیک‌تر شده و از پراکندگی داده‌ها کاسته می‌شود. به‌عبارت دیگر می‌توان بیان داشت که اگرچه دقت نتایج پیش‌بینی هدایت الکتریکی خاک با استفاده از ورودی‌های حاصل از تجزیه مؤلفه‌های اصلی و همبستگی با یکدیگر مشابه می‌باشند اما دقت برآورد با

آماره‌ها با توجه به بهترین مقادیر p و K تعیین شدند. نتایج نشان داد که در کمترین مقدار مجذور میانگین مربعات خطا (۱۸/۷۳ دسی‌زیمنس بر متر) در برآورد هدایت الکتریکی خاک با استفاده از K نزدیک‌ترین همسایگی زمانی به‌دست آمد که از متغیرهای ورودی انتخاب‌شده براساس تجزیه مؤلفه‌های اصلی استفاده شده بود. در روش k نزدیک‌ترین همسایگی در بدترین حالت، هدایت الکتریکی را با ۵۱/۱۵ دسی‌زیمنس بر متر خطا پیش‌بینی نمود. این شرایط در روش مدل شبکه عصبی مصنوعی با تخمین ضعیف‌تر و میزان خطای ۷۷/۶۹ دسی‌زیمنس بر متر همراه بود.

مقادیر زیاد ضریب همبستگی در هر دو روش k نزدیک‌ترین همسایگی و مدل شبکه عصبی مصنوعی نشان داد که این دو روش در برآورد هدایت الکتریکی خاک‌های مورد آزمون از دقت بالا و یکسانی برخوردارند مقایسه مقادیر کارایی مدل نشان داد که هر دو روش دارای کارایی بالایی (EF=۰/۷۹) در برآورد شوری خاک از روی متغیرهای محیطی می‌باشند و هیچ اختلافی از این لحاظ بین دو روش مشاهده نمی‌شود (جدول ۲). همچنین مقایسه نتایج نشان داد که کمترین ریشه میانگین مربعات خطا با هر دو روش زمانی حاصل می‌شود که از متغیرهای ورودی مستخرج شده از روش تجزیه مؤلفه‌های اصلی در برآورد هدایت الکتریکی خاک استفاده شده باشد. در این مطالعه و به‌طور کلی در تولید نقشه‌های رقومی استخراج



شکل ۵. رابطه بین مقادیر پیش بینی و مشاهده ای هدایت الکتریکی با استفاده از روش k نزدیک ترین همسایگی در بانک داده هدف الف) با استفاده از تمامی متغیرها، ب) با استفاده از روش تجزیه مؤلفه های اصلی و ج) با استفاده از متغیرهای معنی دار در همبستگی



شکل ۶. رابطه بین مقادیر پیش بینی و مشاهده ای هدایت الکتریکی با استفاده از روش شبکه عصبی مصنوعی در بانک داده هدف الف) با استفاده از تمامی متغیرها، ب) با استفاده از روش تجزیه مؤلفه های اصلی و ج) با استفاده از متغیرهای معنی دار در همبستگی

مصنوعی در برخی از خاک ها مقدار هدایت الکتریکی را با دقت بسیار بالایی پیش بینی می کند اما در پیش بینی هدایت الکتریکی برخی از خاک ها دارای خطای بسیار بالایی است. به عنوان مثال نتایج جدول ۲ نیز نشان می دهد که در بدترین حالت شبکه عصبی مقدار هدایت الکتریکی را با ۷۱/۱۵ دسی زیمنس بر متر خطا پیش بینی می کند در حالی که این مقدار برای روش k نزدیک ترین همسایگی برابر با ۵۱/۱۵ دسی زیمنس بر متر است. از آنجاکه الگوریتم روش شبکه عصبی مصنوعی از تمام داده های بانک مرجع و ارتباط بین آنها جهت آموزش استفاده می نماید، همگنی داده ها می تواند کمک قابل توجهی به بهبود نتایج

استفاده از ورودی های حاصل از همبستگی در برخی از نمونه بسیار خوب و در برخی دیگر دارای خطای زیاد است؛ که این عامل ممکن است که سبب ایجاد خطای زیادی در تهیه نقشه های رقومی شوری خاک شود. به عبارت دیگر این نتایج نشان می دهد که استفاده از ورودی های حاصل از تجزیه مؤلفه های اصلی سبب تولید دقیق ترین نقشه های رقومی خواهد شد.

شکل (۶) مقادیر پیش بینی شده و اندازه گیری شده هدایت الکتریکی خاک را با استفاده از روش شبکه عصبی مصنوعی نشان می دهد. نتایج این نمودار نشان می دهد که شبکه عصبی

رگرسیون چند متغیره خطی قادر به برآورد رطوبت خاک در مکش های مختلف می باشد. این محققین بیان کردند که روش k نزدیک ترین همسایگی انعطاف پذیری بالاتری نسبت به روش رگرسیون چند متغیره خطی دارد. به عنوان مثال می توان داده های جدیدی را به بانک داده مرجع اضافه کرد بدون این که لازم باشد معادله جدیدی برای برآورد خصوصیات خاک توسعه داده شود. ذوالفقاری و همکاران (۲۶) روش k نزدیک ترین همسایگی و شبکه عصبی مصنوعی را برای برآورد ظرفیت تبادل کاتیونی خاک (CEC) مورد استفاده قرار دادند. آنها نشان دادند که کارایی دو روش برای برآورد ظرفیت تبادل کاتیونی خاک مشابه می باشد. این محققین نشان دادند که روش k نزدیک ترین همسایگی برتری اندکی نسبت به روش شبکه عصبی مصنوعی در برآورد ظرفیت تبادل کاتیونی خاک دارد.

نتیجه گیری

این تحقیق به منظور مقایسه دو تکنیک مدل سازی نقشه برداری رقمی شامل شبکه عصبی مصنوعی و الگوریتم k نزدیک ترین همسایه جهت استفاده از داده های محیطی کمکی شامل داده های استخراج شده توپوگرافی از مدل رقمی ارتفاعی، داده های سنجش از دور و داده های هدایت الکتریکی ظاهری (ECa) برای پیش بینی هدایت الکتریکی خاک به عنوان شاخصی از شوری خاک استفاده شد. نتایج کلی این تحقیق نشان داد که در منطقه مورد مطالعه الگوریتم k نزدیک ترین همسایه دارای دقت و کارایی بیش تری برای پیش بینی شوری خاک نسبت به مدل شبکه عصبی مصنوعی هستند. لازم است تا کارایی دیگر تکنیک های مدل سازی نقشه برداری رقمی خاک نظیر درخت تصمیم گیری یا نرو فازی نیز در پیش بینی شوری خاک در منطقه مورد ارزیابی قرار گیرد تا کارایی این الگوریتم در مقایسه با آنها مشخص گردد.

برآوردها با این روش نماید این شرایط برای روش k نزدیک ترین همسایگی که از تعداد محدودتری از داده ها که کمترین فاصله را با داده های هدف دارند (k)، به مراتب منافع کمتری را به دنبال خواهد داشت. به عبارت دیگر در صورتی که همگنی کمی بین خاک های بانک داده مرجع وجود داشته باشد دقت برآورد شبکه عصبی مصنوعی کاهش خواهد یافت. در حالی که ناهمگنی در داده های مرجع تأثیر زیادی بر دقت برآورد هدایت الکتریکی با استفاده از روش k نزدیک ترین همسایگی نخواهد داشت. حقوردی و همکاران (۲) در تخمین توابع هیدرولیکی خاک با مقایسه اثر کیفیت و نوع داده ها بر عملکرد دو روش شبکه عصبی مصنوعی و k نزدیک ترین همسایگی گزارش کردند که حساسیت روش شبکه عصبی مصنوعی به ناهمگنی داده ها بیشتر از روش k نزدیک ترین همسایگی بوده به طوری که همگن نبودن داده ها در تحقیق مذکور باعث افزایش ۱۰۰ درصدی خطای برآوردها در این روش شد. مشابه روش k نزدیک ترین همسایگی روش شبکه عصبی نیز نشان داد که استفاده از تجزیه مؤلفه های اصلی برای جداسازی متغیرهای ورودی سبب بهبود برآورد هدایت الکتریکی خواهد شد (شکل ب-۶).

نمس و همکاران (۱۵) روش k نزدیک ترین همسایگی و شبکه عصبی مصنوعی را برای برآورد رطوبت در مکش ۳۳ و ۱۵۰۰ کیلو پاسکال مورد استفاده قرار دادند. این محققین نشان دادند که اختلاف معنی داری بین مقدار خطای برآورد روش های روش k نزدیک ترین همسایگی و شبکه عصبی مصنوعی وجود ندارد. اما این محققین بیان کردند که با توجه به سادگی روش k نزدیک ترین همسایگی این روش می تواند در برآورد خصوصیات دیر یافت خاک به خوبی مورد استفاده قرار گیرد. اخیراً بتولا و همکاران (۷) روش k نزدیک ترین همسایگی را برای ایجاد توابع انتقالی در برآورد منحنی رطوبتی خاک مورد استفاده قرار دادند. آنها نشان دادند که این روش بهتر از روش

منابع مورد استفاده

۱. صالحی، م. ح. و ح. خادمی. ۱۳۸۷. نقشه برداری خاک. انتشارات جهاد دانشگاهی دانشگاه صنعتی اصفهان. ۲۱۰ صفحه
۲. حق وردی، ا. ب. قهرمان، م. جلیلی، ع. ا. خشنود یزدی و ز. عربی. ۱۳۹۰. برآورد رطوبت در نقطه پژمردگی دائم و ظرفیت زراعی خاک‌های شمال و شمال شرق ایران با استفاده از روش‌های نزدیک‌ترین همسایه و شبکه‌های عصبی مصنوعی. آب و خاک ۲۴(۴): ۸۱۴-۸۰۴
۳. خدادادی، م. م. ص. عسکری، ف. سرمیدان، ا. حیدری، ح. رفاهی، ع. ا. نوروزی. و ح. ر. متین فر. ۱۳۸۷. تهیه نقشه خاک‌های تحت تأثیر شوری و قلیائیت با استفاده از داده‌های سنجنده ETM+ در بخشی از دشت قزوین، پژوهش و سازندگی در زراعت و باغبانی ۸۰: ۷۷-۹۰
4. Anguilli, F. 2005. Fast condensed nearest neighbor rule. Proceedings of the 22nd International Conference on Machine Learning, Bonn, Germany.
5. Akramkhanov. A. and P. G. Vlek. 2011. The assessment of spatial distribution of soil salinity risk using neural network. Environ. Monit. Assess. DOI 10.1007/s10661-011-2132-5
6. Bannayan, M. and G. Hoogenboom. 2009. Using pattern recognition for estimating cultivar coefficients of a crop simulation model. Field Crop. Res. 111: 290-302.
7. Botula , Y. D., A. Nemes, M. Mbe-Mpie, E. Van Ranst and W. M. Cornelis. 2013. Prediction of water retention of soils from the humid tropics by the nonparametric k-nearest neighbor approach. Vadose Zone J. 12:70-78.
8. Farifteh, J., F. Van der Meer, C. Atzberger and E. J. M. Carranza. 2007. Quantitative analysis of saltaffected soil reflectance spectra: A comparison of two adaptive methods (PLSR and ANN). Remote Sens. Environ. 110: 59-78.
9. Fernández-Buces, N., C. Siebe, S. Cram and J. L. Palasio. 2006. Mapping soil salinity using a ombined spectral response index for bare soil and vegetation (a case study in the former lake Texcoco, Mexico). J. Arid Environ. 65: 644-667.
10. Hengl, T. and D. G. Rossiter. 2003. Supervised landform classification to enhance and replace photointerpretation in semi-detailed soil survey. Soil Sci. Soc. Am. J. 67:1810-1822.
11. Lall, U. and A.Sharma. 1996. A nearest-neighbor bootstrap for resampling hydrologic time series. Water Resour. Res. 32:679-693.
12. Menezes, M. D., S. H. G. Silva, P. R. Owens and N. Curi. 2013. Digital soil mapping approach based on fuzzy logic and field expert knowledge. Ciência e Agrotecnol. 37: 287-29
13. McBratney, A. B., B. Minasny, S. R. Cattle and R. W. Vervoort. 2002. From pedotransfer functions to soil inference systems. Geoderma 109:41-73
14. McKay, J., S. Grunwald, X. Shi and R. F. Long. 2010. Evaluation of the transferability of a knowledge-based soil-landscape model. PP. 165-177. In: Boettinger, J. L.; Howell, D. W. Moore, A. C. Hartemink, A. E. Kienast-Brown, S., eds. Digital soil mapping: bridging research, environmental application, and operation. Springer, New York, NY, USA.
15. Nemes, A., W. J. Rawls and Ya. A. Pachepsky. 2006. Use of the nonparametric nearest neighbor approach to estimate soil hydraulic properties. Soil Sci. Soc. Am. J. 70: 327-336.
16. Nemes A., D. J Timlin, Ya. A. Pachepsky and W. J. Rawls. 2009. Evaluation of the Rawls et al. (1982) pedotransfer functions for their applicability at the U.S. national scale. Soil. Sci. Soc. Am. J. 73:1638-1645.
17. Patel, R., S. Prasher, P. God and R. Bassi. 2002. Soil salinity prediction using artificial neural networks. J. Am. Water Resour. Assoc. 38: 91-100.
18. Pachepsky, Ya. A., D. Timlin and G. Varallyay. 1996. Artificial neural networks to estimate soil water retention from easily measurable data. Soil Sci. Soc. Am. J. 6:727-733.
19. Taghizadeh-Mehrjardi, R. B Minasny, F. Sarmadian and B. P. Malone. 2014. Digital mapping of soil salinity in Ardakan region, central Iran. Geoderma 213: 15-28
20. Tajgardan T, S. Ayoubi, S. Shataee and K. L. Sahrawat. 2010. Soil surface salinity prediction using ASTER Data: Comparing statistical and geostatistical models. Aust. J. Basic Appl. Sci. 4(3): 457-467.
21. Tamari, S., J. H. M. Wosten and J. C. Ruize. 1996. Testing an arificial neural networks for predicting soil hydraulic conductivity. Soil. Sci. Soc. Am. J. 60: 1732-1741.
22. Triantafilis. J. and S. M. Lesch. 2005. Mapping clay content variation using electromagnetic induction techniques. Comput. Electron. Agric. 46: 203-237.
23. Triantafilis, J. and M. Santos F. A. 2009. 2-dimensional soil and vadose zone representation using an EM38 and EM34 and a laterally constrained inversion model. Aust. J. Soil Res. 47: 809-920.
24. Yates, D., S. Gangopadhyay, B. Rajagopalan and K. Strzepak. 2003. A technique for generating regional climate scenarios using a nearest-neighbor algorithm. Water Resour. Res. 39: 712-716.

25. Zhu, Q. and H. S. Lin, 2010. Comparing ordinary kriging and regression kriging for soil properties in contrasting landscapes. *Pedosphere*. 20: 594-606.
26. Zolfaghari, A. A. Tirgar Soltani, M. T., Dyck, M. and A. Weldeyohannes. 2013. Comparison of K-nearest neighbor and artificial neural network methods for predicting cation exchange capacity of soil. 50th anniversary Alberta soil science workshop. *Book of Abstracts*. p. 48.

The Comparison of *k*-NN and ANN for Digital Mapping of Salinity in Chahafzal, Ardekan

S. Ayoubi^{1*}, R. Taghizadeh², Z. Namazi¹, A. Zolfaghari³ and F. Roustae Sadrabadi²

(Received: Nov. 25-2014 ; Accepted : Jan. 20-2016)

Abstract

Digital soil mapping techniques which incorporate the digital auxiliary environmental data to field observation data using software are more reliable and efficient compared to conventional surveys. Therefore, this study has been conducted to use *k*- Nearest Neighbors (*k*-NN) and artificial neural network (ANN) to predict spatial variability of soil salinity in Ardekan district in an area of 700 km², in Yazd province. In this study, 180 soil samples were collected in a grid sampling manner and then soil chemical and physical properties were measured in laboratory. Environmental auxiliary variables were included topographic attributes, remote sensing data (ETM+) and apparent electrical conductivity (ECa). The result of the study showed that the K-mean nearest neighborhood had higher accuracy than ANN models for predicting soil electrical conductivity (ECe). Overall, *k*-NN models could provide significant relationships between soil salinity data and environmental auxiliary variables. The *k*-NN model had the root mean square and coefficient of determination of 12.10 and 0.92, respectively, between predicted and observed ECe data. Also, apparent EC, and remotely sensed indices and wetness index were identified as the most important factors for predicating the soil salinity in the studied area.

Keywords: auxiliary data, digital soil mapping, soil salinity.

1. Dept. of Soil Sci., College of Agric., Isf. Univ. of Techno., Isfahan, Iran.

2. Dept. of Arid Land Management, College of Agric. and Natu. Resour. Ardekan Univ. Ardekan, Iran.

3. Faculty of Desert Sci., Semnan Univ., Semnan, Iran.

*: Corresponding Author, Email: ayoubi@cc.iut.ac.ir