

## استفاده از مدل داده کاوی CANFIS در تخمین ظرفیت تبادل کاتیونی برخی خاک‌های مناطق خشک و نیمه خشک\*\*

❖ فریدون سرمدیان\*؛ استاد گروه علوم و مهندسی خاک، دانشکده مهندسی و فناوری کشاورزی، دانشگاه تهران، ایران.

❖ علی کشاورزی؛ استادیار گروه علوم و مهندسی خاک، دانشکده مهندسی و فناوری کشاورزی، دانشگاه تهران، ایران.

### چکیده

داده کاوی این فرصت را فراهم می‌کند تا داده‌های موجود از خاک، به مناطق دور از دسترس تعمیم داده شوند و داده‌های خاک را در طیفی از مقیاس‌ها، متراکم کرده و یا گسترش داد؛ بطوریکه می‌توان آن را به عنوان یکی از دستاوردهای با ارزش در جهت کمک به تصمیم‌گیری صحیح مدیران اجرایی تلقی نمود. ظرفیت تبادل کاتیونی (CEC) یکی از مهمترین ویژگی‌های شیمیایی خاک‌هاست که توانایی خاک در ذخیره عناصر غذایی و یا عناصر آلاینده در خاک را نشان می‌دهد. اندازه‌گیری CEC خاک‌های مناطق گرم و خشک با دارا بودن خصوصیات مانند مواد آلی پایین و کانی شناسی خاص، به روش‌های معمول سخت و زمان بر است، لذا در این تحقیق از روش CANFIS (Coactive Neuro-Fuzzy Inference Systems) جهت برآورد ظرفیت تبادل کاتیونی برخی خاک‌های مناطق خشک و نیمه خشک استفاده گردید. در این تحقیق از ۸۵ نمونه خاک (بانک داده‌های خاک هدف) موجود در پایگاه داده‌های خاک (۴۴۰ نمونه خاک مرجع) به نسبت ۱:۵ استفاده شد. به منظور بررسی همراستایی در داده‌ها، همبستگی بین متغیرهای مستقل مورد بررسی قرار گرفت و با استفاده از روش رگرسیون حذف پیشرو، مهمترین و تاثیرگذارترین مولفه‌های ورودی بر نتایج خروجی، انتخاب گردید. نتایج این تحقیق نشان داد که روش CANFIS دارای قابلیت و کارایی زیادی در تخمین ظرفیت تبادل کاتیونی خاک با استفاده از ویژگی‌های زودیافتی مانند بافت خاک، ماده آلی و تصاویر ماهواره ای می‌باشد.

واژگان کلیدی: پایگاه داده‌های خاک، داده کاوی، ظرفیت تبادل کاتیونی، ویژگی‌های زودیافت خاک، CANFIS.

## ۱. مقدمه

داده کاوی<sup>۱</sup> طریقه کشف الگوهای معنادار و تشخیص روند از میان داده ها با استفاده از تکنیک تشخیص الگو مانند شبکه عصبی، آموزش ماشین و الگوریتم ژنتیک می باشد که توانایی تشخیص خودکار دانش موجود را دارد و از دیگر روش های تجزیه و تحلیل داده ها به دلیل کشف ساختار پنهان، نسبت ها، الگوها و علائم متفاوت است و می تواند روابط را بازیابی کند [۱۰]. داده کاوی یکی از پیشرفت های اخیر در زمینه فن آوری داده هاست که قابلیت زیادی برای کشف اطلاعات مهم از پایگاه داده را دارد. این فناوری جزئی از تکنولوژی کشف دانش<sup>۲</sup> بوده که شامل مجموعه ای از روش های پیش بینی و گروه بندی است که روابط، الگوها و دانش پنهان شده در درون انبوهی از داده ها را استخراج می کند، بطوریکه در بعضی موارد، داده کاوی را مترادف با تحلیل آماری پیشرفته برای کشف الگوهای مفید، قلمداد می کنند. تفاوت اصولی که بین داده کاوی و آمار وجود دارد این است که داده کاوی یک رهیافت بدون پیش فرض است؛ در حالی که بیشتر تکنیک های آماری معمول نیاز به پیش فرض دارند و آماردان ها در جستجوی معادلاتی برای مطابقت دادن با پیش فرض ها هستند. در مقابل الگوریتم های داده کاوی، می توانند این معادلات را به طور اتوماتیک از اطلاعات موجود در مجموعه داده ها توسعه دهند [۶].

در سال های اخیر، استفاده از تکنیک های داده کاوی در علوم و مهندسی رشد چشمگیر و شایان توجهی داشته و در مسائل خاک و محیط زیست نیز به خوبی جایگاه خود را پیدا کرده است. ظرفیت تبادل کاتیونی<sup>۳</sup> از مهم ترین خصوصیات خاک در ارتباط با توانایی خاک برای نگهداری آب و مواد غذایی و همچنین شاخص خوبی برای کیفیت و بهره وری خاک و مدیریت آلودگی است [۲۵]. همچنین CEC از مهم ترین خصوصیات خاک است که در پایگاه

داده ها مورد نیاز است [۱۷] و به عنوان ورودی در مدل های خاک و زیست محیطی مورد استفاده قرار می گیرد [۱۴]. مقدار CEC بسته به میزان مواد آلی و مقدار و نوع رس و شرایط خاک، متغیر است. وجود مشکلاتی از قبیل تغییر اسیدیته خاک و قدرت یونی محلول خاک در حین اندازه گیری CEC به دلیل وجود مقادیر قابل توجهی کلوئیدهای با بار وابسته به pH مثل اکسیدهای هیدراته آلومینیوم، منگنز، آلفان ها و ماده آلی از یک سو [۲۹] و از سوی دیگر میزان بالای آهک در خاک های مناطق خشک و نیمه خشک مانند ایران، منجر به مشکل تر شدن اندازه گیری ظرفیت تبادل کاتیونی می شود [۲۴]. علاوه بر آن، حضور گچ در خاک های مناطق خشک، اندازه گیری مستقیم را دوچندان مشکل می سازد. با توجه به اندازه گیری دشوار، وقت گیر و پرهزینه CEC، تعیین آن از طریق دیگر خصوصیات زود یافت خاک، مطلوب می باشد.

استفاده از روش های غیرمستقیم در تخمین خصوصیات دیر یافت خاک مانند ظرفیت تبادل کاتیونی اصطلاحاً توابع انتقالی خاک نامیده شده که در نتیجه ی استفاده از این توابع، ضمن صرفه جویی در وقت و هزینه ها، می توان مقایسه مناسبی را برای تعیین بهترین روش تخمین برای پارامترهای خاک، انجام داد. تعیین توابع انتقالی، به صورتی انجام می شود که کاربر این توابع، بتواند با سطوح متفاوت اطلاعاتی، تخمین پارامترهای خاک را با دقت مناسب انجام دهد. مرسوم ترین توابع در تخمین CEC خاک، معادلات رگرسیونی چند متغیره می باشند که توسط محققین مختلف مورد استفاده قرار گرفته و توسعه داده شدند [۳ و ۱۸]. با این وجود، مدل های توسعه داده شده برای یک ناحیه ممکن است نتواند تخمین مناسبی برای یک ناحیه دیگر فراهم کنند [۳۶].

اخیراً شبکه های عصبی به طور موفقیت آمیزی در پیش بینی بسیاری از خصوصیات دیر یافت خاک مورد استفاده قرار گرفته اند [۱۵ و ۲۲]. یکی از مزیت های

<sup>1</sup> Data Mining

<sup>2</sup> Knowledge discovery

<sup>3</sup> Cation Exchange Capacity, CEC

## ۲. روش‌شناسی تحقیق

### ۱.۲. معرفی منطقه مورد مطالعه

منطقه کوهین واقع در شمال غربی استان قزوین برای انجام این طرح انتخاب گردید. حدود تغییرات ارتفاع از ۱۲۰۰ متر تا ۱۶۰۰ متر از سطح دریای آزاد متغیر است. دامنه تغییرات شیب نیز از حدود ۱ تا ۲۵ درصد می‌باشد. وسعت منطقه حدود ۵۱۰۰ هکتار است و در حدفاصل عرض جغرافیایی ۳۶ درجه و ۲۰ دقیقه تا ۳۶ درجه و ۲۳ دقیقه شمالی و طول جغرافیایی ۴۹ درجه و ۳۴ دقیقه تا ۴۹ درجه و ۳۸ دقیقه شرقی می‌باشد. میانگین بارندگی سالیانه ۲۹۰/۳۰ میلی‌متر و میانگین درجه حرارت سالیانه ۱۲/۰۵ درجه سانتی‌گراد است. اقلیم منطقه نیمه خشک بوده و رژیم حرارتی خاک، مزیک و رژیم رطوبتی خاک، زیریک می‌باشد. خاک‌های این منطقه، بر روی رسوبات آبرفتی و مواد مادری آهکی و مارنی به وجود آمده‌اند و در رده‌های انتی سول و اینسپتی سول قرار دارند. این منطقه اغلب تحت کاربری کشت دیم می‌باشد که در اطراف آن، به صورت پراکنده، باغات انگور، زالزالک، سماق و سنجد نیز به چشم می‌خورد. نمونه برداری بر اساس یک شبکه نمونه برداری منظم صورت گرفت. ابعاد شبکه نمونه برداری حدود ۳۰۰ متر در ۳۰۰ متر انتخاب گردید. چند نمونه هم خارج از شبکه نمونه برداری برداشته شد تا نمونه‌ها از اشکال فیزیوگرافی مختلف برداشت شوند. در نهایت ۸۵ نمونه خاک سطحی از عمق صفر تا ۲۰ سانتی‌متری برداشت شد؛ چراکه این عمق، عمق مؤثر زراعی می‌باشد [۱۶]. موقعیت جغرافیایی نقاط نمونه برداری شده به وسیله سیستم موقعیت‌یاب جهانی<sup>۳</sup> ثبت گردید. پس از انتقال نمونه‌ها به آزمایشگاه، نمونه‌های خاک هوا خشک شده و پس از کوبیدن از الک ۲ میلی‌متر عبور داده شدند. خصوصیات اندازه‌گیری شده خاک در این مطالعه شامل توزیع اندازه ذرات، مقدار ماده آلی، ظرفیت تبادل کاتیونی بوده که

استفاده از شبکه‌های عصبی مصنوعی نسبت به مدل‌های قدیمی این است که نیازمند تعیین یک تابع خاص برای بیان رابطه میان داده‌های ورودی و خروجی نیست و رابطه بین داده‌های ورودی و خروجی از طریق فرایند آموزش به دست می‌آید [۳۰]. یک نوع از شبکه‌های عصبی مصنوعی، شبکه‌های پرسپترون چندلایه می‌باشد که از یک الگوریتم آموزش پس انتشار خطا استفاده می‌کند و از آن برای توسعه بسیاری از توابع انتقالی شبکه عصبی استفاده شده است [۳۳].

استفاده از سیستم‌های مبتنی بر قوانین فازی به دلیل پیوسته بودن تغییرات خاک‌ها و وجود عدم قطعیت در داده‌ها رو به گسترش بوده بطوریکه برخی محققین برای ایجاد توابع انتقالی خاک، دو روش رگرسیون آماری و فازی را مورد مقایسه قرار دادند [۲۷]. نتایج این تحقیق نشان داد که روش رگرسیون فازی، در شرایطی که روابط فازی بین متغیرها و به طور کلی در مواردی که با خطاهای ناشی از ابهام در ساختار معادلات رگرسیونی مواجه است، می‌تواند به‌عنوان مکمل و یا جایگزین مناسبی برای روش رگرسیون آماری تلقی شود. قابلیت منطق فازی در علوم خاک مانند فرسایش خاک، نقشه برداری، ارزیابی و مدل کردن فرآیندهای فیزیکی توسط پژوهشگران متعددی به اثبات رسیده است [۱، ۱۲، ۱۹ و ۲۳]. در مورد کاربرد روش تلفیقی عصبی- فازی در علوم خاک که اصطلاحاً نروفازی<sup>۱</sup> نامیده می‌شود، تحقیقات محدودی انجام گرفته است. به‌عنوان نمونه می‌توان به پژوهشی اشاره داشت که جهت پیش بینی مقاومت در برابر فشار از توزیع اندازه ذرات خاک استفاده نموده و برای مدل سازی از دو روش نروفازی و شبکه عصبی بهره بردند. نتایج نشان از کارایی بالاتر روش نروفازی در این تحقیق داشت [۱۲]. لذا با توجه به مشکلات اندازه‌گیری مستقیم ظرفیت تبادل کاتیونی خاک، اهداف این مطالعه عبارت‌اند از: ۱- توسعه مدل داده کاوی CANFIS<sup>۲</sup> با استفاده از دانش داده کاوی و متغیرهای زود یافت خاک ۲- ارزیابی کارایی مدل CANFIS در تخمین ظرفیت تبادل کاتیونی خاک.

1 ANFIS

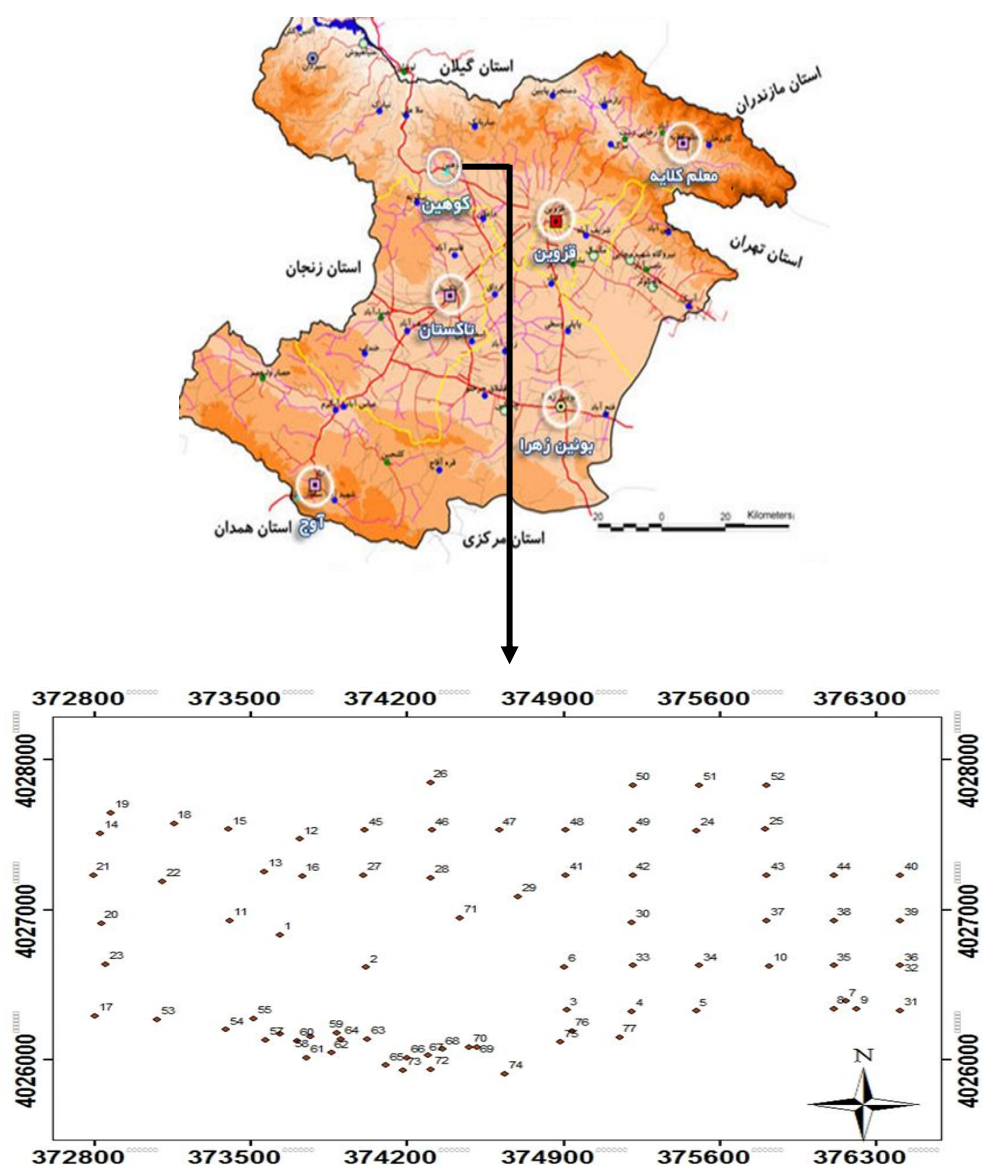
2 Coactive Neuro-Fuzzy Inference Systems

3 GPS

استفاده از مدل رقومی ارتفاع انجام شد. سپس تصویر ماهواره‌ای، با روش‌هایی نظیر نسبت‌گیری طیفی، تجزیه مؤلفه‌های اصلی و تسلدکپ<sup>۱</sup> مورد پردازش و بارزسازی قرار گرفت. طبقه‌بندی تصویر به صورت تلفیقی (رقومی - چشمی) صورت گرفت و مقادیر باندهای یک، دو و سه تصاویر ماهواره، توسط نرم‌افزار PCI Geomatica در هر نقطه نمونه برداری استخراج گردید.

دامنه مقادیر خصوصیات اندازه‌گیری شده در جدول ۱ نشان داده شده است.

برای اندازه‌گیری درصد ذرات خاک از روش هیدرومتری [۸]، کربن آلی خاک به روش و الکی و بلاک [۴] و ظرفیت تبادل کاتیونی نمونه‌های خاک به روش باور [۵] اندازه‌گیری شدند. در این تحقیق، از تصاویر ماهواره لندست ۷ قطعه ۳۵-۱۶۶ مربوط به سال ۲۰۱۳ استفاده شد. تصحیح هندسی تصویر به روش ارتو و با



شکل ۱. موقعیت جغرافیایی منطقه مورد مطالعه (سیستم مختصات UTM) و نمایی از شبکه نمونه برداری

1 Tasseled Cap

جدول ۱. مقادیر خصوصیات اندازه‌گیری شده خاک در این تحقیق (N=۸۵)

خصوصیت	واحد	ماکزیمم	می نیمم	میانگین	انحراف معیار	ضریب تغییرات (درصد)
رس		۵۹	۲۵	۴۰/۷۷	۹/۲۱	۲۲/۵۹
سیلت		۴۴	۱۶	۲۶/۳۵	۵/۷۲	۲۱/۷۰
شن	درصد	۵۷	۱۰	۳۲/۸۴	۱۱/۴۸	۳۴/۹۵
کربن آلی		۱/۳۳	۰/۱۳	۰/۶۸	۰/۲۹	۴۲/۶۴
ظرفیت تبادل کاتیونی	$\text{cmol}^+\text{kg}^{-1}$	۲۹/۴۳	۱۷/۰۳	۲۳/۰۸	۲/۸۵	۱۲/۳۵
باند ۱		۱۵۱	۷۷	۱۱۲/۷۶	۱۸/۳۱	۱۶/۲۳
باند ۲	ندارد	۱۴۲	۵۴	۹۹/۰۲	۲۲/۴۹	۲۲/۷۱
باند ۳		۱۴۴	۶۵	۱۰۹/۳۳	۲۲/۹۲	۲۰/۹۶

## ۲.۲. پیش پردازش داده‌ها

برای انجام پیش‌پردازش داده‌ها، از نرم‌افزار Sigma Plot 12.0 استفاده شد. قبل از انتخاب عوامل ورودی (متغیرهای مستقل) و خروجی (متغیر وابسته)، همه داده‌ها تحت آزمون هم‌راستایی، نرمالیتیه (کولموگروف-اسمیرنوف) و جستجوی داده‌های پرت قرار

گرفتند (شکل ۲) و پس از حذف داده‌های پرت، نرمال بودن داده‌ها تأیید شد. سپس همبستگی بین متغیرهای مستقل مورد بررسی قرار گرفت (جدول ۲) و با استفاده از روش رگرسیونی حذف پیش‌رو، مهم‌ترین و تأثیرگذارترین مؤلفه‌های ورودی بر نتایج خروجی، انتخاب گردید [۹].

جدول ۲. ضریب همبستگی پیرسون (r) بین متغیرهای منتخب در منطقه مورد مطالعه

متغیر	رس (درصد)	کربن آلی (درصد)	باند ۳	ظرفیت تبادل کاتیونی ( $\text{cmol}^+\text{kg}^{-1}$ )
رس (درصد)	۱	۰/۰۵۹	-۰/۳۴۹**	۰/۴۷۶**
کربن آلی (درصد)		۱	-۰/۴۷۷**	۰/۷۲۷**
باند ۳			۱	-۰/۵۸۴**
ظرفیت تبادل کاتیونی ( $\text{cmol}^+\text{kg}^{-1}$ )				۱

\*\* معنی‌دار در سطح یک درصد

## ۳.۲. داده کاوی بانک اطلاعاتی خاک با استفاده

### از روش CANFIS

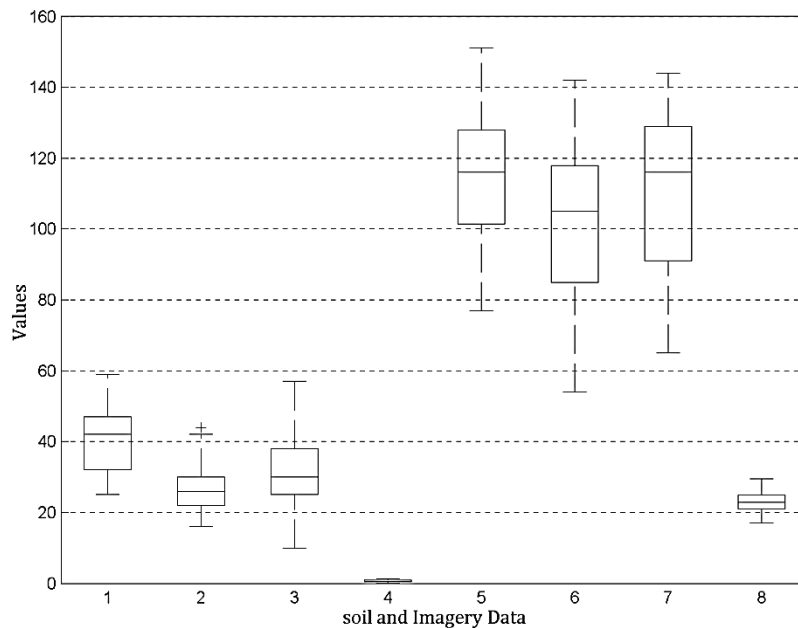
پس از انجام پیش‌پردازش داده‌ها و حذف داده‌های پرت، تعداد داده‌ها به ۷۷ داده کاهش یافت. ساختار CANFIS مشابه مدل ANFIS (سیستم عصبی-فازی یا نروفازی تطبیقی) بوده که اولین بار توسط ژانگ معرفی شد [۱۱]. این مدل یک سیستم فازی سوگنو را در یک

ساختار شبکه عصبی اجرا می‌کند و برای فرآیند آموزش، از ترکیبی از روش‌های آموزش پس انتشار خطا و حداقل مربعات خطا استفاده می‌کند.

مدل CANFIS یک شبکه چند لایه پیش‌رونده بوده که از الگوریتم‌های یادگیری شبکه‌های عصبی مصنوعی و استدلال‌های فازی برای توصیف مکان ورودی به یک مکان خروجی استفاده می‌کند [۷]. CANFIS دارای خصوصیتی از قبیل یادگیری، نوشتن، تدوین کردن و

مؤثر می‌باشد. CANFIS از سویی از توانایی یادگیری شبکه‌های عصبی برای تعریف نحوه ارتباط ورودی-خروجی بهره گرفته و از سویی دیگر، با استفاده از قوانین فازی، ساختار ورودی سیستم را تهیه و پایه‌ریزی می‌کند.

طبقه‌بندی است. هم‌چنین دارای این حسن بوده که اجازه استخراج قوانین فازی از داده‌های متعدد و تخصصی را داده و به‌گونه‌ای متناسب، قوانین پایه و اساسی را می‌سازد. مشکل اصلی این سیستم، مدت زمان لازم جهت تهیه ساختار و نحوه آموزش سیستم و تعیین پارامترهای



شکل ۲. روند جستجوی داده‌های پرت

(مقادیر ۱ تا ۸ در محور افقی به ترتیب نشان‌دهنده داده‌های رس، سیلت، شن، کربن آلی، باند ۱، باند ۲، باند ۳ و ظرفیت تبادل کاتیونی خاک هستند)

به هر یک از قوانین، نرمال می‌شوند. در لایه چهارم، گره‌های موجود به صورت تطبیق پذیر درآمده و هر تابع موجود، یک مدل درجه اول با پارامترهای حاصله ارائه می‌دهد و لایه پنجم، لایه خروجی است. به منظور بهینه‌نموده مدل CANFIS، لازم است که پارامترهای مقدماتی در لایه اول و پارامترهای حاصله در لایه چهارم تعریف شوند. پارامترهای لایه اول، به‌عنوان بخش مرکزی و دامنه تابع عضویت مقدم مطرح می‌شوند.

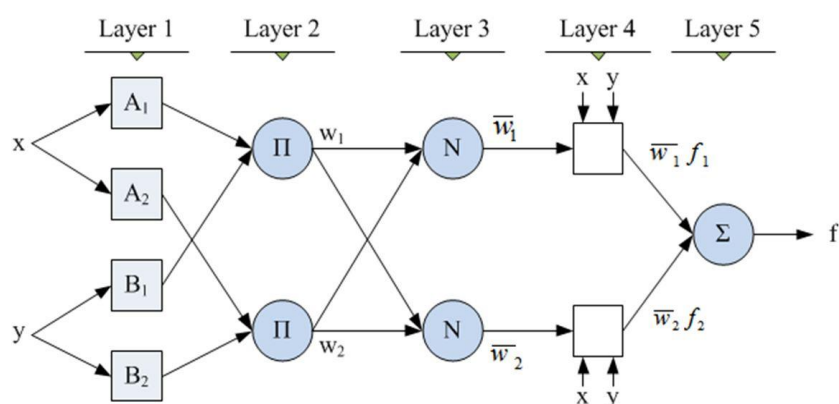
بیشتر محققین الگوریتم آموزش هیبرید را برای تعیین پارامترهای CANFIS پیشنهاد دادند [۱۱]. نحوه آموزش شبکه عصبی در ساختار CANFIS به‌صورت یادگیری نظارتی است. از این‌رو، با آموزش شبکه، توابع

مدل CANFIS انتخاب شده شامل پنج لایه است (شکل ۳). در لایه اول یا عناصر ورودی، گره‌های موجود در این لایه، سری‌های فازی در ارتباط با بخش مقدم قوانین فازی را ارائه می‌دهند. این لایه دارای پارامترهایی است که موقعیت بخش مرکزی و اصلی هر سری فازی را کنترل می‌کند. در این تحقیق، از تابع عضویت گوسین برای تعیین درجات عضویت استفاده شد. خروجی‌های این لایه، مقادیر مقدم توابع عضویت، متناسب با ورودی‌های فازی سیستم هستند.

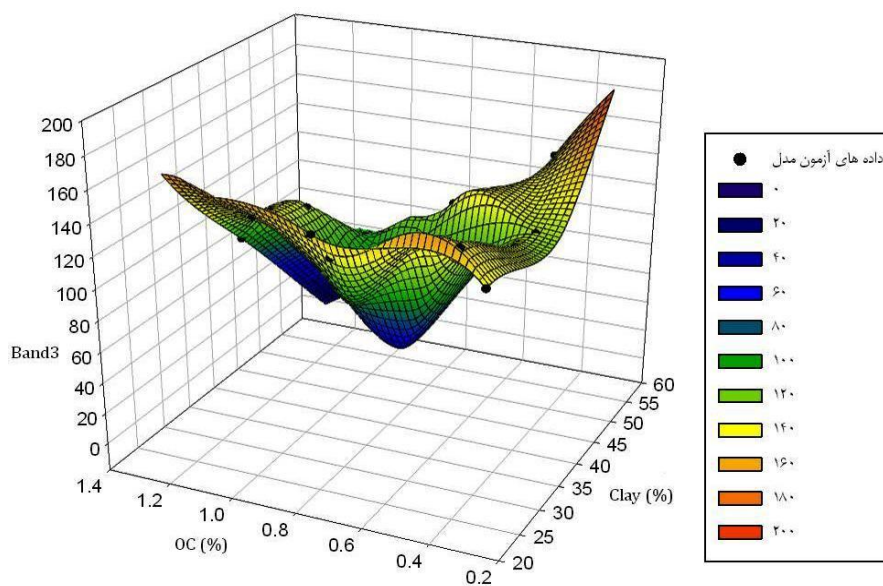
در لایه دوم، هر گره موظف به محاسبه نتایج ورودی‌های خود است و میزان فعال‌سازی قوانین خاص فازی تعیین می‌شود. در لایه سوم، درجه فعالیت مربوط

گرفتند (شکل ۴). در طی فرآیند یادگیری شبکه در ساختار CANFIS، میزان فراگیری توسط شبکه به وسیله توابع هدف به طور مرتب سنجیده می‌شود و در نهایت شبکه‌ای با کم‌ترین میزان خطا مورد پذیرش قرار می‌گیرد [۲۴]. در این مطالعه، به علت کارایی، سادگی و سرعت بالا، الگوریتم آموزشی لوببرگ-مارکوارت مورد استفاده قرار گرفت.

ناشناخته ارائه شده توسط داده‌های آموزشی یاد گرفته شده و متناسب با ارزش پارامترهای ورودی، خروجی تعیین می‌شود. به همین دلیل، در ساختار CANFIS از الگوریتم آموزش هیبرید در محیط برنامه‌نویسی MATLAB (R2013a) استفاده شد. در این مطالعه، داده‌ها با استفاده از نرم‌افزار Excel تصادفی شده و ۸۰ درصد داده‌ها به عنوان داده‌های آموزشی و ۲۰ درصد به عنوان داده‌های آزمون یا تست مورد استفاده قرار



شکل ۳. ساختار مدل CANFIS به همراه لایه‌های آن



شکل ۴. تغییرات سه بعدی داده‌های آزمون در فرآیند مدل‌سازی ظرفیت تبادل کاتیونی خاک (مثال موردی: باند ۳)

## ۴.۲. ارزیابی کارایی مدل

عملکرد مدل، به وسیله داده‌های آزمونی که در آموزش شبکه استفاده نشده‌اند، ارزیابی شد. معیارهای مورد استفاده در ارزیابی مدل، ریشه دوم میانگین مربعات خطا (RMSE) و ضریب تعیین ( $R^2$ ) بود که با استفاده از روابط زیر [۳۷] محاسبه شد:

$$RMSE = \sqrt{\sum_{i=1}^N (y_o - y_p)^2 / N} \quad (1)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_o - y_p)^2}{\sum_{i=1}^N (y_o - \bar{y}_i)^2} \quad (2)$$

که در آن‌ها:  $y_o$  مقادیر اندازه‌گیری شده،  $y_p$  مقادیر پیش‌بینی شده،  $\bar{y}_i$  میانگین مقادیر اندازه‌گیری شده و  $N$  تعداد کل داده‌هاست. فاکتور RMSE دقت و اعتبار دسته داده‌های آموزش و آزمون را اندازه‌گیری کرده که هر چقدر به صفر نزدیک تر باشد، بیانگر خطای کمتر روش

می‌باشد.

## ۳. نتایج

## ۱.۳. خلاصه آماری و توصیف داده‌ها

از آنجایی که شبکه عصبی در ساختار نروفازی، فاقد قدرت برون‌یابی بالا می‌باشد و توانایی تعمیم‌دهی آن، فقط در چارچوب درون‌یابی مطرح است، لذا داده‌های آموزشی طوری باید انتخاب گردند که نماینده کل داده‌ها باشند، یعنی تمامی حالات ممکن در داده‌ها از قبیل مقادیر حدی (حداکثر و حداقل) را در خود داشته باشند. در این تحقیق، تقسیم داده‌ها به وسیله تصادفی کردن نمونه‌ها و سپس انتخاب تصادفی ۸۰ درصد از داده‌ها به منظور آموزش مدل و ۲۰ درصد باقیمانده برای آزمون مدل بود [۲]. خلاصه آماری خصوصیات خاک‌های مورد استفاده برای آموزش و آزمون مدل، در جدول ۳ ارائه شده است.

جدول ۳. خلاصه آماری داده‌های مورد استفاده در آموزش و آزمون مدل CANFIS

ضریب تغییرات (درصد)	انحراف معیار	میانگین	می نیمم	ماکزیمم	واحد	خصوصیت	
۲۱/۹۱	۹/۰۳	۴۱/۲۰	۲۸	۵۶/۳۵	درصد	رس	داده‌های آموزش (N=۶۲)
۴۵/۷۵	۰/۳۰۲	۰/۶۶	۰/۱۸	۱/۰۲		کربن آلی	
۱۲/۲۶	۲/۸۲	۲۲/۹۹	۱۹/۲۵	۲۷/۶۰	cmol <sup>+</sup> kg <sup>-1</sup>	ظرفیت تبادل کاتیونی	
۱۹/۷۰	۲۲/۱۸	۱۱۲/۵۶	۷۰	۱۴۲	ندارد	باند ۳	داده‌های آزمون (N=۱۵)
۲۵/۸۲	۱۰/۰۷	۳۹	۲۵	۵۹	درصد	رس	
۳۵/۰۵	۰/۲۷۱	۰/۷۷۳	۰/۱۳	۱/۳۳		کربن آلی	
۱۲/۹۲	۳/۰۳	۲۳/۴۵	۱۷/۰۳	۲۹/۴۳	cmol <sup>+</sup> kg <sup>-1</sup>	ظرفیت تبادل کاتیونی	
۲۴/۴۶	۲۶/۶۳	۱۰۸/۸۵	۶۵	۱۴۴	ندارد	باند ۳	

نتایج این تحقیق نشان می‌دهد که CEC خاک با میزان رس ( $r=0/476$ ) و کربن آلی خاک ( $r=0/727$ )

ضریب همبستگی پیرسون ( $r$ ) بین متغیرهای منتخب در منطقه مورد مطالعه در جدول ۲ نشان داده شده است.



متغیرهایی به عنوان ورودی به مدل استفاده شد که همبستگی معنی داری با ظرفیت تبادل کاتیونی داشتند (جدول ۲) و بدین منظور، از متغیرهای رس، کربن آلی و باند ۳ برای تخمین ظرفیت تبادل کاتیونی خاک استفاده شد. مدل ابتدائاً با دسته داده‌های آموزشی، آموزش داده شده و سپس مقادیر خطای پیش بینی در مقابل تعداد نرون‌های لایه پنهان بررسی شد. پیدا کردن تعداد بهینه نرون‌ها در لایه پنهان، یکی از مراحل مهم در توسعه مدل CANFIS می‌باشد. در طراحی ساختار CANFIS، تعداد نسبتاً زیاد واحدهای مخفی، باعث بیش برآورد می‌شود، در حالی که تعداد کم واحدهای مخفی باعث کم برآورد می‌شود. به منظور پیدا کردن تعداد بهینه نرون‌های مخفی، از روش سعی و خطا استفاده شد.

در ساختار CANFIS، مجموعه‌های فازی را می‌توان به عنوان وزن ارتباطات و متغیرهای ورودی و خروجی را به عنوان نرون‌ها تفسیر کرد و الگوریتم یادگیری، ساختارها، پارامترها و یا هر دوی آن‌ها را اصلاح می‌کند. نتایج خصوصیات بهترین مدل CANFIS با توجه به آماره‌های ریشه دوم میانگین مربعات خطا (RMSE) و ضریب تعیین ( $R^2$ ) در جدول ۴ ارائه شده است.

همبستگی مستقیم و معنی دار و با باند ۳ ( $r = -0.584$ ) همبستگی معکوس و معنی دار دارد. اصولاً وارد کردن داده‌ها به صورت خام باعث کاهش سرعت و دقت شبکه می‌شود. برای اجتناب از چنین شرایطی و همچنین به منظور یکسان نمودن ارزش داده‌ها، قبل از آموزش CANFIS، داده‌ها استاندارد (نرمال) شدند. این کار مانع از کوچک شدن بیش از حد وزن‌ها می‌گردد و همچنین این کار بدین علت صورت می‌گیرد که داده‌ها به اعدادی ما بین صفر تا ۱ تبدیل گردند، زیرا خروجی اکثر توابع آستانه، اعدادی بین صفر تا یک می‌باشد. داده‌ها با استفاده از فرمول زیر نرمال شدند و به مقادیر بین صفر تا یک تبدیل شدند:

$$y_{normal} = (y_o - y_{min}) / (y_{max} - y_{min}) \quad (3)$$

که در آن،  $y_{max}$  ماکزیمم داده‌ها،  $y_{min}$  حداقل داده‌ها و  $y_o$  داده‌های مشاهده‌ای هستند.

### ۲.۳. تخمین ظرفیت تبادل کاتیونی خاک با

#### استفاده از مدل CANFIS

به منظور توسعه مدل CANFIS در این تحقیق

جدول ۴. ویژگی‌های مدل CANFIS انتخابی برای تخمین CEC خاک

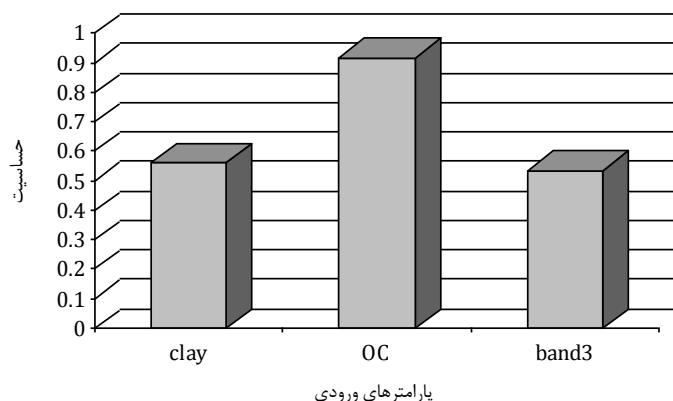
خصوصیت	روش دفازی کردن	روش یادگیری	نوع تابع عضویت	تعداد توابع عضویت	تکرار	RMSE	$R^2$
ظرفیت تبادل کاتیونی	میانگین وزنی	هیبرید	گوسین	۳	۸۹۴	۱/۳۴	۰/۸۰۶

حساسیت برای ورودی‌های انتخابی در شکل ۵ آمده است.

همان طور که در شکل ۵ مشاهده می‌شود، مدل CANFIS بیشترین حساسیت را به کربن آلی خاک از خود نشان می‌دهد که نشان دهنده تأثیر زیاد مواد آلی بر افزایش CEC خاک به دلیل سطح ویژه زیاد و داشتن گروه‌های عاملی می‌باشد [۱۷].

### ۳.۳. بررسی آنالیز حساسیت مدل

به منظور تعیین اثر هر یک از پارامترهای ورودی بر ظرفیت تبادل کاتیونی، لازم است آنالیز حساسیت انجام گیرد. آنالیز حساسیت بر روی پارامترهای ورودی شبکه به روش Statsoft نیز انجام شد تا تأثیر پارامترهای ورودی بر میزان خروجی بهتر نشان داده شود. نتایج آنالیز

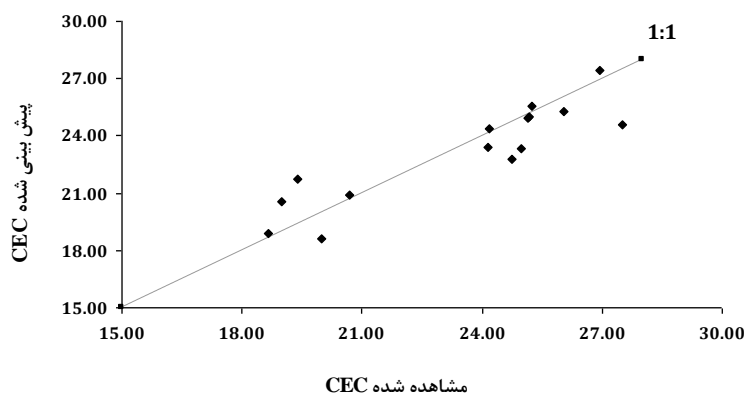


شکل ۵. نتایج آنالیز حساسیت مدل برای پارامترهای ورودی

داده‌های آزمون ترسیم شد (شکل ۶). با توجه به این شکل ملاحظه می‌شود بهترین خط برازش داده شده دارای زاویه‌ای نزدیک به ۴۵ درجه (خط یک به یک) می‌باشد که نشان از دقت بالای برآورد توسط مدل CANFIS در تخمین پارامتر مورد بررسی دارد.

#### ۴.۳. ارزیابی عملکرد مدل

در این مرحله به منظور ارزیابی کارایی مدل CANFIS، از داده‌های آزمون برای پیش بینی ظرفیت تبادل کاتیونی استفاده شد و مقادیر ظرفیت تبادل کاتیونی پیش‌بینی شده در مقابل اندازه‌گیری شده، برای



شکل ۶. دیاگرام پراکنش برای مقادیر مشاهداتی و پیش‌بینی شده ظرفیت تبادل کاتیونی با استفاده از مدل

خصوصیات زود یافت خاک شامل متغیرهای رس، کربن آلی و باند ۳ تصاویر ماهواره لندست ۷ و با توجه به آماره‌های ریشه دوم میانگین مربعات خطا (RMSE) و ضریب تعیین ( $R^2$ ) استفاده شد. همان‌گونه که در جداول

#### ۴. بحث و نتیجه‌گیری

در این پژوهش، از مدل داده‌کاوی CANFIS به منظور امکان‌سنجی تخمین ظرفیت تبادل کاتیونی خاک در اراضی تحت کاربری کشت دیم با استفاده از

مقایسات مبتنی بر توسعه مدل‌های رگرسیونی و شبکه‌های عصبی مصنوعی و اجرای آنالیز حساسیت بیان شده است [۱۳ و ۲۱].

کاربرد مدل CANFIS در تخمین CEC خاک با استفاده از متغیرهای زود یافت توانست حدود ۸۱ درصد تغییرات در این مطالعه را توجیه کند. فاکتورهای مختلفی می‌تواند بر جوابگویی بهینه این مدل دخیل باشند که از جمله می‌توان به متفاوت بودن نوع و مورفولوژی رس در نواحی نمونه‌برداری و همچنین تغییرپذیری مکانی بالای خصوصیات خاک در ناحیه مورد مطالعه اشاره کرد. به نظر می‌رسد یکی از عواملی که می‌تواند دقت نتایج تخمین را افزایش دهد، همگنی بیشتر خاک‌های مورد مطالعه در کاربری یکسان می‌باشد [۲]. استفاده توأم از دو ویژگی میزان رس و کربن آلی خاک، منجر به افزایش چشمگیر و قابل توجه آماره  $R^2$  و نزدیکی پراکنش نقاط پیش‌بینی شده به خط یک به یک (شکل ۶) توسط مدل CANFIS شده که با نتایج اکثر محققین در استفاده از دو ویژگی میزان رس و کربن آلی خاک هم‌خوانی دارد [۲۶]. یکی از نقاط قوت در ساختار مدل CANFIS، حضور شبکه عصبی بوده و طبق نظر برخی پژوهشگران در طراحی شبکه‌های عصبی نوع خاصی از معادلات لازم نیست و با ایجاد رابطه مناسب بین داده‌های ورودی و خروجی می‌توان به نتایج مناسب دست پیدا کرد [۳۱]. همچنین به علت وجود روابط غیرخطی میان متغیرهای مستقل و پیش‌بینی شونده (ظرفیت تبادل کاتیونی) مدل CANFIS عملکرد خوبی را نشان داده است. نکته دیگر این است که انتخاب صحیح مجموعه داده‌های آموزش و آزمون که جزئی از فرآیند داده کاوی است، می‌تواند منجر به افزایش دقت مدل CANFIS انتخابی در پیش‌بینی CEC خاک شده بطوریکه برخی پژوهشگران نشان دادند که اگر ناپایداری داده‌ها بالا باشد (وجود داده‌های پرت، اشتباه و اریب) شبکه عصبی عملکرد خوبی نخواهد داشت، اما زمانی که داده‌هایی با دقت بالا به کار برده شوند، شبکه عصبی

۱ و ۳ مشاهده می‌شود، خاک‌های مورد مطالعه دامنه به نسبت گسترده‌ای از تغییرات را نشان می‌دهند. دامنه تغییرات ذرات رس از ۲۵ تا ۵۹ درصد در ناحیه مورد مطالعه متغیر می‌باشد. مقدار کربن آلی خاک نیز بین ۰/۱۳ تا ۱/۳۳ درصد با میانگین ۰/۶۸ در ناحیه مورد مطالعه متغیر است. با توجه به اینکه این منطقه اغلب تحت کاربری کشت دیم می‌باشد، ضریب تغییرات کربن آلی خاک بالا بوده که می‌تواند در اثر کاربرد کود و اصلاح خاک در ناحیه مورد مطالعه باشد. برخی محققین نیز در نتایج مشابه، بالا بودن ضریب تغییرات ماده آلی خاک‌ها را در نتیجه استفاده از کود و مواد اصلاحی خاک گزارش نمودند [۲۱]. گزارش‌ها محدودی در ارتباط با تأثیر اطلاعات طیفی و انعکاسی خصوصیات خاک، در پیش‌بینی ظرفیت تبادل کاتیونی خاک وجود دارد، به طوری که پاره‌ای از پژوهشگران [۲۶ و ۲۸] نیز به همبستگی بالا و معنی‌دار CEC خاک با میزان رس و کربن آلی خاک اشاره نموده‌اند.

مقادیر آماره RMSE تحقیق حاضر به نتایج مطالعات برخی محققین نزدیک بوده ( $RMSE = 1/04$ ) و هم‌خوانی دارد [۲۰]. بررسی‌های انجام شده در یک مطالعه نشان داد که با به کارگیری دو ویژگی ماده آلی و رس، در توابع انتقالی پیشنهادی به RMSE نزدیکی (در حدود ۱/۵) و  $R^2$  بالاتری (در حدود ۰/۸۵) می‌توان دست یافت [۳]. در پژوهشی دیگر، در خاک‌های اندی‌سول آمریکا، محققین با برازش مدل‌های نمایی و با استفاده از سه ویژگی کربن آلی خاک، اسیدیته و رطوبت پژمردگی دائم، به میانگین RMSE برابر با ۰/۳ و میانگین  $R^2$  برابر با ۰/۷ رسیدند [۳۲]. از سوی دیگر، تأثیر مثبت استفاده از پارامتر سطح ویژه در افزایش میزان  $R^2$  و کاهش میزان RMSE در بهبود تخمین ظرفیت تبادل کاتیونی خاک نیز توسط تعدادی از محققین گزارش گردیده است [۲۶]. اثرات مثبت و معنی‌دار مواد آلی و میزان رس خاک بر افزایش CEC خاک‌ها نیز در مطالعات متعددی توسط سایر پژوهشگران در اراضی بکر و کشاورزی و از طریق

داشت که با توجه به ماهیت ابهامی پدیده های مرتبط با خاک یا تقریبی بودن مقادیر اندازه گیری شده خصوصیات مختلف خاک و عدم قطعیت موجود در داده ها، استفاده از مدل های هیبرید مانند CANFIS که از مجموعه های فازی استفاده می کنند، در پردازش و برازش توابع انتقالی خاک می توانند تا حد زیادی مفید واقع گردند.

کارایی بهتری را نشان می دهد [۳۵]. با توجه به مشکلاتی که در اندازه گیری ظرفیت تبادل کاتیونی به طور مستقیم به خصوص در مناطق خشک و نیمه خشک وجود دارد، پیشنهاد می گردد در مطالعات آتی از مدل CANFIS نیز برای به دست آوردن روابطی منطقی در هر منطقه برای پیش بینی سایر پارامترهای خاک، به خصوص خصوصیات هیدرولیکی خاک استفاده گردد. در پایان باید اذعان

## References

- [1] Ahamed, T., Rao, K. and Murthy, J. (2000). Fuzzy class membership approach to soil erosion modeling. *Agricultural Systems*, 63(2), 97-110.
- [2] 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. *European Journal of Soil Science*, 53, 748-757.
- [3] Bell, M.A. and Van Keulen, H. (1995). Soil pedotransfer functions for four Mexican soils. *Soil Science Society of America Journal*, 59, 865-871.
- [4] Black, C.A. (1982). Method of soil analysis, Chemical and microbiological properties. *American Society of Agronomy, INC*, 2: 1965.
- [5] Bower, C.A., Reitmeir, R.F. and Fireman, M. (1952). Exchangeable cation analysis of saline and alkali soils. *Soil Science*, 73, 251-261.
- [6] Cabena, P.H., Stadler, R., Verhees, J. and Zanasi, A. (1998). *Discovering Data Mining: From Concept to Implementation*, IBM, New Jersey, 195 pp.
- [7] Firat, M. and Gungor, M. (2007). River flow estimation using adaptive neuro fuzzy inference system. *Mathematics and Computers in Simulation*, 75, 87-96.
- [8] Gee, G.W. and Bauder, J.W. (1986). *Particle size analysis*. Pp. 383-411. In: A. Klute (Ed.), *Methods of Soil Analysis. Part 1*. Am. Soc. Agron., Madison, Wisconsin, USA.
- [9] Ghorbani Dashtaki, Sh. and Homaei, M. (2002). Prediction of parametric hydraulic function in unsaturated soils using pedotransfer functions. *Agriculture Engineering Research Journal*, 3, 3-15.
- [10] Ghorbani, Kh., Khalili, A., Alavipanah, S.K. and Nakhaezadeh, Gh. (2010). Comparative Study of the Meteorological Drought Indices (Spi and Siap) Using Data Mining Method (Case Study of Kermanshah Province). *Journal of Water and Soil*, 24(3), 417-426.
- [11] Jang, J., Sun, C. and Mizutani, E. (1997). *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*. Prentice Hall, Upper Saddle River, New Jersey, USA.
- [12] Kalkan, E., Akbulut, S. and Tortum, A. (2008). Prediction of the unconfined compressive strength of compacted granular soils by using inference systems. *Environmental Geology*, 5, 324-330.
- [13] Kashi, H., Ghorbani, H., Emamgholizadeh, S. and Hashemi, S.A.A. (2013). The Estimation of Soil Cation Exchange Capacity in Disturbed and Undisturbed Soils Using Artificial Neural Networks and Multiple Regressions. *Journal of Water and Soil*, 27(3), 472-484.
- [14] Keller, A., Von Steiger, B., Van der Zee, S.T. and Schuline, R. (2001). A stochastic empirical model for regional heavy metal balances in agroecosystems. *Journal of Environmental Quality*, 30, 1976-1989.
- [15] Landeras, G., Ortiz-Barredo, A. and Lopez, J.J. (2008). Comparison of artificial neural network models and empirical and semi-empirical equations for daily reference evapotranspiration estimation in the Basque Country (Northern Spain). *Agricultural Water Management*, 95, 553-565.
- [16] Mahmoodi, Sh. and Hakimian, M. (2011). *Fundamentals of soil science*. University of Tehran Press, 704 pp.
- [17] Manrique, L.A., Jones, C.A. and Dyke, P.T. (1991). Predicting cation exchange capacity from soil physical and chemical properties. *Soil Science Society of America Journal*, 55, 787-794.
- [18] McBratney, A.B., Minasny, B., Cattle, S.R. and Vervoort, R.W. (2002). From pedotransfer functions to soil inference systems. *Geoderma*, 109, 41-73.
- [19] McBratney, A.B. and Odeh, I.O.A. (1997). Application of fuzzy sets in soil science: fuzzy logic, fuzzy measurements and fuzzy decisions. *Geoderma*, 77, 85-113.
- [20] Mehrabanian. M., Taghizadeh Mehrjardi, R. and Dehghani, F. (2010). Assessing the Efficiency of Pedotransfer Functions for Estimating CEC in some Calcareous and Gypsiferous Soils of Yazd Province. *Journal of Water and Soil Conservation*, 17(1), 113-127.

- [21] Memarian Fard, M. and Beigi Harchagani, H. (2009). Comparison of artificial neural network and regression pedotransfer functions models for prediction of soil cation exchange capacity in Chaharmahal - Bakhtiari province. *Journal of Water and Soil*, 23(4), 90-99.
- [22] Merdun, H., Meral, O.C. and Apan, R.M. (2006). Comparison of artificial neural network and regression Pedotransfer functions for predict of water retention and saturated hydraulic conductivity. *Soil and Tillage Research*, 90, 108-116.
- [23] Metternicht, G. and Gonzalez, S. (2004). Foundations of a fuzzy exploratory model for soil erosion hazard prediction. *Environmental Modelling & Software*, 20 (6), 715-728.
- [24] Minasny B. and Mcbratney, A.B. (2002). The neuro-m method for fitting neural network parametric pedotransfer functions. *Soil Science Society of America Journal*, 66, 352-361.
- [25] Mirkhani, R., Shabanpour, M. and Saadat, S. (2005). Using particle size distribution and organic carbon percentage to predict the cation exchange capacity of soils of Lorestan province. Iran. *Journal of Soil and Water Sciences*, 19(2), 235-242.
- [26] Mohajjer, R., Salehi, M.H. and Beigi Herchegani, H. (2009). Estimating Soil Cation Exchange Capacity (in View of Pedotransfer Functions) Using Regression and Artificial Neural Networks and the Effect of Data Partitioning on Accuracy and Precision of Functions. *JWSS - Isfahan University of Technology*, 13 (49), 83-97.
- [27] Mohammadi, J. and Taheri, M. (2005). Estimation of pedotransfer function using fuzzy regression. *Journal of Agriculture Science and Technology*, 2, 51-60.
- [28] Nourbakhsh, F., Jalalian, A. and Shariatmadari, H. (2003). Estimation of Cation Exchange Capacity from Some Soil Physical and Chemical Properties. *JWSS - Isfahan University of Technology*, 7 (3), 107-118.
- [29] Sayegh, A.H., Khan, N.A., Khan, P. and Ryan, J. (1978). Factors affecting gypsum and cation exchange capacity determinations in gypsiferous soils. *Soil Science*, 125(5), 294-300.
- [30] Schaap, M.G. and Bouten, W. (1996). Modelling water retention curves of sandy soils using neural networks. *Water Resources Research*, 32, 3033-3040.
- [31] Schaap, M.G., Leij, F.J. and Van Genuchten, M.T. (1998). Neural network analysis for hierarchical prediction of soil hydraulic properties. *Journal of Soil Science Society of America*, 62, 847-855.
- [32] Seybold, C.A., Grossman, R.B. and Reinsch, T.G. (2005). Predicting Cation Exchange Capacity for soil survey using linear models. *Soil Science Society of America Journal*, 69, 856-863.
- [33] Sunil, K., Sinha, M. and Wang, C. (2008). Artificial Neural Network Prediction Models for Soil Compaction and Permeability. *Geotechnical and Geological Engineering*, 26, 47-64.
- [34] Tadesse, T., Wilhite, D., Harms, S., Hayes, H. and Oddard, S. (2004). Drought Monitoring using Data Mining Techniques: A case study for Nebraska, USA. *Natural Hazards*, 33(1), 137-159.
- [35] Tamari, S., Wosten, J.H.M. and Ruz-suarez, J.C. (1996). Testing an artificial neural network for predicting soil hydraulic conductivity. *Soil Science Society of America Journal*, 60, 1732-1741.
- [36] Wagner, B., Tranawski, V.R., Hennings, V., Muller, U., Wessolek, G. and Playge, R. (2001). Evaluation of pedo-transfer function for unsaturated soil hydraulic conductivity using an independent data set. *Geoderma*, 102, 275-279.
- [37] Wosten, J.H.M., Pachepsky, Y.A. and Rawls, W.J. (2001). Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics. *Journal of Hydrology*, 251, 123-150.