



Estimating Indices of Irrigation Water Quality (TH, SAR, and RSC) by EC and pH Using Machine Learning Algorithms

M. Tazeh , **M. Shirmardi**, and **S. Kalantari**

Associate Prof., Department of Nature Engineering, Faculty of Agriculture & Natural Resources, Ardakan University, P.O. Box 184, Ardakan, Iran. mtazeh@ardakan.ac.ir

Associate Prof., Department of Horticulture, Faculty of Agriculture and Natural Resources, Ardakan University, Ardakan, Iran. shirmardi@ardakan.ac.ir

Associate Prof., Department of Nature Engineering, Faculty of Agriculture & Natural Resources, Ardakan University, P.O. Box 184, Ardakan, Iran. skalandari@ardakan.ac.ir

Article Info	Abstract
Research Article	The parameters sodium adsorption ratio (SAR), total hardness (TH), and residual sodium carbonate (RSC) play a vital role in agricultural activities, but their measurement requires specialized equipment and considerable time. The aim of this study was to predict SAR, TH, and RSC values based on measured electrical conductivity (EC) and pH using different Machine Learning algorithms. For this purpose, seven algorithms including Decision Tree, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Cubist, Random Forest, Super Learning, and XGBoost were examined. In this study, data from 132 water samples with laboratory-measured EC, pH, SAR, RSC, and TH values were used. Results showed that the use of Super Learning significantly improved prediction accuracy compared to the other algorithms. Specifically, the R^2 values obtained using both EC and pH simultaneously for predicting TH, RSC, and SAR were 0.91, 0.88, and 0.68, respectively, which were higher than those of other methods. Using EC alone yielded R^2 values of 0.87 for TH, 0.86 for RSC, and 0.73 for SAR, while using pH alone reduced the R^2 values to 0.12, 0.11, and 0.9, respectively. The findings of the present study indicate that with simple EC and pH measurements combined with machine learning algorithms, especially Super Learning, it is possible to accurately predict TH, RSC, and SAR parameters, thereby reducing costs and enabling faster decision-making in water resources management.
Received: August 05, 2025	
Accepted: November 01, 2025	
Keywords: Model performance, Model validation, Irrigation water quality	
Corresponding author's email: mtazeh@ardakan.ac.ir	

Cite this article: Tazeh, M., Shirmardi, M. and Kalantari, S., 2025. Estimating Indices of Irrigation Water Quality (TH, SAR, and RSC) by EC and pH Using Machine Learning Algorithms. *Journal of Water Research in Agriculture*, 39(3), pp.217-236.



DOI:<https://doi.org/10.22092/jwra.2025.369886.1086>



نشریه پژوهش آب در کشاورزی

<https://wra.areeo.ac.ir/>

برآورد شاخص‌های کیفیت آب آبیاری (SAR، TH و RSC) با استفاده از

شاخص‌های EC و pH توسط الگوریتم‌های یادگیری ماشین

مهدی تازه* , مصطفی شیرمردی و سعیده کلانتری

دانشیار گروه مهندسی طبیعت، دانشکده کشاورزی و منابع طبیعی، دانشگاه اردکان، اردکان، یزد، ایران.

mtazeh@ardakan.ac.ir

دانشیار گروه مهندسی باغبانی، دانشکده کشاورزی و منابع طبیعی، دانشگاه اردکان، اردکان، یزد، ایران.

shirmardi@ardakan.ac.ir

دانشیار گروه مهندسی طبیعت، دانشکده کشاورزی و منابع طبیعی، دانشگاه اردکان، اردکان، یزد، ایران.

skalanteri@ardakan.ac.ir

اطلاعات مقاله	چکیده
مقاله پژوهشی	پارامترهای نسبت جذب سدیم (SAR)، سختی کل (TH) و کربنات سدیم باقیمانده (RSC) آب، نقش عمده ای در فعالیت‌های کشاورزی ایفا می‌کنند، اما اندازه‌گیری آن‌ها نیازمند تجهیزات تخصصی و زمان زیادی است. هدف این تحقیق، پیش‌بینی مقادیر SAR، TH و RSC بر اساس مقادیر اندازه‌گیری شده EC و pH با استفاده از الگوریتم‌های مختلف یادگیری ماشین است. به این منظور، روش‌های مختلف شامل درخت تصمیم، نزدیک‌ترین همسایه، ماشین بردار پشتیبان، کوبیزت، جنگل تصادفی، سوپرلرنینگ و XGBoost بررسی شد. در این مطالعه، داده‌های مربوط به ۱۳۲ نمونه آب که مقادیر EC، pH، SAR، RSC و TH آن‌ها در آزمایشگاه اندازه‌گیری شده بود، استفاده گردید. نتایج نشان داد که استفاده از روش سوپرلرنینگ، دقت پیش‌بینی را در مقایسه با سایر الگوریتم‌ها به‌طور قابل توجهی افزایش می‌دهد. بدین ترتیب که مقادیر R^2 با استفاده هم‌زمان از EC و pH برای پیش‌بینی TH، RSC و SAR به ترتیب به ۰/۹۱، ۰/۸۸ و ۰/۶۸ رسید که بیشتر از سایر روش‌ها بود. استفاده از EC به تنهایی R^2 را به ۰/۸۷ در TH، ۰/۸۶ در RSC و ۰/۷۳ در SAR سوق داد. استفاده از pH به تنهایی نیز موجب کاهش مقادیر R^2 به ۰/۱۲ و ۰/۱۱ و ۰/۹ شد. یافته‌های پژوهش حاضر نشان داد که با استفاده از اندازه‌گیری‌های ساده EC و pH و الگوریتم‌های یادگیری ماشین، به‌ویژه سوپرلرنینگ، می‌توان پیش‌بینی دقیقی از پارامترهای RSC، TH و SAR انجام داد. این روش باعث کاهش هزینه‌ها و تصمیم‌گیری سریع در مدیریت منابع آب خواهد شد.
واژه‌های کلیدی: عملکرد مدل، اعتبارسنجی مدل، کیفیت آب آبیاری	
آدرس ایمیل نویسنده مسئول: mtazeh@ardakan.ac.ir	

استناد: تازه، مهدی، شیرمردی، مصطفی، و کلانتری، سعیده، ۱۴۰۴. برآورد شاخص‌های کیفیت آب آبیاری (SAR، TH و RSC) با استفاده از شاخص-

های EC و pH توسط الگوریتم‌های یادگیری ماشین. نشریه پژوهش آب در کشاورزی، ۳۹ (۳)، صص ۲۱۷-۲۳۶



DOI: <https://doi.org/10.22092/jwra.2025.369886.1086>

مقدمه

در برخی مناطق کشورمان، کشاورزی به عنوان منبع اصلی اشتغال و ستون اقتصاد منطقه است و همچنین حدود ۸۰٪ آب مصرفی در بخش کشاورزی است؛ بنابراین، برنامه‌ریزی و مدیریت آب برای تضمین کشاورزی پایدار، ضروری است. به طور کلی، شاخص کیفیت آب آبیاری (IWQI) بر اساس پارامترهای اتخاذ شده توسط دستورالعمل‌های FAO 29، ارائه می‌شود. منابع آب نقش مهمی در تأمین نیازهای شرب، صنعت و کشاورزی دارند. در واقع، کیفیت خوب منابع آب، هزینه تصفیه آب برای مصارف مختلف را به میزان قابل توجهی کاهش داده و کشاورزی را بهبود می‌بخشد.

تعیین سختی و کربنات سدیم باقیمانده (RSC)، به عنوان پارامترهای کیفی دارای اهمیت ویژه‌ای است. سختی آب در فعالیت‌هایی از قبیل کشاورزی، صنعت و شرب می‌تواند ایجاد مشکلاتی کند. مسدود شدن نازل‌ها در سیستم آبیاری قطره‌ای در بخش کشاورزی، خرابی تجهیزات صنعتی، ایجاد بیماری‌ها و تهدید سلامتی انسان و حیوانات از جمله خطرات سختی آب به شمار می‌رود. اطلاع از میزان سختی آب می‌تواند منجر به حصول راهکارهایی جهت کنترل کیفیت آب در مصارف خانگی، صنعتی و کشاورزی باشد. دو عنصر کلسیم و منیزیم محلول در آب، رایج‌ترین عناصری هستند که باعث سختی آب می‌شوند و با افزایش میزان آن‌ها سختی آب افزایش می‌یابد. سختی عموماً بر اساس میلی گرم کربنات کلسیم معادل در لیتر بیان می‌شود. در کاربری‌های مختلف، سطوح متفاوتی از میزان سختی مدنظر است. از نظر سلامتی و بهداشت، غلظت کلسیم $40-80 \text{ mg.L}^{-1}$ ، منیزیم $20-30 \text{ mg.L}^{-1}$ و سختی کل $2-4 \text{ mmol.L}^{-1}$ گزارش شده است. سختی کمتر از 1 mg.L^{-1} را نرم، $60-171$ را کمی سخت، $120-180$ را سخت و بیش از 180 را خیلی سخت بیان داشته‌اند. هنگام استفاده از آب زیرزمینی برای آبیاری نیز وضعیت RSC بسیار حائز اهمیت است. این مسئله در مناطق خشک و نیمه خشک دارای

اهمیت بیشتری است. چراکه در این مناطق، علاوه بر کمبود منابع آب، کیفیت آن نیز چندان مناسب نیست. بالا بودن RSC در آب آبیاری باعث تجمع نمک در خاک و شور شدن آن، سدیمی شدن خاک و کاهش رشد گیاه می‌شود. اثر سدیمی شدن خاک، بر رشد گیاه، متفاوت از اثر شوری بر گیاه است. شوری با تأثیر بر شیب پتانسیل آب، بین خاک و سلول ریشه باعث کاهش رشد می‌شود، درحالی‌که سدیم زیاد با تأثیر بر خصوصیات فیزیکی و هیدرولوژیکی خاک، رشد گیاه را کاهش می‌دهد. شوری و سدیمی بودن خاک هر دو با ایجاد اختلالات تغذیه‌ای در گیاه، رشد را کاهش می‌دهند. محدوده‌های متفاوتی از نظر RSC در مطالعات وجود دارد. ویلکوکس و همکاران (۱۹۵۴)، آب با RSC بیش از $2/5 \text{ meq.L}^{-1}$ را برای آبیاری مناسب نمی‌دانند. درحالی‌که گوپتا (۱۹۸۳) گزارش کرد که آبیاری با آبی با RSC برابر 10 meq.L^{-1} ، اثری بر عملکرد گندم در یک خاک آهکی با بافت شنی - لومی ندارد.

اندازه‌گیری سختی آب به روش‌های مختلف شیمیایی و دستگاهی صورت می‌گیرد. یکی از این روش‌ها، تیتراسیون کمپلکسومتری با اتیلن دی آمین تتراسیتیک اسید (EDTA) است. در این روش، مجموع کلسیم و منیزیم را با کمک معرف اریوکروم بلک تی، تعیین می‌کنند. تیتراسیون کمپلکسومتری، یک روش آزمایشگاهی است که نیازمند صرف زمان و استفاده از مواد شیمیایی است. روش دیگر تعیین سختی، فتومتر است که با اضافه کردن مقدار مشخصی از تیتراست و اندازه‌گیری میزان جذب در یک طول موج مشخص است. در این روش، آب را بر اساس سختی در چهار گروه قرار می‌دهد. روش دیگر تعیین سختی استفاده از الکتروود انتخابی یونی، جهت تعیین کلسیم و منیزیم است (آفنداس و مارکالو، ۲۰۱۹). علاوه بر سختی، اندازه‌گیری RSC نیز در آب‌های مناطق خشک ضروری است. برای اندازه‌گیری RSC، علاوه بر کلسیم و منیزیم، غلظت کربنات و بیکربنات نیز باید در آب اندازه‌گیری شود. برای اندازه‌گیری کربنات و بیکربنات به ترتیب از معرف فنولفتالئین و متیل اورانژ و تیتراسیون با اسیدی با نرمالیت

مشخص استفاده می‌شود. باین حال، تحقیقات متعددی برای ارزیابی کیفیت آب با استفاده از یک رویکرد مبتنی بر شاخص، انجام شده است. لازم به ذکر است که می‌توان از برخی پارامترهایی که اندازه‌گیری آن‌ها آسان‌تر و ارزان‌تر است، جهت تخمین پارامترهای دیگر استفاده نمود و در حقیقت با استفاده از روابط ریاضی و آماری می‌توان به این هدف دست یافت (تقی زاده و همکاران، ۱۳۹۵).

از طرفی، در سال‌های اخیر، استفاده از یادگیری ماشین به‌طور وسیع در حال گسترش است. هر مدل، دارای نکات مثبت و منفی است و جهت دستیابی به نتیجه بهتر، محققین چندین مدل را با هم مقایسه کرده و بهترین را گزارش می‌کنند. راهکار دیگر، استفاده از همه مدل‌ها و تکنیک سوپرلرنینگ است که دارای کاربردهای وسیعی در علوم آب‌و خاک است. برای مثال، ابراهیم و همکاران (۲۰۲۳)، از مدل‌های یادگیری شبکه عصبی - فازی تطبیقی و ماشین بردار پشتیبان برای پیش‌بینی شاخص‌های کیفیت آب آبیاری و ارزیابی منابع آب زیرزمینی استفاده کردند. نتایج نشان داد که مدل‌های توسعه‌یافته دقت بالایی در پیش‌بینی کیفیت آب دارند. زو (۲۰۱۰)، از روش فازی برای پیش‌بینی خصوصیات خاک بهره گرفت. ماراک اوغلو و کارمن (۲۰۱۰)، برای پیش‌بینی هدررفت خاک از روش فازی استفاده کردند. مردون و همکاران (۲۰۰۶)، برای پیش‌بینی هدایت هیدرولیکی اشباع، از روش شبکه عصبی مصنوعی استفاده کردند. شو و شارما (۲۰۲۵)، به ارزیابی و پیش‌بینی کیفیت آب‌های زیرزمینی برای آبیاری با استفاده از مدل‌های رگرسیون و یادگیری ماشین پرداختند. نتایج نشان داد که درخت تصمیم و شبکه عصبی مصنوعی عملکرد بهتری نسبت به مدل‌های رگرسیون معمولی داشتند. شبکه عصبی مصنوعی، قادر به مدل‌سازی روابط غیرخطی بین ورودی‌ها و خروجی‌هاست و در بسیاری از مطالعات قدیمی برای پیش‌بینی شاخص‌های کیفیت آب استفاده شده است. باین حال، شبکه عصبی مصنوعی، معمولاً نیاز به حجم بالای داده دارد، احتمال برازش بیش‌ازحد وجود دارد و تفسیر مدل برای کاربران

غیرمتخصص دشوار است. روش‌های مدرن یادگیری ماشین و هوش مصنوعی مانند سوپرلرنینگ و جنگل تصادفی، قابلیت ترکیب چند مدل را دارند، دقت پیش‌بینی بالاتری ارائه می‌کنند و نسبت به حجم داده مقاوم‌تر هستند، همچنین امکان تفسیر بهتر و تعیین اهمیت پارامترهای ورودی را فراهم می‌کنند.

سوپرلرنینگ ظرفیت بالایی برای ایجاد پیش‌بینی‌های دقیق دارد. در واقع، هنگامی که مجموعه‌ای از داده‌ها برای مدل‌سازی مورد استفاده قرار می‌گیرند، روش‌های برآوردی متعددی در دسترس هستند. سوپرلرنینگ با ترکیب نقاط قوت این روش‌ها موجب افزایش دقت پیش‌بینی می‌شود. برای مثال، پالی و همکاران (۲۰۱۹) با استفاده از داده‌های ۵۱۱ حوضه و الگوریتم‌های مختلف نشان دادند که سوپرلرنینگ عملکرد رگرسیون خطی را تا ۲۰٪ بهبود بخشید.

در پژوهش حاضر، هفت الگوریتم یادگیری ماشین مورد بررسی قرار گرفتند. هدف مطالعه، برآورد شاخص‌های TH، RSC و SAR با استفاده از دو پارامتر ساده‌تر یعنی EC و pH از طریق روش‌های داده‌کاوی بود. الگوریتم‌های مورد استفاده شامل درخت تصمیم، نزدیک‌ترین همسایه، ماشین بردار پشتیبان، کوبیزت، جنگل تصادفی، سوپرلرنینگ و XGBoost بودند؛ به عبارت دیگر، هدف اصلی این پژوهش بررسی قابلیت استفاده از مقادیر EC و pH برای تخمین پارامترهای TH، SAR و RSC و تعیین کارآمدترین الگوریتم یادگیری ماشین در این زمینه است.

روش بررسی

منطقه مورد مطالعه

محل نمونه‌برداری پژوهش حاضر، استان یزد در مرکز ایران است. این استان بین عرض‌های جغرافیایی ۳۵° ۲۹' تا ۳۲° ۲۲' شمالی و طول‌های ۴۸' ۵۲° تا ۳۹' ۵۶° شرقی واقع شده است. اقلیم منطقه خشک و بیابانی است و میانگین بارندگی سالانه کمتر از ۱۰۰ mm و میانگین دمای

سالانه 20°C است. بیشترین بارش‌ها در فصول زمستان و اوایل بهار رخ می‌دهد. از نظر زمین‌شناسی، استان یزد شامل سازندهای آبرفتی، سنگ‌های آهکی، گچی، شیل و رسوبات تبخیری است که از دوران پرکامبرین تا دوره‌های زمین‌شناسی جدید در منطقه گسترده شده‌اند. نمونه‌برداری پژوهش حاضر از آب چاه‌های کشاورزی در فصل تابستان انجام شده است.

نمونه‌برداری آب و انتخاب متغیرهای مدل

در پژوهش حاضر، از ۱۳۲ نمونه آب برداشت‌شده از چاه‌های کشاورزی استفاده شد، بطوریکه این تعداد نمونه آب با هدف دستیابی به پوشش مکانی مناسب و با توجه به تنوع زمین‌شناسی و شرایط کیفی منطقه تعیین شد. موقعیت چاه‌های کشاورزی به صورت میدانی شناسایی و نمونه‌برداری طوری انجام گرفت تا از پوشش یکنواخت در کل محدوده مطالعه، اطمینان حاصل شود. این تعداد نمونه به گونه‌ای انتخاب گردید که هم تنوع شرایط هیدروژئوشیمیایی منطقه را در برگیرد و هم حجم کافی داده، برای آموزش و ارزیابی مدل‌های یادگیری ماشین فراهم سازد. نمونه‌ها پس از برداشت به آزمایشگاه منتقل شده و پارامترهای شیمیایی شامل EC و pH در هر نمونه اندازه‌گیری گردید. دو پارامتر EC و pH به عنوان ورودی‌های اصلی مدل‌های یادگیری ماشین انتخاب شدند، زیرا این پارامترها قابلیت اندازه‌گیری سریع و آسان داشته و نقش مهمی در تعیین کیفیت آب و پارامترهای مربوط به سختی کل (TH)، شوری نسبی (RSC) و نسبت جذب سدیم (SAR) ایفا می‌کنند. اگرچه در این مطالعه امکان اندازه‌گیری مستقیم یون‌های محلول مهم شامل Ca^{2+} ، Mg^{2+} ، Na^{+} و HCO_3^{-} و سایر یون‌ها وجود داشت، اما هدف اصلی پژوهش، بررسی قابلیت پیش‌بینی شاخص‌های کیفیت آب (TH، SAR، RSC) با استفاده از پارامترهای پایه و دارای قابلیت اندازه‌گیری سریع مانند EC و pH بود. این رویکرد از چند جنبه حائز اهمیت است؛ نخست، اندازه‌گیری EC و pH نسبت به تعیین دقیق غلظت تک‌تک یون‌ها بسیار سریع‌تر و کم‌هزینه‌تر است. دوم، این روش

امکان پایش میدانی و مکرر کیفیت آب را فراهم می‌کند که در شرایط کشاورزی و نظارت‌های مستمر کاربردی است و سوم، توسعه مدل‌های یادگیری ماشین برای تخمین این شاخص‌ها، امکان ساخت ابزارهای ساده‌تر و قابل استفاده در محیط‌های غیرآزمایشگاهی را فراهم می‌سازد. در نتیجه، هدف از تخمین غیرمستقیم با مدل‌های یادگیری ماشین، ارائه روشی کاربردی و اقتصادی برای ارزیابی کیفیت آب است که مکمل اندازه‌گیری‌های مستقیم و دقیق است. از طرفی EC شاخص کلی، از مجموع املاح محلول در آب است و یون‌های مؤثر در محاسبه شاخص‌های TH، SAR و RSC را شامل می‌شود؛ اما ارتباط بین EC و این شاخص‌ها صرفاً رابطه‌ای ساده و خطی نیست. ترکیب یونی، تعادل‌های شیمیایی و شرایط محیطی باعث ایجاد روابط پیچیده و غیرخطی بین این پارامترها می‌شوند. مدل‌های یادگیری ماشین به دلیل قابلیت شناسایی الگوهای پیچیده و غیرخطی، قادرند از داده‌های EC و pH به عنوان ورودی استفاده کرده و روابط غیرمستقیم میان پارامترها را استخراج کنند. این رویکرد به بهبود دقت پیش‌بینی شاخص‌های کیفیت آب کمک می‌کند، چراکه این مدل‌ها فراتر از روابط خطی ساده عمل می‌کنند و می‌توانند دانش نهفته در داده‌ها را به طور کامل‌تری استخراج نمایند. از سوی دیگر، pH به عنوان پارامتری مستقل که مستقیماً به غلظت یون هیدروژن مرتبط است، ارتباط فیزیکوشیمیایی مستقیمی با یون‌های کلسیم، منیزیم یا سدیم ندارد. با این حال، pH یکی از پارامترهای کلیدی در بررسی کیفیت آب است که می‌تواند به صورت غیرمستقیم از طریق تعادل‌های شیمیایی آب (مانند تعادل کربنات‌ها و دی‌اکسیدکربن محلول) تحت تأثیر یون‌های محلول قرار گیرد؛ بنابراین، هدف از استفاده از pH در مدل‌سازی، تحلیل آماری و بررسی دقت پیش‌بینی در شرایط میدانی است. جهت برآورد شاخص‌های TH، SAR و RSC می‌توان از مدل‌های مختلفی استفاده کرد که دامنه‌ای از رگرسیون خطی ساده تا روش‌های پیچیده داده‌کاوی مانند شبکه‌های عصبی مصنوعی را شامل می‌شود. هر یک از این مدل‌ها دارای نقاط قوت و ضعف خاصی

محاسبه RSC

پس از اندازه‌گیری غلظت آنیون‌های کربنات و بیکربنات و کاتیون‌های کلسیم و منیزیم، RSC با استفاده از معادله (۱)، بر حسب meq.L^{-1} به دست می‌آید.

$$\text{RSC (meq. L}^{-1}\text{)} = (\text{HCO}_3^{-1} + \text{CO}_3^{-2}) - (\text{Ca}^{+2} + \text{Mg}^{+2}) \quad (1)$$

غلظت تمام کاتیون‌ها و آنیون‌ها در فرمول بالا بر حسب meq.L^{-1} است.

محاسبه سختی

سختی آب بر حسب میلی‌گرم کربنات کلسیم معادل در لیتر محاسبه می‌شود. سختی کل مجموع سختی کلسیم و منیزیم است. سختی کلسیم و منیزیم، به غلظت یون‌های کلسیم و منیزیم گفته می‌شود که به‌صورت معادل کربنات کلسیم (CaCO_3) بیان می‌شود (کاگیلاری و همکاران، ۲۰۱۱). سختی کل آب، با استفاده از رابطه (۲) قابل محاسبه است.

$$\begin{aligned} \text{Total water hardness (equivalent of CaCO}_3\text{)} \left(\frac{\text{mg}}{\text{L}}\right) \\ = \left[\text{Ca}^{+2} \left(\frac{\text{meq}}{\text{L}}\right) \times 20 \times 2.5 \right] \\ + \left[\text{Mg}^{+2} \left(\frac{\text{meq}}{\text{L}}\right) \times 12 \times 4.1 \right] \end{aligned} \quad (2)$$

نسبت جذب سدیم

نسبت جذب سدیم، به‌عنوان شاخصی مؤثر برای بررسی خطر بالقوه سدیم در محلول متعادل شناخته می‌شود. فرمول محاسبه SAR به‌صورت رابطه (۳) تعریف می‌شود:

$$\text{SAR} = \frac{\text{Na}^+}{\sqrt{\frac{1}{2}(\text{Ca}^{2+} + \text{Mg}^{2+})}} \quad (3)$$

در فرمول بالا، غلظت یون‌های کلسیم، منیزیم و سدیم بر حسب میلی‌اکی‌والان در لیتر بیان می‌شود. در این زمینه، نسبت جذب سدیم نشان‌دهنده احتمال ایجاد پدیده تجمع یا پراکندگی ذرات رس است. یون‌های سدیم و پتاسیم موجب پراکندگی ذرات رس می‌شوند، درحالی‌که یون‌های کلسیم و منیزیم باعث تجمع و ایجاد ساختار پایدار در خاک می‌گردند.

هستند و بخشی از دانش نهفته در داده‌ها را استخراج می‌کنند. با استفاده از تکنیک‌های ترکیبی، می‌توان از قدرت مجموع مدل‌ها بهره گرفت. برای مثال، جنگل تصادفی از تعداد زیادی مدل درختی استفاده می‌کند و نتایج را میانگین‌گیری می‌کند، یا در مدل تقویت گرادیان هر مدل درختی ساخته‌شده در مرحله بعد اصلاح می‌شود تا بهترین نتیجه حاصل گردد. یک روش دیگر ترکیب کردن مدل‌ها این است که از یک مدل دیگر در سطح بالاتر استفاده می‌شود. روش‌های پیشرفته‌تر مانند سوپر لرنینگ نیز به‌کار گرفته می‌شوند که در این تحقیق مورد استفاده قرار گرفته است.

پس از نمونه‌برداری، نمونه‌های آب جهت انجام آنالیزهای شیمیایی به آزمایشگاه منتقل و پارامترهای زیر در هر نمونه اندازه‌گیری گردید.

pH و EC

قابلیت هدایت الکتریکی آب با استفاده از هدایت‌سنج مدل JENWAY-4320 اندازه‌گیری شد. pH آب نیز با دستگاه pH متر مدل METROHM-620 اندازه‌گیری شد.

کربنات و بیکربنات

اندازه‌گیری کربنات با استفاده از معرف فنولفتالین و تیتراسیون با 0.05 N HCl انجام شد. برای اندازه‌گیری بیکربنات از معرف متیل اورانژ استفاده شد و تیتراسیون با 0.05 N HCl صورت گرفت (میناسنی و همکاران، ۲۰۱۱).

کلسیم و منیزیم

غلظت کلسیم و منیزیم، به روش کمپلکسومتری و تیتراسیون با EDTA اندازه‌گیری شد. برای اندازه‌گیری مجموع کلسیم و منیزیم از معرف اریوکروم بلک تی استفاده شد. برای اندازه‌گیری کلسیم از معرف کالکون استفاده و در نهایت از تفاضل آن‌ها غلظت منیزیم به‌دست آمد (کار و همکاران، ۲۰۱۵).

به نام گره ریشه است. درختان تصمیم‌گیری می‌توانند داده-های دسته‌ای و عددی را کنترل کنند (فتحی‌زاد و همکاران، ۱۳۹۴).

نزدیک‌ترین همسایه K

نزدیک‌ترین همسایه K یک روش یادگیری موردی است و از جمله ساده‌ترین الگوریتم‌های یادگیری ماشین است. در این الگوریتم، یک نمونه با رأی اکثریت از همسایه‌هایش دسته‌بندی می‌شود و این نمونه در عمومی‌ترین کلاس مابین k همسایه نزدیک تعیین می‌شود. K یک مقدار مثبت صحیح و عموماً کوچک است. اگر $k=1$ باشد نمونه به‌سادگی در کلاس همسایگان نزدیکش تعیین می‌گردد. فرد بودن مقدار k مفید است، چون با این کار جلوی آراء برابر گرفته می‌شود. این روش، برای بسیاری موارد کاربرد دارد، زیرا اثربخش، غیرپارامتریک و دارای پیاده‌سازی راحت است. با این حال، زمان دسته‌بندی‌اش طولانی است و یافتن مقدار k بهینه مشکل است. بهترین انتخاب از k، وابسته به داده‌ها است. به‌طورکلی، مقدار بزرگ از k اثر نویز روی دسته‌بندی را کاهش می‌دهد؛ اما مرز مابین کلاس‌ها کمتر متمایز می‌شود (دکا و همکاران، ۲۰۱۴).

ماشین بردار پشتیبان

الگوریتم ماشین بردار پشتیبان اولیه در سال ۱۹۶۳ توسط ولادیمیر وپنیک ابداع شد و در سال ۱۹۹۵ توسط وپنیک و کورینا برای حالت غیرخطی تعمیم داده شد. مبنای کاری دسته‌بندی‌کننده آن، دسته‌بندی خطی داده‌ها است و در تقسیم خطی داده‌ها سعی می‌شود خطی انتخاب شود که حاشیه اطمینان بیشتری داشته باشد. حل معادله پیدا کردن خط بهینه برای داده‌ها به‌وسیله روش‌های برنامه‌نویسی درجه دوم که روش‌های شناخته‌شده‌ای در حل مسائل محدودیت‌دار هستند صورت می‌گیرد. ماشین بردار پشتیبان از ترفند کرنل برای تبدیل داده‌ها استفاده می‌کند و سپس بر اساس این تبدیل، مرز بهینه بین خروجی‌های ممکن را پیدا می‌کند. به عبارت

نسبت جذب سدیم (SAR) عمدتاً به‌عنوان معیاری جهت سنجش و ارزیابی کیفیت آب در کاربردهای آبیاری به کار می‌رود. به‌طورکلی، هرچه مقدار SAR بالاتر باشد، آب برای آبیاری نامناسب‌تر است. استفاده طولانی‌مدت از آب آبیاری با SAR بالا ممکن است نیاز به اقدامات اصلاحی در خاک را ایجاد کند. وقتی چنین آبی به مدت چند سال استفاده شود، یون‌های سدیم می‌توانند جایگزین یون‌های کلسیم و منیزیم در خاک شوند. این تبادل یونی باعث کاهش توانایی خاک در تشکیل ذرات پایدار و در نتیجه از بین رفتن ساختار خاک می‌شود. در نتیجه، نفوذپذیری خاک نسبت به آب کاهش می‌یابد و این مسئله می‌تواند منجر به بروز مشکلاتی در تولید محصول شود. اگرچه SAR عمدتاً تأثیر سدیم بر پایداری ساختار خاک را نشان می‌دهد، اما غلظت‌های بالای پتاسیم و منیزیم نیز می‌توانند تأثیر منفی بر نفوذپذیری خاک داشته باشند. تأثیر پتاسیم نیز در برخی موارد می‌تواند از طریق اصلاح محاسبه SAR لحاظ شود.

الگوریتم‌های مورد استفاده

الگوریتم‌های داده‌کاوی کاربردهای گسترده‌ای در شاخه‌های مختلف علوم، از جمله علوم آب‌و‌خاک دارند. در این پژوهش، به بررسی و مقایسه چند الگوریتم داده‌کاوی و ارزیابی توانایی آن‌ها در برآورد مقادیر شاخص‌های TH، RSC و SAR بر اساس داده‌های اندازه‌گیری‌شده EC و pH پرداخته شده است. الگوریتم‌های مورد استفاده عبارت‌اند از:

الگوریتم درخت تصمیم

الگوریتم درخت تصمیم، مدل‌های طبقه‌بندی یا رگرسیون را به شکل ساختار درخت می‌سازد. درخت تصمیم، مجموعه داده را به زیرمجموعه‌های کوچک‌تر و کوچک‌تر تجزیه می‌کند و یک درخت تصمیم مرتبط به‌صورت تدریجی توسعه می‌یابد. نتیجه نهایی یک درخت با گره‌های تصمیم‌گیری و گره‌های برگ است. بالاترین گره تصمیم‌گیری در یک درخت، مطابق با بهترین پیش‌بینی‌کننده

XGBoost

الگوریتم XGBoost (Extreme Gradient Boosting)

نوعی پیاده‌سازی کارآمد از درختان تصمیم مبتنی بر تقویت گرادیان است که با سرعت بالاتر و عملکرد بهتری نسبت به روش‌های مشابه عمل می‌کند. در این روش، مدل‌های جدید (در اینجا درختان تصمیم) به‌منظور پیش‌بینی خطاهای مدل‌های قبلی ایجاد می‌شوند و مدل نهایی حاصل جمع تمامی مدل‌های برازش‌یافته است. برای افزودن درختان تصمیم جدید، الگوریتم نزول گرادیان به‌منظور کمینه‌سازی تابع زیان به‌کار گرفته می‌شود. افزون بر این، این مدل با استفاده از یک چارچوب منظم‌سازی شده طراحی شده است که امکان کنترل پدیده بیش‌برازش را فراهم می‌کند (پاپاکارالامپوس و همکاران، ۲۰۱۹).

سوپر لرنینگ

سوپر لرنینگ، یک روش یادگیری مبتنی بر تابع زیان کلی است و روشی برای پیش‌بینی است که به دنبال یافتن ترکیب بهینه‌ای از مجموعه‌ای از الگوریتم‌های پیش‌بینی است. الگوریتم سوپر لرنینگ ترکیبی از الگوریتم‌ها را پیدا می‌کند که ریسک اعتبارسنجی متقابل را به حداقل می‌رساند. چارچوب این روش، بر پایه نظریه اعتبارسنجی متقابل ساخته شده و امکان در نظر گرفتن کلاس کلی از الگوریتم‌های پیش‌بینی برای مجموعه را فراهم می‌کند (سیم و همکاران، ۲۰۱۴).

مدل‌سازی و اعتبارسنجی متقابل

برای آموزش مدل‌ها، داده‌های نمونه به دو بخش تقسیم شدند: بخش آموزش شامل ۷۰٪ از داده‌ها بود و در بخش اعتبارسنجی ۳۰٪ باقی‌مانده برای ارزیابی عملکرد مدل استفاده شد. نتایج ارائه‌شده در جداول ۲ تا ۴ مربوط به عملکرد مدل‌ها در مراحل اعتبارسنجی و واسنجی است. اعتبارسنجی متقابل، یک تکنیک بسیار مفید برای ارزیابی کارایی مدل‌های یادگیری ماشین است. این روش، کمک می‌کند تا مدل یادگیری ماشین که ایجاد شده است، به یک مجموعه داده مستقل تعمیم داده شود. در حقیقت، می‌توان

ساده، تبدیلات بسیار پیچیده را انجام می‌دهد. سپس مشخص می‌کند چگونه داده‌ها را بر اساس برجسب‌ها یا خروجی‌هایی که تعریف شده‌اند، جدا نمود (فتحی‌زاد و همکاران، ۲۰۱۷).

کوبیزت

مدل کوبیزت بر اساس مدل درخت M5 کوینلان پیشنهاد شده است. در مدل رگرسیون مفهومی کوبیزت، درخت رشد می‌کند و در برگ‌های انتهایی یک مدل رگرسیون خطی برای پیش‌بینی قرار دارد. روش کوبیزت مجموعه‌ای از قواعد «اگر-آنگاه-بعدی» ایجاد می‌کند که هر قاعده به یک مدل رگرسیون خطی چندمتغیره مرتبط است. تا زمانی که مجموعه‌ای متغیرهای مستقل شرایط قاعده را برآورده کنند، مدل مربوطه برای محاسبه مقدار پیش‌بینی شده استفاده می‌شود. مزیت اصلی روش کوبیزت در افزودن چندین کمیته آموزشی و «تقویت» وزن‌ها برای متعادل‌تر کردن آن‌ها است. به این ترتیب، مجموعه‌ای از درختان تولید می‌شود تا مدل کوبیزت شکل گیرد (ساگی و روکاج، ۲۰۱۸).

جنگل تصادفی

جنگل تصادفی یک روش غیرپارامتری و مبتنی بر مجموعه‌ای از درختان تصمیم است. برخلاف روش‌های آماری سنتی که از مدل‌های پارامتری استفاده می‌کنند، جنگل‌های تصادفی شامل تعداد زیادی مدل درخت تصمیم ساده و قابل تفسیر هستند. با ترکیب نتایج تحلیل این درختان تصمیم، می‌توان یک مدل پیش‌بینی جامع‌تر به دست آورد. رگرسیون جنگل تصادفی یک روش رگرسیون غیرپارامتری است که از مجموعه‌ای شامل K درخت تشکیل شده است. ورودی این روش یک بردار چندبعدی است که تشکیل‌دهنده یک جنگل است. این مجموعه، خروجی‌هایی به تعداد درختان تولید می‌کند و خروجی نهایی با میانگین‌گیری از پیش‌بینی‌های همه درختان به دست می‌آید (تیراليس و همکاران، ۲۰۱۹).

تبیین تغییرات و مقدار نزدیک به یک بیانگر برازش کامل مدل است.

میانگین قدر مطلق خطا (MAE) و ریشه میانگین مربعات خطا (RMSE)

برای سنجش دقت پیش‌بینی، از دو معیار میانگین قدر مطلق خطا (MAE) و ریشه میانگین مربعات خطا (RMSE) استفاده شد. این شاخص‌ها میزان اختلاف بین مقادیر مشاهده‌شده و مقادیر پیش‌بینی‌شده را کمی می‌کنند؛ به‌گونه‌ای که مقادیر کوچک‌تر نشان‌دهنده دقت بیشتر مدل هستند.

ضریب همبستگی تطابق

ضریب همبستگی تطابق، میزان توافق بین مقادیر مشاهده‌شده (X) و مقادیر پیش‌بینی‌شده (Y) را ارزیابی می‌کند. این آماره، علاوه بر قدرت همبستگی، انحراف از خط برازش کامل را نیز در نظر می‌گیرد و معیاری دقیق‌تر نسبت به ضریب همبستگی ساده ارائه می‌دهد. مراحل پژوهش حاضر در شکل ۱ نشان داده شده است.

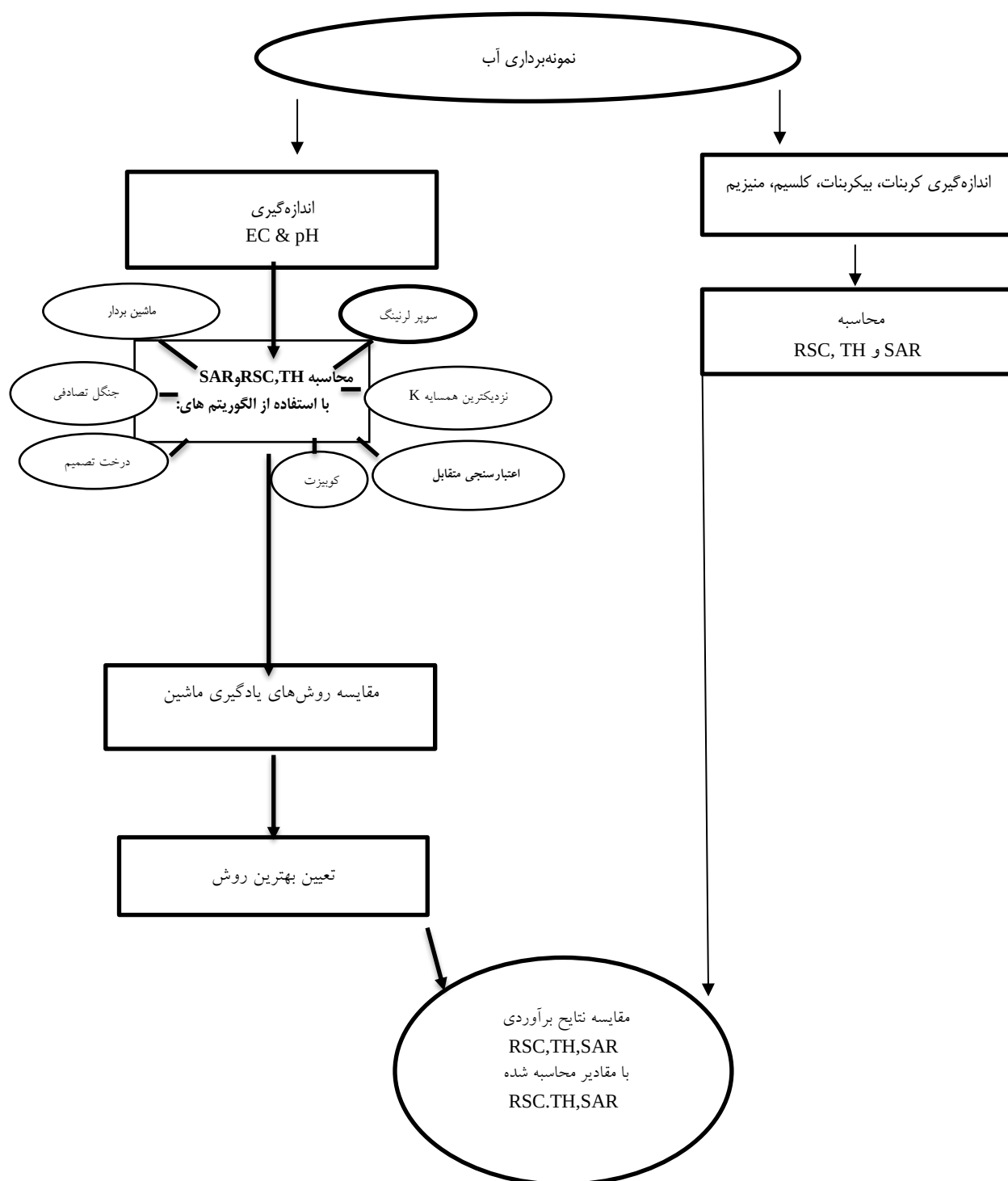
از این تکنیک برای پیش‌بینی و تخمین دقت و کارایی مدل استفاده کرد. در مسائل یادگیری ماشین، معمولاً با دو مجموعه داده در ارتباط است که داده‌های شناخته‌شده و داده‌های ناشناخته می‌باشند. با استفاده از ارزیابی متقابل، می‌توان مدل یادگیری ماشین را در فاز آموزش برای آزمون کارایی و به‌دست آوردن یک ایده و نظر از چگونگی تعمیم مدل یادگیری ماشین به داده‌های مستقل تست نمود (آفندراس و همکاران، ۲۰۱۹).

معیارهای عملکرد مدل

برای مقایسه کارایی مدل‌ها، از معیارهای آماری زیر استفاده شد:

ضریب تعیین (R^2)

ضریب تعیین (R^2) نشان‌دهنده درصد واریانس متغیر وابسته است که توسط متغیر مستقل تعیین می‌شود؛ به عبارت دیگر، با محاسبه این ضریب می‌توان گفت که متغیر مستقل چه درصدی از واریانس کل متغیر Y را توضیح می‌دهد. مقدار عددی این ضریب بین صفر تا یک متغیر است. مقدار نزدیک به صفر بیانگر عدم توانایی مدل در



شکل ۱- نمودار انجام کار

یافته‌ها

از آنجاکه هدف پژوهش حاضر، امکان‌سنجی استفاده از یادگیری ماشین در دستیابی به روش دقیق و آسان برای برآورد مقادیر RSC, SAR, TH است، لذا در مرحله اول مقادیر EC و pH به صورت مجزا و مرحله‌ای و در

مرحله دوم مقادیر EC و pH هم‌زمان به عنوان ورودی در مدل‌های مختلف در برآورد مقادیر SAR, TH, SCH مورد استفاده قرار گرفت. پس از مقایسه روش‌های مختلف و انتخاب روش بهینه، اقدام به برآورد مقادیر مذکور بر اساس ورودی‌های EC و pH گردید. جدول ۱، خلاصه آماری مقادیر اندازه‌گیری را نشان می‌دهد.

جدول ۱- خلاصه آماری نمونه‌ها

parameter	Min	Max	Mean	Std. deviation	Variance	Skewness	Kurtosis
EC ($\mu\text{S}/\text{cm}$)	44.99	2902.1	525.54	619.12	383321.30	2.28	4.71
pH	6.32	10.00	8.16	1.06	1.12	0.046	1.80
RSC (meq/l)	-55.15	34.99	-5.52	12.51	156.53	-1.88	5.85
TH (mg/L as CaCO_3)	0	1902.1	517.42	615.32	378629	2.30	4.85
SAR	0.22	603.71	7.28	26.32	692.96	10.98	125.58

به منظور بررسی قابلیت پیش‌بینی پارامترهای مورد نظر با استفاده از الگوریتم‌های مختلف، فرایند آموزش، اعتبارسنجی و ارزیابی دقت آن‌ها انجام شد. عملکرد مدل‌های پیش‌بینی با بهره‌گیری از چهار معیار آماری شامل ضریب تعیین (R^2)، ریشه میانگین مربعات خطا (RMSE)، میانگین قدر مطلق خطا (MAE) و ضریب همبستگی تطبیقی (CCC) مورد سنجش قرار گرفت. نتایج حاصل از آموزش مدل‌ها برای پیش‌بینی پارامترهای SAR، TH و RSC به عنوان شاخص‌های اصلی مطالعه، در جداول ۲ تا ۴ ارائه شده است. در ادامه، نتایج واسنجی و اعتبارسنجی مدل‌ها برای پیش‌بینی مقادیر SAR، TH و RSC ارائه شده است.

برآورد SAR با استفاده از pH, EC، PH + EC

همزمان

همان‌گونه که در جدول ۲ مشاهده می‌شود، در برآورد SAR بر اساس EC، مدل سوپر لرنینگ با ضریب

تعیین ($R^2=0.73$) و مقادیر پایین تر خطا MAE و RMSE بالاترین قابلیت پیش‌بینی را نشان داد. پس از آن، مدل جنگل تصادفی با $R^2=0.67$ در رتبه دوم قرار گرفت که حاکی از پایداری نسبی آن در مقایسه با سایر مدل‌هاست. در بخش دوم، مقادیر pH به عنوان ورودی مدل‌های مختلف استفاده شد. در این مورد نیز بالاترین R^2 متعلق به مدل سوپر لرنینگ بود، این امر بیانگر برتری نسبی این مدل در استخراج الگوهای غیرخطی و تخمین SAR بر مبنای pH است.

در نهایت، در حالتی که EC و pH به صورت هم‌زمان به عنوان ورودی مورد استفاده قرار گرفتند، نتایج جدول ۲ نشان داد که مدل سوپر لرنینگ با $R^2=0.76$ و شاخص توافق (CCC) بالاترین دقت و کارایی را داشته است. پس از آن، مدل جنگل تصادفی با $R^2=0.69$ در جایگاه دوم قرار گرفت که نشان‌دهنده توان پیش‌بینی قابل قبول آن در شرایط ترکیب متغیرها است.

جدول ۲- عملکرد مدلها در برآورد SAR با استفاده از pH، EC، SAR و PH همزمان

models	مرحله	SAR											
		EC				pH				PH و EC			
		MAE	RMSE	R2	CCC	MAE	RMSE	R2	CCC	MAE	RMSE	R ²	CCC
Decision Tree model	واسنجی	2.55	3.85	0.72	0.75	17.2	19.5	0.09	0.10	2.10	3.05	0.72	0.75
	اعتبارسنجی	2.85	4.13	0.66	0.68	18.85	23.50	0.06	0.06	2.45	3.43	0.68	0.71
K-nearest Neighbors	واسنجی	2.40	3.70	0.62	0.68	17	18	0.04	0.04	2.20	3.10	0.63	0.69
	اعتبارسنجی	2.79	4.00	0.57	0.65	17.88	18.53	0.03	0.03	2.29	3.50	0.61	0.67
Support vector regression	واسنجی	2.75	3.30	0.68	0.70	18	20.0	0.08	0.07	2.18	3.0	0.68	0.70
	اعتبارسنجی	3.02	3.65	0.64	0.65	19.85	24.50	0.06	0.01	2.32	3.25	0.66	0.69
Cubist	واسنجی	2.60	3.10	0.70	0.74	16.0	19.0	0.05	0.04	2.05	2.90	0.7	0.73
	اعتبارسنجی	2.96	3.45	0.64	0.69	16.85	20.50	0.06	0.01	2.16	3.05	0.68	0.72
Random forest	واسنجی	2.50	3.60	0.71	0.76	17.5	20.0	0.05	0.06	2.0	2.95	0.7	0.75
	اعتبارسنجی	2.87	3.98	0.67	0.71	18.85	22.50	0.06	0.05	2.07	3.18	0.69	0.75
Extreme Gradient Boosting	واسنجی	2.45	3.20	0.69	0.74	16.8	19	0.04	0.05	2.05	3.00	0.68	0.74
	اعتبارسنجی	2.85	3.50	0.63	0.69	16.88	20.53	0.03	0.02	2.15	3.20	0.67	0.73
Super Learning	واسنجی	2.10	3.05	0.78	0.82	11.00	13	0.01	0.12	1.90	2.80	0.77	0.81
	اعتبارسنجی	2.51	3.38	0.73	0.79	11.82	14.47	0.09	0.11	2.01	2.88	0.76	0.83

داد. در نهایت، زمانی که EC و pH به صورت همزمان به عنوان ورودی استفاده شدند، نتایج نشان داد که مدل سوپر لرنینگ همچنان بهترین گزینه بوده و با $R^2=0.11$ و مقادیر نسبتاً مناسب CCC برتری نسبی خود را نسبت به سایر مدلها حفظ کرده است. با این حال، باید تأکید کرد که در این سناریو نیز مقادیر R^2 به طور کلی کم بوده و نشاندهنده کارایی محدود متغیرهای ترکیبی EC و pH در پیش‌بینی دقیق RSC است.

برآورد TH با استفاده از pH، EC، SAR و PH همزمان

همان‌گونه که در جدول ۴ مشاهده می‌شود، در برآورد سختی کل (TH) بر اساس مقادیر EC، مدل سوپر لرنینگ با ضریب تعیین (R^2) برابر با ۰/۸۷ در مرحله اعتبارسنجی، عملکرد برتر را نسبت به سایر الگوریتم‌ها نشان داد. در برآورد TH با استفاده از pH نیز اگرچه مقدار

برآورد RSC با استفاده از pH، EC، SAR و PH همزمان

در این بخش، ابتدا مقادیر EC به تنهایی برای تخمین RSC به کار گرفته شد. بر اساس جدول ۳، مدل سوپر لرنینگ با ضریب تعیین ($R^2=0.86$) و مقادیر کمتر MAE و RMSE بهترین عملکرد را در این برآورد نشان داد. این موضوع بیانگر توان بالای این مدل در شناسایی روابط پیچیده میان هدایت الکتریکی و شاخص RSC است.

در گام بعدی، تنها مقادیر pH به عنوان ورودی مدلها استفاده شدند. نتایج نشان داد که در این حالت، مقادیر R^2 برای تمامی الگوریتم‌ها به طور کلی کم بوده و دقت مدلها قابل توجه نیست. با این حال، سوپر لرنینگ نسبت به سایر مدلها در مرحله اعتبارسنجی عملکرد بهتری داشت و بیشترین مقدار $R^2=0/۱۱$ را به خود اختصاص

R^2 در مقایسه با سایر ورودی‌ها کمتر است، اما مدل سوپر لرنینگ همچنان بهترین عملکرد را ارائه داد. در حالی که هر دو متغیر EC و pH به طور همزمان به عنوان ورودی مدل‌ها مورد استفاده قرار گرفتند، مدل سوپر لرنینگ با دستیابی به مقدار R^2 برابر با ۰/۱۲ در مرحله اعتبارسنجی، بیشترین دقت را در میان الگوریتم‌ها داشت؛ هرچند این مقدار نسبتاً کم بوده و بیانگر محدودیت این ورودی‌ها در برآورد سختی کل است.

جدول ۳- عملکرد مدل‌ها در برآورد RSC با استفاده از EC، PH، EC و PH همزمان

models	مرحله	RSC											
		EC				pH				EC و PH			
		MAE	RMSE	R2	CCC	MAE	RMSE	R2	CCC	MAE	RMSE	R2	CCC
Decision tree model	واسنجی	4.10	6.60	0.70	0.73	12	16.5	0.12	0.015	11.90	16.40	0.12	0.15
	اعتبارسنجی	4.75	7.45	0.61	0.64	13.63	18.21	0.06	0.08	13.63	18.21	0.06	0.08
K-nearest Neighbors	واسنجی	3.90	6.20	0.74	0.78	12	19	0.11	0.10	12.10	18.9	0.12	0.10
	اعتبارسنجی	4.41	7.57	0.67	0.74	13.62	22.20	0.07	0.04	13.62	22.20	0.07	0.04
Support vector regression	واسنجی	4.20	6.30	0.74	0.76	11	18	0.12	0.07	11.20	17.8	0.12	0.08
	اعتبارسنجی	4.71	6.99	0.67	0.72	12.62	20.20	0.07	0.02	12.62	20.20	0.07	0.02
Cubist	واسنجی	3.7	6.1	0.82	0.86	13.10	14.70	0.12	0.08	13.2	14.6	0.12	0.08
	اعتبارسنجی	4.10	6.70	0.76	0.82	14.62	16.20	0.07	0.03	14.62	16.20	0.07	0.03
Random forest	واسنجی	3.5	6	0.83	0.87	11.4	17	0.12	0.09	11.3	17.1	0.12	0.1
	اعتبارسنجی	3.99	6.79	0.76	0.82	12.62	19.20	0.07	0.06	12.62	19.20	0.07	0.06
Extreme Gradient Boosting	واسنجی	3.6	5.3	0.83	0.86	9.5	18	0.12	0.1	9.6	18.10	0.12	0.1
	اعتبارسنجی	4.03	6.06	0.76	0.81	10.63	20.21	0.06	0.06	10.63	20.21	0.06	0.06
Super Learning	واسنجی	3.2	4.4	0.9	0.92	7.2	10.30	0.2	0.18	7.3	10.4	0.2	0.18
	اعتبارسنجی	3.86	5.09	0.84	0.86	8.58	12.16	0.11	0.12	8.58	12.16	0.11	0.12

جدول ۴- عملکرد مدل‌ها در برآورد TH با استفاده از EC، PH، EC و PH همزمان

Models	مرحله	TH											
		EC				pH				PH و EC			
		MAE	RMSE	R ²	CCC	MAE	RMSE	R ²	CCC	MAE	RMSE	R ²	CCC
Decision tree model	واسنجی	190.25	295.6	0.75	0.77	500.12	720.4	0.08	0.1	500.12	720.4	0.08	0.1
	اعتبارسنجی	202.88	310.73	0.70	0.73	518.16	748.51	0.06	0.08	518.16	748.51	0.06	0.08
K-nearest Neighbors	واسنجی	165.4	280.35	0.77	0.82	510	720	0.08	0.08	510	720	0.08	0.08
	اعتبارسنجی	176.10	296.72	0.73	0.80	531.18	739.53	0.06	0.06	531.18	739.53	0.06	0.06
Support vector regression	واسنجی	158.2	260.45	0.8	0.83	490	710	0.11	0.07	490	710	0.11	0.07
	اعتبارسنجی	169.42	271.08	0.76	0.80	510.19	735.54	0.09	0.05	510.19	735.54	0.09	0.05
Cubist	واسنجی	150.1	245.3	0.82	0.85	530	690	0.09	0.09	530	690	0.09	0.09
	اعتبارسنجی	162.22	257.16	0.79	0.81	550.17	715.52	0.07	0.07	550.17	715.52	0.07	0.07
Random forest	واسنجی	145.8	238.2	0.84	0.86	520	685	0.08	0.05	520	685	0.08	0.05
	اعتبارسنجی	159.88	250.07	0.81	0.82	547.21	713.56	0.06	0.03	547.21	713.56	0.06	0.03
Extreme Gradient Boosting	واسنجی	138.4	228.5	0.87	0.88	520	700	0.09	0.07	520	700	0.09	0.07
	اعتبارسنجی	152.06	241.45	0.84	0.83	537.19	720.54	0.07	0.05	537.19	720.54	0.07	0.05
Super Learning	واسنجی	130	210	0.9	0.93	460	630	0.14	0.14	460	630	0.14	0.14
	اعتبارسنجی	143.26	225.72	0.87	0.91	485.12	650.47	0.12	0.12	485.12	650.47	0.12	0.12

بحث

می‌کند و هیچیک از الگوریتم‌های دیگر در مقایسه با سوپر لرنینگ دارای برتری نشدند.

برای ارزیابی مدل در این تحقیق از اعتبارسنجی متقاطع (CV) استفاده شد. نتایج حاصل از تحلیل‌ها نشان داد که افزودن متغیر pH به عنوان ورودی در برخی مدل‌ها موجب کاهش نسبی دقت پیش‌بینی پارامترهای کیفیت آب شده است. این امر بیانگر نقش محدود pH در پیش‌بینی پارامترهای مورد مطالعه می‌باشد. با این وجود، pH یکی از پارامترهای کلیدی کیفیت شیمیایی آب است که می‌تواند به صورت غیرمستقیم و از طریق تأثیر بر تعادل‌های شیمیایی، بر ویژگی‌های دیگر آب مؤثر باشد. با توجه به هدف اصلی این مطالعه، بررسی قابلیت پیش‌بینی پارامترهای شیمیایی آب بر اساس داده‌های پایه‌ای و در دسترس مانند EC و pH بوده است. کاهش دقت مدل در حضور pH نیز به عنوان یافته‌ای علمی می‌باشد. همانگونه که شکل ۳ نشان می‌دهد، بالاترین همبستگی میان TH و EC با مقدار ۰/۹۱ است و کمترین همبستگی را RSC و TH با میزان ۰/۹۱- دارند. در این شکل، هر دایره نمایانگر میزان و جهت همبستگی بین دو پارامتر است. رنگ آبی نشان‌دهنده همبستگی مثبت و رنگ قرمز نشان‌دهنده همبستگی منفی می‌باشد، در حالی که اندازه دایره‌ها بیانگر شدت همبستگی (قدر مطلق ضریب همبستگی) است. اعداد داخل ماتریس مقادیر عددی ضریب همبستگی پیرسون را نمایش می‌دهند و نام پارامترها به رنگ قرمز در قطر ماتریس درج شده‌اند.

در پژوهش حاضر، کاربرد الگوریتم‌های مختلف، در برآورد مقادیر RSC, TH, SAR بر اساس مقادیر pH, EC به مطالعه گذاشته شد و روش قوی‌تر برای تخمین TH به دست آمد. مقادیر تخمینی مربوط به RSC و SAR در رده‌های بعدی قرار داشتند. نتایج حاصل از مدل‌سازی نشان داد که استفاده از pH به عنوان ورودی در پیش‌بینی شاخص‌های کیفیت آب، به‌ویژه RSC، کارایی بسیار محدودی دارد؛ به طوری که مقدار ضریب تعیین (R^2) کمتر از ۱۰ درصد به دست آمد. این مقدار کم R^2 تأیید می‌کند که pH به دلیل خصوصیات فیزیکوشیمیایی مستقل خود، فاقد توانایی لازم برای پیش‌بینی مستقیم پارامترهایی مانند SAR، TH و RSC است که عمدتاً بر اساس غلظت یون‌های کلسیم، منیزیم و سدیم تعریف می‌شوند. از سوی دیگر، pH با وجود عدم ارتباط مستقیم با یون‌های محلول، همچنان پارامتری کلیدی است که وضعیت شیمیایی کلی آب را تحت تأثیر تعادل‌های پیچیده‌ای مانند تعادل کربنات‌ها و دی‌اکسید کربن محلول قرار می‌دهد؛ بنابراین، ممکن است نقش غیرمستقیم و حمایتی در تعیین کیفیت آب داشته باشد که در پژوهش حاضر، مدل‌های یادگیری ماشین تلاش کردند این ارتباط‌های آماری را کشف کنند. با توجه به این محدودیت‌ها، توصیه می‌شود که در تحقیقات آتی، pH به عنوان بخشی از مجموعه‌ای جامع‌تر از پارامترهای ورودی مورد استفاده قرار گیرد تا بتوان به مدلی با دقت بالاتر و تبیین علمی قوی‌تر دست یافت. همچنین، تحلیل‌های بیشتر برای درک بهتر نقش pH در تعادل شیمیایی آب و ارتباطات غیرمستقیم آن ضروری است.

همانگونه که شکل ۲ نشان می‌دهد، در تمامی حالت‌های تخمین، مدل سوپرلرنینگ بالاترین قابلیت را داشته است و این الگوریتم یک روش عملی مناسب برای پیش‌بینی مقادیر مورد نظر است و یک روش سریع و اقتصادی برای این هدف است، از سایر الگوریتم‌های مورد مطالعه برتری فاحشی دارد، به طوری که در حالت‌های مختلف رتبه دوم به الگوریتم‌های متفاوتی اختصاص پیدا

EC	SL	0.73	0.79	0.84	0.86	0.87	0.91	0.58
	EGB	0.63	0.69	0.76	0.81	0.84	0.83	0.61
	RF	0.67	0.71	0.76	0.82	0.81	0.82	0.64
	C	0.64	0.69	0.76	0.82	0.79	0.81	0.67
	SVR	0.64	0.65	0.67	0.72	0.76	0.8	0.7
	KNN	0.57	0.65	0.67	0.74	0.73	0.8	0.73
	DT	0.66	0.68	0.61	0.64	0.7	0.73	0.76
		R ²	CCC	R ²	CCC	R ²	CCC	0.79
								0.82
pH	SL	0.09	0.11	0.11	0.12	0.12	0.12	0.85
	EGB	0.03	0.02	0.06	0.06	0.07	0.05	0.88
	RF	0.06	0.05	0.07	0.06	0.06	0.03	0.91
	C	0.06	0.01	0.07	0.03	0.07	0.07	R ² , CCC >0.5
	SVR	0.06	0.01	0.07	0.02	0.09	0.05	0.01
	KNN	0.03	0.03	0.07	0.04	0.06	0.06	0.02
	DT	0.06	0.06	0.06	0.08	0.06	0.08	0.03
		R ²	CCC	R ²	CCC	R ²	CCC	0.04
								0.05
EC,PH	SL	0.76	0.83	0.06	0.08	0.12	0.12	0.06
	EGB	0.67	0.73	0.07	0.04	0.07	0.05	0.07
	RF	0.69	0.75	0.07	0.02	0.06	0.03	0.08
	C	0.68	0.72	0.07	0.03	0.07	0.07	0.09
	SVR	0.66	0.69	0.07	0.06	0.09	0.05	0.1
	KNN	0.61	0.67	0.06	0.06	0.06	0.06	0.11
	DT	0.68	0.71	0.11	0.12	0.06	0.08	0.12
		R ²	CCC	R ²	CCC	R ²	CCC	R ² , CCC <0.5
		SAR		RSC		TH		

شکل ۲- مقادیر آماری محاسبه شده برای ۷ الگوریتم جهت تخمین RSC ، SAR ، TH با استفاده از EC و pH

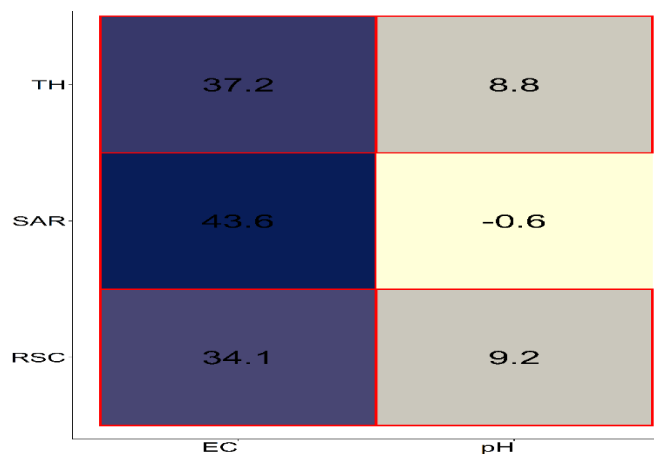
(SL=Super Learning, EGB=Extreme Gradient Boosting, RF=Random Forest, C=Cubist, SVR=Support Vector Regression, K-NN=K-Nearest Neighbors, DT=Decision Tree)



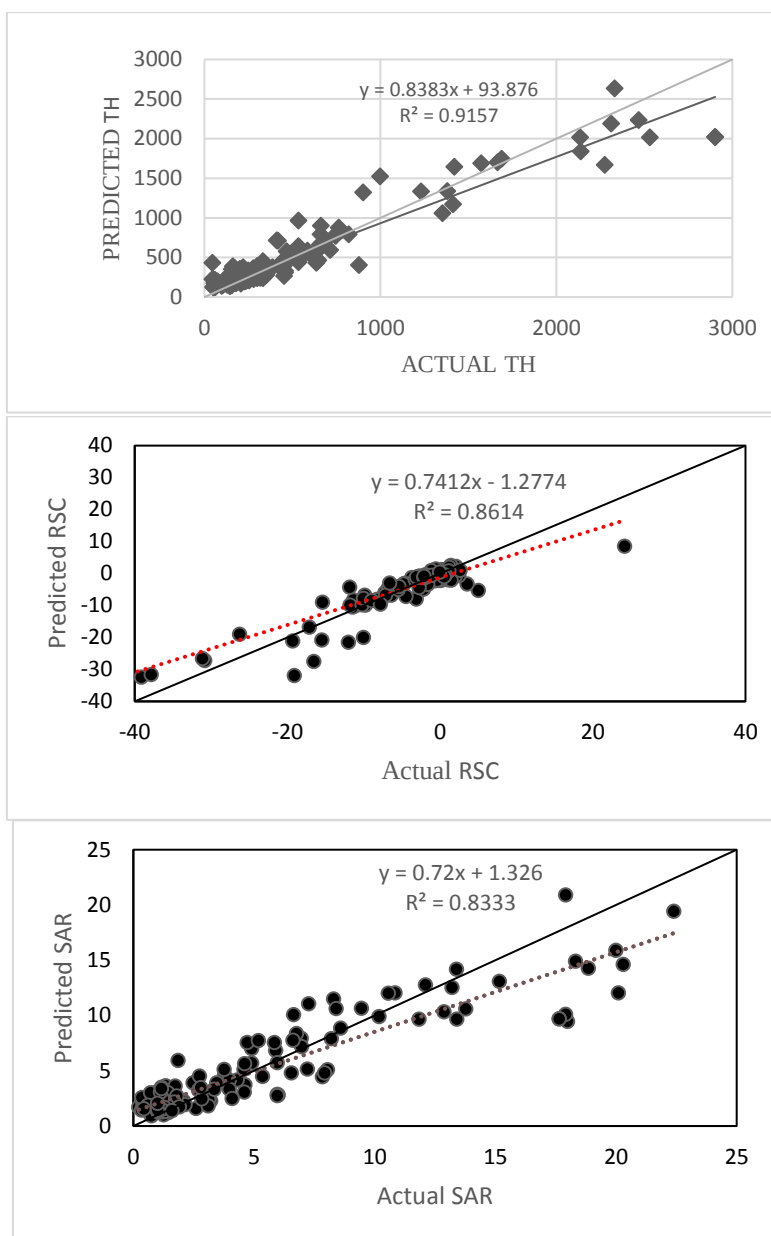
شکل ۳- ماتریس همبستگی بین پارامترهای شیمیایی آب

همچنین برای بررسی اینکه کدام ورودی مهم است، شکل ۴ نشان می‌دهد که EC برای پیش‌بینی مهمتر است. در مقایسه کارایی این مدل، هرچه نتایج پیش‌بینی، بیشتر منطبق بر نتایج مشاهده شده باشد، مدل از کارایی مطلوب‌تری برخوردار است. هنگام مقایسه این مدل‌ها از نظر کارایی، هرچه نتایج پیش‌بینی سازگارتر باشند، مدل از نظر نتایج مشاهده شده، کارآمدتر است. برای بررسی نزدیکی مقدار واقعی و پیش‌بینی نیز از چهار معیار

در شکل ۵، ارتباط بین مقادیر برآورد شده و مقادیر واقعی، با استفاده از این مدل ارائه داده شده است.



شکل ۴- نتیجه الگوریتم



شکل ۵- همبستگی بین مقادیر پیش‌بینی شده و مشاهده شده

و همکاران (۲۰۲۰)، دی و همکاران (۲۰۱۹)، لیو و همکاران (۲۰۱۹)، لو و ما (۲۰۲۰) و وانگ و همکاران (۲۰۱۷) ارائه شده است. پژوهش وانگ و همکاران (۲۰۲۳) که در آن از الگوریتم‌های SVM و KNN برای پیش‌بینی TH و SAR استفاده کرده‌اند. همچنین، چن و همکاران (۲۰۲۰) بر این نکته تأکید کرده‌اند که سوپرلرنینگ‌قادر است ویژگی‌های پیچیده‌تر خاک را بهتر مدل‌سازی کند و در مقایسه با درخت تصمیم یا شبکه‌های عصبی عملکرد بهتری ارائه دهد. این نتایج با پژوهش حاضر مطابقت داشت.

نمودارهای مختلف نشان می‌دهد که استفاده از روش سوپرلرنینگ در برآورد مقادیر TH، با R^2 برابر با ۰/۹۱ بین مقادیر TH واقعی و مقادیر TH برآوردی، دارای بیشترین دقت است و پس از آن R^2 بین مقادیر واقعی و برآوردی RSC، با مقدار ۰/۸۶ و در نهایت R^2 بین مقادیر واقعی SAR با مقادیر برآوردی آن، با مقدار ۰/۸۴ است. نتایج پیش‌بینی‌های دقیقی با استفاده از الگوریتم‌های مختلف بیان شده است. از جمله، نتایجی که توسط احمد و همکاران (۲۰۱۹)، کاستریلو و گارسیا (۲۰۲۰)، چن

در مجموع، یافته‌های این پژوهش نشان داد که الگوریتم سوپرلرنینگ با فراهم آوردن دقت زیاد در تخمین شاخص‌های کیفی آب، می‌تواند به‌عنوان یک ابزار کارآمد در مدیریت منابع آب و کشاورزی مورد استفاده قرار گیرد. بر اساس نتایج مدل‌سازی، ارتباط میان متغیرهای ورودی EC و pH و شاخص‌های کیفیت آب SAR، TH و RSC را می‌توان به‌صورت رابطه‌های (۴ تا ۶) بیان کرد:

$$SAR = -382/9 + 489/1 EC + 48/1 pH \quad (4)$$

$$TH = 11/1328 + 36/189 EC - 4/159 pH \quad (5)$$

$$RSC = -0.4/23 EC - 5/3 EC + 27/3 pH \quad (6)$$

نتیجه‌گیری

نتایج پژوهش حاضر نشان داد که سوپرلرنینگ به‌عنوان یک تکنیک ترکیبی توانسته است عملکرد بهتری نسبت به الگوریتم‌های سنتی یادگیری ماشین مانند درخت تصمیم، رگرسیون خطی و شبکه‌های عصبی در پیش‌بینی به‌دست آورد. در این مطالعه، از هفت مدل یادگیری ماشین استفاده شده و برای پیش‌بینی سه پارامتر ارزیابی شد. روش‌های مرسوم ارزیابی کیفیت آب آبیاری، بر اساس تعداد زیادی از پارامترها و شاخص‌ها می‌باشند و ابزارهای کارآمدی محسوب می‌شوند، اما می‌توانند هزینه‌بر باشند. در مقایسه با روش‌های دیگر، سوپرلرنینگ توانست دقت پیش‌بینی بالاتری ارائه دهد. این امر، به دلیل قدرت سوپرلرنینگ در ادغام و بهینه‌سازی مدل‌های مختلف برای رسیدن به بهترین نتیجه ممکن، است. به‌علاوه، زمان آموزش مدل‌های سوپرلرنینگ در مقایسه با مدل‌های پیچیده‌تر مانند شبکه‌های عصبی به مراتب کوتاه‌تر بود که این ویژگی می‌تواند در کاربردهای عملی بسیار ارزشمند باشد. با این حال، پژوهش حاضر نیز محدودیت‌هایی دارد. به‌عنوان مثال، مدل‌های سوپرلرنینگ برای مقیاس‌های بزرگتر داده‌ها و مجموعه‌های پیچیده‌تر ممکن است به زمان محاسباتی بیشتری نیاز داشته باشند. همچنین، پژوهش‌های آینده می‌توانند به بررسی قابلیت سوپرلرنینگ در

مقیاس‌های زمانی مختلف و یا برای داده‌های مختلف از دیگر ویژگی‌ها بپردازند. نتایج این تحقیق همچنین نشان داد که استفاده از داده‌های EC و pH به‌عنوان ورودی‌ها، به‌ویژه برای پیش‌بینی SAR و TH، می‌تواند به‌طور مؤثری برای ارزیابی ویژگی‌های آب استفاده شود. با مقایسه نتایج مدل‌های خطی و الگوریتم سوپرلرنینگ مشاهده شد که مدل‌های خطی دقت پایین‌تری داشتند، در حالی که سوپرلرنینگ با توانایی شناسایی الگوهای پیچیده و غیرخطی بین EC و pH و شاخص‌های کیفیت آب (TH، SAR و RSC)، دقت بالاتری ارائه داد؛ بنابراین ماهیت رابطه بین ورودی‌ها و خروجی‌ها غیرخطی است و مدل‌های ترکیبی پیشرفته برای پیش‌بینی این شاخص‌ها مناسب‌تر می‌باشند. این یافته‌ها می‌توانند به پیشرفت‌های عملی در کشاورزی کمک کنند، به‌ویژه در مناطقی که نیاز به پیش‌بینی دقیق ویژگی‌های آب برای مدیریت و بهینه‌سازی فرآیندهای کشاورزی دارند. به عبارتی دیگر، مدل سوپرلرنینگ ابزاری کارآمد برای پیش‌بینی دقیق کیفیت آب آبیاری است، آن هم تنها با استفاده از پارامترهایی که می‌توانند به‌طور مستقیم در مدت زمان کوتاهی اندازه‌گیری شوند. در نتیجه، پیاده‌سازی فناوری‌های حسگر خودکار همراه با مدل‌های مذکور، کنترل کیفیت آب را بهبود می‌بخشد و به کشاورزان در مدیریت کیفیت آب آبیاری کمک خواهد کرد و با اندازه‌گیری پارامترهای ورودی آب، بر کیفیت آب آبیاری نظارت بیشتری داشته باشند. در نهایت، این تحقیق، به اهمیت انتخاب مدل‌های ترکیبی پیشرفته مانند سوپرلرنینگ برای مسائل پیچیده پیش‌بینی در علوم آب، اشاره دارد و امیدوار است که این نتایج بتواند به توسعه ابزارهای پیشرفته‌تر در این حوزه‌ها کمک کند.

تعارض منافع

در این مقاله تعارض منافی وجود ندارد و این مسئله مورد تأیید نویسندگان مقاله است.

فهرست منابع

۱. آزاد اسلامیه، محمدرضا، کلانتری، سعیده، شیرمردی، مصطفی، و تازه، مهدی، ۱۳۹۵. بررسی کاربری اراضی و خصوصیات شیمیایی-فیزیکی خاک بر سرعت آستانه فرسایش بادی با استفاده از داده‌کاوی، نشریه مهندسی اکوسیستم بیابان، ۹(۲۹)، صص ۱۲-۱. <https://doi.org/10.22052/deej.2020.9.29.1>
۲. تقی‌زاده، روح‌الله، غزالی، آرزو، کلانتری، سعیده، و رحیمیان، محمد حسن، ۱۳۹۵. مکان‌یابی شوری خاک با استفاده از داده‌های محیطی و نمونه‌برداری هایپرکیوب در شهرستان میبد، نشریه خشک بوم، ۱(۶)، صص ۷۹-۶۹. <https://doi.org/20.1001.1.2008790.1395.6.1.6.9>
۳. فتحی‌زاد، حسن، تازه، مهدی، و کلانتری، سعیده، ۱۳۹۴. مقایسه کارایی روشهای طبقه‌بندی پیکسل پایه روشهای شبکه عصبی آرتمپ فازی و تصمیم‌گیری درختی (و شیء‌گرا در تهیه نقشه کاربری اراضی) مطالعه موردی: حوزه خشک و نیمه‌خشک میمه، استان ایلام، نشریه خشک بوم، ۲(۵)، صص ۸۱-۶۹.
4. Afendras, G. and Markatou, M., 2019. Optimality of training/test size and resampling effectiveness in cross-validation. *Journal of Statistical Planning and Inference*, 199, pp.286-301. <https://doi.org/10.1016/j.jspi.2018.07.005>
5. Ahmed, A.N., Binti Othman, F., Abdulmohsin Afan, H., Khaleel Ibrahim, R., Ming Fai, C., Shabbir Hossain, M.D., Ehteram, M., Elshafie, A., 2019. Machine learning methods for better water quality prediction. *J. Hydrol.* 578, 124084. <https://doi.org/10.1016/j.jhydrol.2019.124084>
6. Cagliari, J., Veronez, M.R. and Alves, M.E., 2011. Remaining phosphorus estimated by pedotransfer function. *Revista Brasileira de Ciência do Solo*, 35, pp.203-212. <https://doi.org/10.1590/S0100-06832011000100019>
7. Carr, P.M., Brevik, E.C., Horsley, R.D. and Martin, G.B., 2015. Long-Term No-Tillage Sequesters Soil Organic Carbon in Cool Semiarid Regions. *Soil Horizons*, 56(6), pp.1-7. <https://doi.org/10.2136/sh15-07-0016>
8. Castrillo, M., García, Á.L., 2020. Estimation of high frequency nutrient concentrations from water quality surrogates using machine learning methods. *Water* watres.2020.115490. Res. 172, 115490. <https://doi.org/10.1016/j>
9. Chen, K., Chen, H., Zhou, C., Huang, Y., Qi, X., Shen, R., Liu, F., Zuo, M., Zou, X., Wang, J., Zhang, Y., Chen, D., Chen, X., Deng, Y., Ren, H., 2020. Comparative analysis of surface water quality prediction performance and identification of key water parameters using different machine learning models based on big data. *Water Res.* 171, 115454. <https://doi.org/10.1016/j.watres.2019.115454>
10. Di, Z., Chang, M., Guo, P., 2019. Water quality evaluation of the Yangtze River in China using machine learning techniques and data monitoring on different time scales. *Water (Switzerland)* 11. <https://doi.org/10.3390/w11020339>
11. Fathizad, H., Tazeh, M., Kalantari, S. and Shojaei, S., 2017. The investigation of spatiotemporal variations of land surface temperature based on land use changes using NDVI in southwest of Iran. *Journal of African Earth Sciences*, 134, pp.249-256. <https://doi.org/10.1016/j.jafrearsci.2017.06.007>
12. Gupta, S.K., 1983. Variations of water table in Yamuna drainage basin of Haryana-implications and management strategies. In *Seminar on Strategies for Irrigation Water Management, Patna*.
13. Liu, P., Wang, J., Sangaiah, A., Xie, Y., Yin, X., 2019. Analysis and prediction of water quality using LSTM deep neural networks in IoT environment. *Sustainability* 11, 2058. <https://doi.org/10.3390/su11072058>
14. Lu, H., Ma, X., 2020. Hybrid decision tree-based machine learning models for short term water quality prediction. *Chemosphere* 249, 126169. <https://doi.org/10.1016/j.chemosphere.2020.126169>

15. Ibrahim, H., Yaseen, Z.M., Scholz, M., Ali, M., Gad, M., Elsayed, S., Khadr, M., Hussein, H., Ibrahim, H.H., Eid, M.H. and Kovács, A., 2023. Evaluation and prediction of groundwater quality for irrigation using an integrated water quality indices, machine learning models and GIS approaches: A representative case study. *Water*, 15(4), p.694. <https://doi.org/10.3390/w15040694>
16. Marakoğlu, T. and Çarman, K., 2010. Fuzzy knowledge-based model for prediction of soil loosening and draft efficiency in tillage. *Journal of Terramechanics*, 47(3), pp.173-178. <https://doi.org/10.1016/j.jterra.2009.10.001>
17. McBratney, A.B., Odeh, I.O., Bishop, T.F., Dunbar, M.S. and Shatar, T.M., 2000. An overview of pedometric techniques for use in soil survey. *Geoderma*, 97(3-4), pp.293-327. [https://doi.org/10.1016/S0016-7061\(00\)00043-4](https://doi.org/10.1016/S0016-7061(00)00043-4)
18. Merdun, H., Çınar, Ö., Meral, R. and Apan, M., 2006. Comparison of artificial neural network and regression pedotransfer functions for prediction of soil water retention and saturated hydraulic conductivity. *Soil and Tillage Research*, 90(1-2), pp.108-116. <https://doi.org/10.1016/j.still.2005.08.011>
19. Minasny, B., McBratney, A.B., Brough, D.M. and Jacquier, D., 2011. Models relating soil pH measurements in water and calcium chloride that incorporate electrolyte concentration. *European Journal of Soil Science*, 62(5), pp.728-732. <https://doi.org/10.1111/j.1365-2389.2011.01386.x>
20. Papacharalampous, G., Tyralis, H. and Koutsoyiannis, D., 2019. Comparison of stochastic and machine learning methods for multi-step ahead forecasting of hydrological processes. *Stochastic environmental research and risk assessment*, 33(2), pp.481-514. <https://doi.org/10.1007/s00477-018-1638-6>
21. Peltre, C., Thuriès, L., Barthès, B., Brunet, D., Morvan, T., Nicolardot, B., Parnaudeau, V. and Houot, S., 2011. Near infrared reflectance spectroscopy: A tool to characterize the composition of different types of exogenous organic matter and their behaviour in soil. *Soil Biology and Biochemistry*, 43(1), pp.197-205. <https://doi.org/10.1016/j.soilbio.2010.09.036>
22. Polley, E., LeDell, E., Kennedy, C., Lendle, S. and van der Laan, M.J., 2019. SuperLearner: super learner prediction R packages. R.[Online] Available: <https://github.com/ecpolley/SuperLearner> [2019 Jan. 2].
23. Deka, P.C., 2014. Support vector machine applications in the field of hydrology: a review. *Applied soft computing*, 19, pp.372-386. <https://doi.org/10.1016/j.asoc.2014.02.002>
24. Sagi, O. and Rokach, L., 2018. Ensemble learning: A survey. *Wiley interdisciplinary reviews: data mining and knowledge discovery*, 8(4), p.e1249. <https://doi.org/10.1002/widm.1249>
25. Shaw, S.K. and Sharma, A., 2025. Evaluation and prediction of groundwater quality for irrigation using regression and machine learning models. *Water Quality Research Journal*, p.wqrj2025075. <https://doi.org/10.2166/wqrj.2025.075>
26. Simm, J., De Abril, I.M. and Sugiyama, M., 2014. Tree-based ensemble multi-task learning method for classification and regression. *IEICE TRANSACTIONS on Information and Systems*, 97(6), pp.1677-1681. <https://doi.org/10.1587/transinf.E97.D.1677>
27. Tyralis, H., Papacharalampous, G., Burnetas, A. and Langousis, A., 2019. Hydrological post-processing using stacked generalization of quantile regression algorithms: Large-scale application over CONUS. *Journal of Hydrology*, 577, p.123957. <https://doi.org/10.1016/j.jhydrol.2019.123957>
28. Tyralis, H., Papacharalampous, G. and Langousis, A., 2019. A brief review of random forests for water scientists and practitioners and their recent history in water resources. *Water*, 11(5), p.910. <https://doi.org/10.3390/w11050910>
29. Wang, X., Zhang, F., Ding, J., 2017. Evaluation of water quality based on a machine learning algorithm and water quality index for the Ebinur Lake Watershed. China. *Sci. Rep.* 7, 1–18. <https://doi.org/10.1038/s41598-017-12853-y>.

30. Wang, D., Liu, S., Zhang, C., Xu, M., Yang, J., Yasir, M. and Wan, J., 2023. An improved semantic segmentation model based on SVM for marine oil spill detection using SAR image. *Marine Pollution Bulletin*, 192, p.114981.
<https://doi.org/10.1016/j.marpolbul.2023.114981>
31. Wilcox, L.V., Blair, G.Y. and Bower, C.A., 1954. Effect of bicarbonate on suitability of water for irrigation. *Soil Science*, 77(4), pp.259-266. **<https://doi.org/10.1097/00010694-195404000-00001>**
32. Zhu, A.X., Qi, F., Moore, A. and Burt, J.E., 2010. Prediction of soil properties using fuzzy membership values. *Geoderma*, 158(3-4), pp.199-206.
<https://doi.org/10.1016/j.geoderma.2010.05.001>