

## نقشه برداری رقومی شوری خاک با استفاده از داده‌های کمکی و مدل‌های یادگیری

### ماشین در حوضه آبخیز بدر، استان کردستان

مسلم زرینی بهادر\* و رحمان شریفی

کارشناس تحقیقات خاک و آب مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان تهران، سازمان تحقیقات، آموزش و ترویج کشاورزی (تات)، تهران، ایران. (نویسنده

مسئول مقاله) ایمیل: [Moslem.zarini@gmail.com](mailto:Moslem.zarini@gmail.com)

استادیار زمین شناسی، عضو هیئت علمی مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان تهران، سازمان تحقیقات، آموزش و ترویج کشاورزی (تات)، تهران، ایران:

ایمیل: [rahmansharifi@yahoo.com](mailto:rahmansharifi@yahoo.com)

«مقاله پژوهشی»

دریافت: ۱۴۰۳/۴/۲۴ و پذیرش: ۱۴۰۳/۷/۲۵

### چکیده

استفاده از تکنیک‌های سنجش از دور و یادگیری ماشین به طور فزاینده‌ای به عنوان روش‌های مقرون به صرفه برای نمایش نقشه شوری خاک شناخته می‌شود. در این مطالعه از داده‌های ماهواره‌ای لندست ۸ و تکنیک‌های پیچیده یادگیری ماشین برای ترسیم و ارزیابی سطوح شوری خاک در حوضه آبخیز بدر استفاده گردید. به این منظور، از چندین تکنیک یادگیری ماشین در نرم افزار R برای پیش‌بینی مقادیر شوری در حوضه آبخیز بدر استفاده شد. این الگوریتم‌ها نزدیک ترین همسایه K (KNN)، تحلیل درخت تصمیم (DTA)، شبکه عصبی مصنوعی (ANN)، جنگل تصادفی (R.F.) و رگرسیون چند متغیره خطی ترکیبی (MLR) را در بر می‌گرفت. برای انجام این مطالعه، بر اساس تکنیک ابر مکعب لاتین، محل ۱۲۵ خاک‌رخ در منطقه مطالعاتی انتخاب و حفر شد. نمونه‌های خاک، پس از هوا خشک شدن در محیط آزمایشگاه، کوبیده شد و از الک ۲ میلی‌متری عبور داده شد. سپس شوری خاک اندازه‌گیری گردید. برای برآورد ویژگی‌های خاک، دو حالت مختلف مورد بررسی قرار گرفت. در حالت اول، مدل‌های شبکه عصبی مصنوعی، تحلیل درخت تصمیم و رگرسیون چند متغیره خطی برای پیش‌بینی استفاده شد. همچنین، برای ترکیب نتایج مدل‌ها، از مدل نزدیک‌ترین همسایه استفاده شد. نتایج این مطالعه نشان داد که متغیرهای کمکی مهم در پیش‌بینی شوری خاک به ترتیب اهمیت عبارت‌اند از: ژئومورفولوژی، عمق دره، شاخص همواری قله برآمدگی با درجه تفکیک بالا، شاخص خیزی، جهت شیب، مدل رقومی ارتفاع، شیب حوضه، موقعیت نسبی شیب، مقدار شیب و طول شیب. همچنین، نتایج ارزیابی مدل‌ها نشان داد که در میان مدل‌های استفاده‌شده برای پیش‌بینی شوری، مدل رگرسیون چند متغیره خطی ترکیبی (MLR) با ضریب تعیین ۰/۶۱۱ و ریشه دوم متوسط مربعات خطا ۰/۳۲ از بیشترین دقت برای پیش‌بینی برخوردار بوده است.

واژه‌های کلیدی: لندست ۸ الگوریتم نزدیک ترین همسایه، تحلیل درخت تصمیم، شبکه عصبی مصنوعی، جنگل تصادفی، رگرسیون

چند متغیره خطی ترکیبی

\* - آدرس ایمیل نویسنده مسئول: [Moslem.zarini@gmail.com](mailto:Moslem.zarini@gmail.com)



## مقدمه

شوری خاک یک چالش مهم در بسیاری از مناطق در سراسر جهان از جمله ایران است. پیامدهای بالقوه شامل تأثیرات منفی بر عملکرد محصول، رفاه انسان و سیستم‌های زیست‌محیطی است. استفاده از تکنیک‌های سنجش از دور و یادگیری ماشین به طور فزاینده‌ای به عنوان روش‌های مقرون به صرفه برای نمایش نقشه شوری خاک شناخته می‌شود (الشوی و همکاران، ۲۰۲۴).

شوری خاک در حال حاضر یک مشکل جدی زیست‌محیطی است که در سطح جهانی با آن مواجه هستیم (لی و همکاران، ۲۰۱۴). شوری خاک به شدت بر تولید کشاورزی، تعادل اکولوژیکی و توسعه پایدار تأثیر می‌گذارد (وانگ و همکاران، ۲۰۱۲؛ بوکسیلا و همکاران، ۲۰۱۳؛ کانیدو و همکاران، ۲۰۱۳). بنابراین، درک سطوح شوری خاک برای تعیین زمان بندی بهینه آبیاری، اجرای شیوه‌های مدیریت موثر، تضمین کیفیت خاک، تقویت رشد گیاه قوی و حفظ منابع آب بسیار مهم است (اسکودپرو و همکاران، ۲۰۱۷). با توجه به این زمینه، انجام یک بررسی جامع از پویایی مکانی و زمانی شوری به عنوان یک ضرورت مبرم برای حفاظت از منابع زمین، حفظ تعادل اکولوژیکی و تضمین بهره‌وری کشاورزی پدیدار شده است.

روش‌های مرسوم برای اندازه‌گیری شوری خاک معمولاً شامل جمع‌آوری نمونه‌های خاک مزرعه برای آنالیز آزمایشگاهی برای تعیین غلظت املاح است. با این حال، چنین روش‌هایی زمان بر و پرهزینه هستند (دنت و همکاران، ۱۹۸۱؛ نانی و همکاران، ۲۰۰۶؛ دهان و همکاران، ۲۰۱۰). در سال‌های اخیر، استفاده از تکنیک‌های سنجش از دور برای ارزیابی شوری خاک به دلیل زمان اندازه‌گیری کوتاه، دامنه نظارت گسترده، و ظرفیت به دست آوردن داده‌های چند زمانی افزایش یافته است (آسفاو و همکاران، ۲۰۱۸؛ گرجی و همکاران، ۲۰۱۷). به طور خاص، فناوری سنجش از دور ابزارهای کارآمد و مقرون به صرفه را برای پایش سریع شوری خاک ارائه می‌دهد. تکنیک‌های پایش را می‌توان به دو دسته تقسیم کرد، یعنی ارزیابی مستقیم خاک

لخت و ارزیابی غیرمستقیم از طریق پوشش گیاهی تحت تأثیر شوری خاک (موگنوت، ۱۹۹۳). اخیراً محققان ارتباطی بین اطلاعات رشد پوشش گیاهی و غلظت شوری خاک از طریق شاخص‌های پوشش گیاهی تعیین کرده‌اند (وانگ و همکاران، ۲۰۲۱). علاوه بر این، تعداد فزاینده‌ای از مطالعات شاخص شوری و شاخص ترکیبی آن را برای پایش شوری خاک اتخاذ کرده‌اند (ژانگ و همکاران، ۲۰۲۲). دینگ و همکاران، ۲۰۱۲). استفاده از متغیرهای محیطی مانند ارتفاع، شیب و جهت، پتانسیل بالایی برای شناسایی دقیق خاک‌های شور دارد؛ که با افزایش ارتفاع، رطوبت خاک افزایش یافته و شوری کاهش می‌یابد و همچنین در شیب‌های رو به شمال به دلیل نگهداشت بیشتر رطوبت در خاک سبب کاهش شوری خاک می‌گردد (تریکی فوراتی و همکاران، ۲۰۱۷). تحقیقات قبلی همبستگی بین این شاخص‌ها و شاخص‌های شوری خاک را نشان داده است (اله‌ارتی و همکاران، ۲۰۱۶). بسیاری از محققان شوری خاک را با ایجاد رابطه بین شاخص‌های مختلف (به عنوان مثال، شاخص‌های طیفی) و شوری خاک اندازه‌گیری شده، انتخاب عوامل وارونگی بهینه و ساخت مدل‌های وارونگی از طریق داده‌های سنجش از دور ماهواره‌ای متنوع را پیش-بینی کرده‌اند (وو و همکاران، ۲۰۲۱؛ شی و همکاران، ۲۰۲۱).

اکثر تحقیقات مرتبط قبلی، بر مدل‌های تجربی تکیه دارند. با این حال، Google Earth Engine (GEE) داده‌های سری Landsat تصحیح شده را همراه با رابط‌های الگوریتم یادگیری ماشین ارائه می‌دهد. این منابع امکان استفاده از روش‌های یادگیری ماشینی در مقیاس بزرگ (جیانگ و همکاران، ۲۰۱۹. وانگ و همکاران، ۲۰۱۲. باربوچی و همکاران، ۲۰۱۴) و یادگیری عمیق (ژانگ و همکاران، ۲۰۲۰؛ تیان و همکاران، ۲۰۱۹) را فراهم می‌کند که از داده‌های ابری گسترده و قدرت محاسباتی استفاده می‌کند. چنین مدل‌هایی توانایی نظارت بر شوری خاک را افزایش می‌دهند. علاوه بر این، محققان از شاخص‌های پوشش گیاهی برای ساخت مدل‌هایی در GEE برای مطالعه شوری خاک

مطابق با آمار طولانی‌مدت هواشناسی، متوسط درجه حرارت سالیانه‌ی هوا ۱۲/۰۷ درجه سانتی‌گراد و میانگین بارندگی سالیانه ۳۴۵/۸ میلی‌متر است. الگوی بارش، مدیترانه‌ای بوده و بخش اعظم بارندگی در فصول سرد سال اتفاق می‌افتد. بر اساس اقلیم‌نمای آمبرژه، اقلیم این ناحیه نیمه‌خشک و سرد است. در منطقه مورد مطالعه، رژیم رطوبتی خاک زیریک (Xeric) و رژیم حرارتی آن مزیک (Mesic) تعیین شده است.

قدیمی‌ترین واحد چینه‌شناسی در منطقه، تریاس است. رسوبات تریاس از دولومیت کاملاً بلورین و خردشده تشکیل شده است. روی این واحد چینه‌ای را شیل، شیست و ولکانیک‌های مشابه با سازند شمشک در البرز مرکزی، به سن ژوراسیک، می‌پوشانند. آخرین توالی رسوبی به ظاهر پیوسته مزوزوئیک- ترشیاری، شیل‌ها و ماسه‌سنگ‌های دارای عدسی‌های آهکی به همراه سنگ‌های آتش‌فشانی هستند که به‌طور عمده همساز بر روی رسوبات کرتاسه بالائی (پالئوسن) قرار دارند (حریری، ۱۳۷۴). اصلی‌ترین رخداد کوهزایی در این منطقه مربوط به کوهزایی کیمیرین میانی است که سبب دگرگونی سنگ‌های رسوبی- آتش‌فشانی (تریاس-ژوراسیک) گردیده و حاصل آن تشکیل مرمر، آمفیبولیت، شیست و فیلیت است (شکل ۲) (حسینی، ۱۳۷۶).

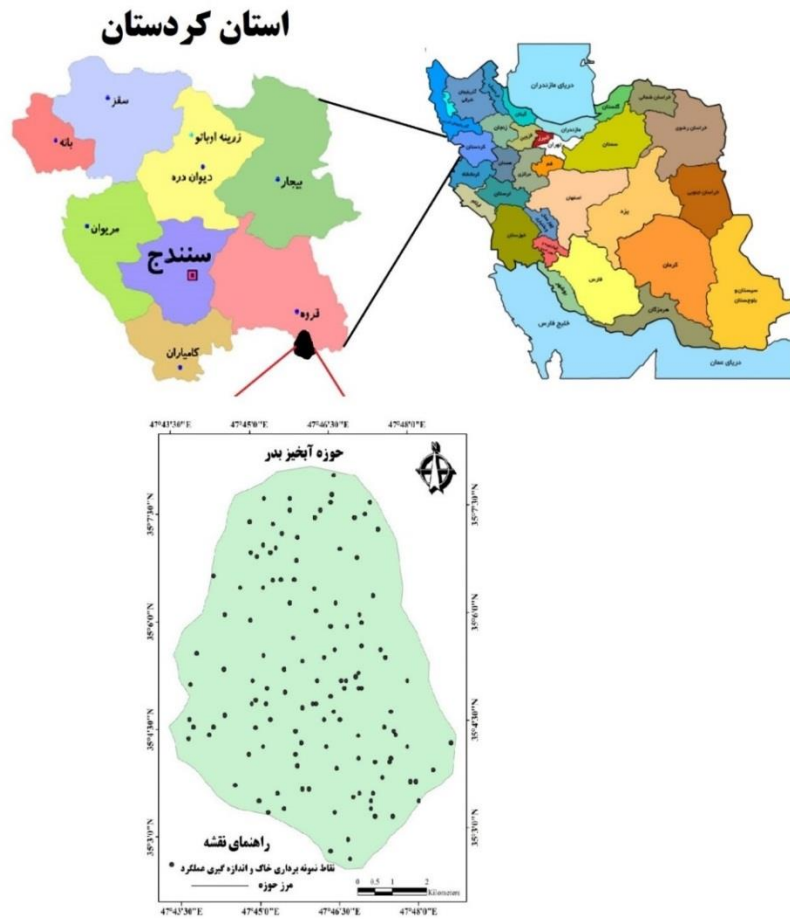
استفاده کرده‌اند (ما و همکاران، ۲۰۲۳). در حال حاضر، ادغام GEE و یادگیری ماشین برای نقشه‌برداری زمین‌های زراعی (دانگ و همکاران، ۲۰۱۶؛ ژیانگ و همکاران، ۲۰۱۷)، تخمین دمای سطح (پاراستاتیدیس و همکاران، ۲۰۱۷) و نظارت بر طبقه‌بندی (براولی و همکاران، ۲۰۲۰) استفاده می‌شود. البته در این مطالعه ما از نرم افزار R برای تجزیه و تحلیل داده‌ها و به عنوان پلت فرم اصلی در این مطالعه استفاده نمودیم.

اهداف این مطالعه عبارتند از: انتخاب بهترین مدل پیش‌بینی کننده شوری خاک، بررسی دقت و صحت مدل‌های یادگیری ماشین، تهیه نقشه‌های توزیع مکانی شوری خاک در حوضه آبخیز بدر و تحلیل تغییرات شوری خاک در حوضه آبخیز بدر.

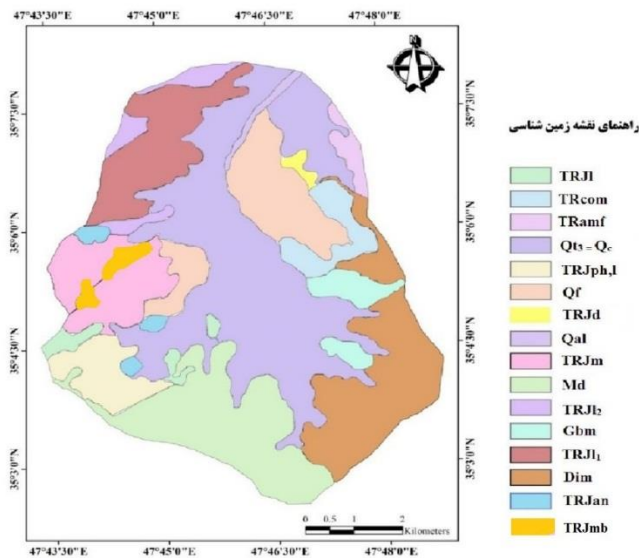
## مواد و روش‌ها

### وضعیت عمومی منطقه مورد مطالعه

اراضی مورد مطالعه با وسعتی حدود ۶۷۰۰ هکتار در جنوب شهر ستان قروه و بین مختصات جغرافیائی ۴۷ درجه و ۴۳ دقیقه و ۴۷ درجه و ۴۹ دقیقه طول شرقی و ۳۵ درجه و ۲ دقیقه و ۳۵ درجه و ۸ دقیقه عرض شمالی واقع شده است (شکل ۱).



شکل ۱ - موقعیت منطقه مطالعاتی و پراکنش مکانی نقاط نمونه برداری از لحاظ ژئومرفولوژی، حوزه‌ی آبخیز مورد مطالعه شامل کوه و اراضی پای کوهی (پیدمونت) است. کاربری‌های رایج اراضی کشاورزی، گندم دیم و آبی، جو دیم و آبی، نخود، لوبیا، یونجه، شبدر و... است.



شکل ۲- زمین شناسی حوضه آبخیز بدر شهرستان قروه، استان کردستان

## روش اجرا

برای اجرای این پژوهش مراحل زیر، طبق شکل ۳ انجام گرفت:

در مرحله‌ی اول، داده‌های کمکی از جمله: تصاویر ماهواره‌ای لندست ۸ و مدل رقومی ارتفاع با قدرت تفکیک مکانی ۱۰ متر منطقه از سازمان نقشه‌برداری کشور تهیه گردید. نقشه‌ی زمین‌شناسی قروه از سایت زمین‌شناسی کشور تهیه و نقشه زمین‌شناسی حوضه آبخیز بدر از آن استخراج و در محیط سامانه اطلاعات جغرافیایی رقومی گردید. نقشه‌ی ژئومرفولوژی با استفاده از نقشه زمین‌شناسی و بر اساس روش ژئوپدولوژی زینک در محیط سامانه اطلاعات جغرافیایی ترسیم گردید. در مرحله‌ی دوم، محل نقاط مشاهداتی تعیین، شناسایی خاک‌ها در صحرا، نمونه‌برداری از لایه‌های مختلف خاک، اندازه‌گیری‌های فیزیکی و شیمیایی و کانی‌شناسی خاک‌ها انجام و خاک‌ها طبقه‌بندی شدند. در مرحله سوم، مدل‌سازی انجام، نقشه‌ی ویژگی‌های خاک تهیه و ارزیابی مدل‌ها صورت گرفت.

## جمع‌آوری داده‌های متغیرهای کمکی، پردازش و

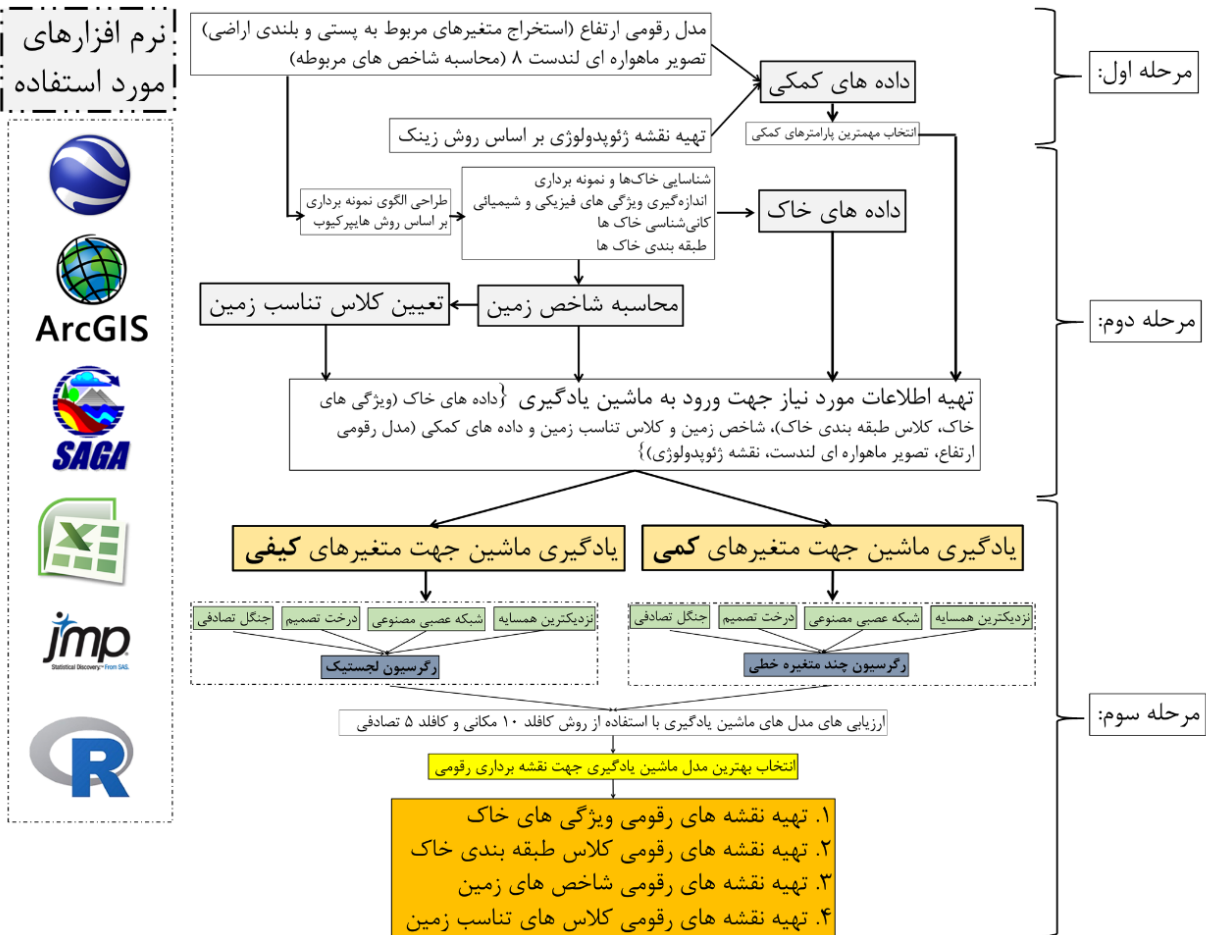
### فرآیند انتخاب مهم‌ترین آن‌ها

در پژوهش حاضر، از سه منبع برای تهیه داده‌های کمکی بهره گرفته شد: تصاویر ماهواره‌ای، مدل رقومی ارتفاع و نقشه ژئومورفولوژی. مدل رقومی ارتفاع با قدرت تفکیک مکانی ۱۰ متر مبنای کار قرار گرفت. در این مرحله، با استفاده از این مدل و زیربرنامه‌های تعریف‌شده‌ی استاندارد موجود در نرم‌افزار Arc-GIS10 و (System for Automated Geoscientific Analyses, SAGA) مشتقات مختلف مرتبه‌های اول و دوم مدل تهیه شد. شاخص‌های سنجش‌از دور نیز با استفاده از داده‌های

باند‌های ۱ تا ۶ ماهواره‌ی لندست ۸ (سال ۲۰۱۳) محاسبه شدند.

تفکیک واحدهای ژئومرفیک موجود در منطقه مطالعاتی با توجه به تفاوت رنگ، الگوی آبراهه‌ای، شیب، شکل شیب و ارتفاع که در تصاویر سه‌بعدی گوگل ارث قابل مشاهده هستند، بر اساس نقشه زمین‌شناسی و سامانه سلسله‌مراتبی زینک (زینک، ۱۹۸۹) انجام شد. این سامانه از بالاترین تا پایین‌ترین سطح طبقه‌بندی شامل زمین‌نما (Landscape)، پستی‌وبلندی (Relief) و یا قالب (Molding)، سنگ و مواد مادری (Lithology) و شکل زمین (Landform) است. به‌منظور نشان دادن رابطه بین زمین ریخت (Geoform) و خاک، خاک (های) مربوط به هر شکل زمین نیز معرفی گردید. در پایان، نقشه ژئومورفولوژی در سامانه اطلاعات جغرافیایی (GIS) رقومی گردید.

برای انتخاب پارامترهای موثر در مدل‌سازی، از روش «انتخاب ویژگی به ترتیب اهمیت» استفاده شد. به این معنی که تمام اطلاعات کمکی پس از استخراج از نرم‌افزار ساگا، وارد نرم‌افزار وکا (Weka) گردید و با استفاده از الگوریتم ReliefAttributeEval، اطلاعات کمکی مهم و مؤثر در پیش‌بینی ویژگی‌ها و کلاس‌های طبقه‌بندی خاک و رده‌بندی کیفی تناسب اراضی به ترتیب اهمیت، انتخاب و در مدل‌سازی مورد استفاده قرار گرفت. این الگوریتم، ارزش یک زیرمجموعه از ویژگی‌ها، با در نظر گرفتن توانایی پیش‌بینی فردی هر یک از آن‌ها و همچنین میزان فراوانی بین آن‌ها را ارزیابی می‌کند. در این الگوریتم، زیرمجموعه فوق‌الذکر که با ویژگی‌ها و کلاس‌های مورد پیش‌بینی همبستگی بالایی دارند، درحالی‌که همبستگی درونی آن‌ها کم است، ترجیح داده می‌شوند.



شکل ۳- مراحل اجرای پژوهش

### جمع‌آوری داده‌های خاک و عملکرد محصول

در نقشه‌برداری رقومی خاک، شبکه نمونه‌برداری (مکان‌های نمونه‌برداری) خاک برای ایجاد روابط کمی بین خاک و متغیرهای محیطی تأثیر بنیادین دارد. با توجه به اهمیت این موضوع، در تحقیق حاضر، برای افزایش کارایی مدل‌های نقشه‌برداری رقومی، نسبت به استفاده از روش نمونه‌برداری ابرمکعب لاتین (میناسنی و مک براتنی، ۲۰۰۶) اقدام شد. این تکنیک، یک روش نمونه‌برداری مونت کارلوی محدودشده و یک طرح تصادفی طبقه‌بندی‌شده است که به کمک توزیع چند متغیره، نمونه‌برداری قابل قبولی را باعث می‌شود. بر اساس تکنیک ابرمکعب لاتین، محل ۱۲۵ خاک رخ در منطقه مطالعاتی انتخاب شد. در این تحقیق از نرم افزار R برای مشخص کردن نقاط حفر خاکرخ‌ها استفاده و داده‌های استفاده شده برای وارد کردن به نرم افزار مشتعل بر DEM، RS و نقشه ژئومورفولوژی

بودند. موقعیت تمامی خاک رخ‌ها توسط سیستم موقعیت-یاب جهانی (GPS) ثبت گردید. مطالعات خاکشناسی برای تهیه نقشه‌های رقومی کلاس-های خاک و ویژگی‌های آن انجام گردید و فاصله پروفیل‌ها از یکدیگر بر اساس روش ابر مکعب لاتین تعیین شد. خاکرخ‌ها بر اساس راهنمای تشریح و نمونه‌برداری خاک‌ها در صحرا (شوئنبرگر و همکاران، ۲۰۱۲) تشریح و از تمامی افق‌های آن‌ها نمونه‌برداری صورت گرفت. سطوح مختلف ژئومورفولوژی نیز برای هر پروفیل یادداشت‌برداری شد.

### اندازه‌گیری مشخصات فیزیکی و شیمیایی خاک

برای برداشت نمونه‌های خاک از تمامی افق‌های خاکرخ خاک نمونه برداشته شد. بعضی خاکرخ‌های خاک کم عمق و بعضی‌ها عمیق‌تر بودند که نمونه‌برداری از تمامی افق‌های خاکرخ‌ها برای انجام رده‌بندی مورد نیاز بود. اما

تخمین نهایی برگزیده شد. لازم به ذکر می‌باشد که تفاوت ارزیابی متقابل تصادفی و ارزیابی تصادفی مکانی، در انتخاب مجموعه داده‌های خاک جهت آموزش و ارزیابی می‌باشد. در روش ارزیابی متقابل تصادفی، داده‌های به طور تصادفی به دو دسته تقسیم می‌شوند: داده هائی که برای آموزش به‌کار می‌روند و داده‌هائی که برای اعتبارسنجی یا آزمون مورد استفاده قرار می‌گیرند. در چنین شرایطی ممکن است داده‌های آموزش و آزمون از لحاظ مکانی در کنار یکدیگر در فاصله بسیار نزدیک قرار گیرند. اما در روش ارزیابی متقابل مکانی، در ابتدا منطقه مطالعاتی به تعدادی قطعه، تقسیم‌بندی شده و سپس داده‌های قرار گرفته در یک قطعه جهت ارزیابی مدل استفاده شده و بقیه قطعه‌ها برای آموزش مدل به‌کار می‌روند. در چنین شرایطی، می‌توان مطمئن شد که مدل آموزش دیده، در کل منطقه مورد ارزیابی قرار گرفته است (ولای و همکاران، ۲۰۱۸).

در این پژوهش، برای بیان کمی میزان خطا از معیارهای معمول در یادگیری ماشین استفاده گردید. بدین منظور، جهت بررسی دقت و صحت مقادیر پیش‌بینی شده داده‌های کمی خاک مانند درصد رس توسط مدل‌های یادگیری ماشین شاخص‌هایی از قبیل جذر میانگین مربع خطا (Root Mean Square Error: RMSE) و جذر میانگین مربع خطای نسبی (Relative RMSE: RMSE%)، ضریب تعیین ( $R^2$ ) و ضریب همبستگی ( $r$ ) محاسبه گردیدند. روابط ریاضی شاخص‌های مذکور بدین صورت می‌باشد:

$$RMSE = \sqrt{\frac{\sum(x' - x)^2}{n}}$$

در این رابطه،  $x'$  مقدار برآورد شده در نقطه‌ی  $x_i$  مقدار مشاهده در نقطه‌ی  $x_i$  و  $n$  تعداد نقاط است.

$$RMSE\% = \frac{RMSE}{\bar{x}} * 100$$

در این رابطه،  $\bar{x}$  میانگین مقادیر واقعی مشاهدات است.

نهایتاً برای اندازه‌گیری شوری خاک متوسط وزنی عمق ۵۰ سانتیمتری اندازه‌گیری و محاسبه گردید. نمونه‌های خاک، پس از هوا خشک شدن در محیط آزمایشگاه، کوبیده و از الک ۲ میلی‌متری عبور داده شدند. سپس ویژگی‌های خاک از جمله شوری خاک اندازه‌گیری گردید. شوری خاک یا همان EC را به طریقه‌ی عصاره‌گیری از گل اشباع به دست آوردیم و آن از طریق قرار دادن پروب EC متر در عصاره‌ی اشباع حاصل از گل به دست آمد.

### پردازش اطلاعات

متوسط وزنی عمق ۵۰ سانتیمتری هدایت الکتریکی خاک (شوری) برای نقشه‌برداری رقومی ویژگی‌های خاک بکار رفت. به منظور محاسبه متوسط وزنی، مقدار هر ویژگی در افق‌های خاک، ضخامت هر افق و ضریب وزنی عمق (گیوی، ۱۳۷۶) مورد استفاده قرار گرفت.

### یادگیری ماشین

#### الف- مدل‌های یادگیری ماشین، بکار رفته در این

##### تحقیق

در این پژوهش، جهت برآورد داده‌های کمی خاک مانند شوری خاک از مدل‌های شبکه عصبی مصنوعی، نزدیک-ترین همسایه، آنالیز تشخیصی، درخت تصمیم و جنگل تصادفی در نرم افزار R استفاده گردید. علاوه بر این، از مدل رگرسیون چند متغیره خطی و رگرسیون لجستیک جهت ترکیب مدل‌های مذکور نیز بهره گرفته شد.

#### ب- ارزیابی مدل‌های یادگیری ماشین

در این پژوهش، جهت ارزیابی مدل‌های یادگیری ماشین از دو روش بهره گرفته شد: ارزیابی متقابل تصادفی و ارزیابی تصادفی مکانی. در این نوع اعتبارسنجی‌ها، داده‌ها به  $K$  زیرمجموعه دسته‌بندی شدند. از این  $K$  زیرمجموعه، هر بار یکی برای اعتبارسنجی و  $K-1$  تای دیگر برای آموزش بکار رفتند. این روال،  $K$  بار تکرار شد و همه داده‌ها دقیقاً یک‌بار برای آموزش و یک‌بار برای اعتبارسنجی بکار رفتند. درنهایت، میانگین نتیجه این  $K$  بار اعتبارسنجی به‌عنوان یک

برخی از مدل‌های به‌کاربرده شده در نقشه‌برداری

رقومی خاک

رگرسیون لوجیستیک

در تحلیل‌های چندمتغیره، چنانچه متغیر وابسته، دو حالتی (دو ارزشی) باشد از مدل رگرسیون لوجیستیک استفاده می‌شود (هاسمر و لم‌شو، ۲۰۰۰). در این مدل، متغیر وابسته بیانگر دو وضعیت است که این وضعیت می‌تواند به‌صورت بلی یا خیر و یا وجود و عدم وجود وضعیت به‌صورت صفر و یک بیان گردد. مدل لوجیستیک، به‌طور موفقیت‌آمیزی در علوم خاک و ارتباط خاک با انواع ویژگی‌های محیطی استفاده می‌شود. این مدل، احتمال حضور یک کلاس خاک را به متغیرهای محیطی پیش‌بینی کننده ارتباط می‌دهد. هنگل و همکاران (۲۰۰۷) موفقیت رگرسیون لوجیستیک چند جمله‌ای را در پیش‌بینی گروه‌های مرجع خاک سامانه‌ی طبقه‌بندی جهانی (WRB) گزارش نمودند. نتایج آن‌ها نشان داد که موفقیت مدل، شدیداً به همبستگی کلاس‌های خاک با پیش‌بینی کننده‌ها (عوامل محیطی به‌کاربرده شده در مدل) بستگی دارد. جعفری و همکاران (۱۳۹۱) با استفاده از مدل رگرسیون لوجیستیک دوتایی، احتمال حضور افق سالیک را در سیمای اراضی پلایا، افق جیپسیک را در تپه‌های گچی و افق کلسیک را در مخروط‌افکنه، در منطقه‌ی زرنند کرمان پیش‌بینی و سپس با ترکیب حضور یا عدم حضور افق‌های مشخصه، کلاس‌های خاک را پیش‌بینی نمودند.

شبکه‌های عصبی مصنوعی

رایانه‌ها، دارای سرعت قابل توجهی نسبت به انسان می‌باشند؛ ولی قدرت تفکر ندارند و در هنگام کمبود اطلاعات و یا وجود داده‌های پرت، عملکرد آن‌ها به شدت کاهش می‌یابد و رضایت‌بخش نیست. از آنجاکه داده‌هایی که در دسترس پژوهشگران هستند، معمولاً محدود و گاه پرت می‌باشند؛ ایده‌ی شبکه‌های عصبی مصنوعی که یک سامانه‌ی محاسباتی هوشمند و دارای قدرت یادگیری است، مطرح گردید. مبنای شبکه‌های عصبی مصنوعی

روابط پیچیده‌ی ریاضی منطبق بر سامانه‌ی عصبی انسان است. در روش شبکه‌های عصبی مصنوعی، سه لایه‌ی ورودی، پنهان و خروجی وجود دارد و این لایه‌ها با یکدیگر در ارتباط هستند. در این روش، از طریق وزن دهی داده‌های ورودی توسط فرد و در حقیقت آموزش سیستم، وزن نهایی به دست می‌آید (صالحی و خادمی، ۱۳۸۷). بهرنس و همکاران (۲۰۰۵) برآورد واحدهای خاک را با استفاده از شبکه‌های عصبی مصنوعی و با به‌کارگیری داده‌های رقومی ارتفاع انجام داده و بیان نمودند که ترکیب تکنیک شبکه‌های عصبی مصنوعی و اطلاعات به‌دست‌آمده از مدل رقومی ارتفاع، علاوه بر صرفه‌جویی در زمان و هزینه‌ها، نتایج قابل قبولی ارائه می‌دهد.

روش خطی

مدل‌های خطی، شامل مدل رگرسیون برای پیش‌بینی ویژگی‌های خاک و پیش‌بینی کلاس‌های خاک می‌باشند. هنگل و همکاران (۲۰۰۲) بر اساس اطلاعات مدل رقومی ارتفاع و با به‌کار بردن مدل رگرسیون خطی، ضخامت افق سطحی و میزان ماده‌ی آلی را پیش‌بینی کردند. محنت کش و همکاران (۲۰۱۳) مهم‌ترین ویژگی‌های توپوگرافی (شیب، شاخص خیزی، سطح حوضه و شاخص انتقال رسوب) در تعیین عمق خاک را با استفاده از مدل رگرسیون خطی چندمتغیره در مناطق تپه‌ماهوری غرب ایران (منطقه‌ی کوه‌رنگ استان چهارمحال و بختیاری) مشخص نمودند.

رگرسیون درختی توسعه‌یافته

مهم‌ترین مزیت این روش نسبت به سایر روش‌ها، تفسیر آسان نتایج است. به همین دلیل، این روش به‌صورت گسترده‌ای در پیش‌بینی ویژگی‌های خاک به‌کار می‌رود. پاچ‌پسکی و همکاران (۲۰۰۱) با استفاده از مدل رگرسیون درختی و مشتقات مدل رقومی ارتفاع (شیب و انحنای خالص)، میزان شن، سیلت و ظرفیت نگهداری آب را پیش‌بینی نمودند. جعفری و همکاران (۱۳۹۱) با استفاده



انواع روش‌های داده‌کاوی در منطقه‌ی اردکان یزد، مدل آنالیز تشخیصی را برای پیش‌بینی کلاس‌های خاک به کار بردند و به نتایج مثبتی در این زمینه دست یافتند، لازم به ذکر است که در مدل آنالیز تشخیصی متغیرهای کمکی پلی-گونی (نقشه ژئومورفولوژی) را نمی‌توان وارد کرد (لیو و همکاران، ۲۰۰۸).

### مدل میانگین‌گیری نزدیک‌ترین همسایه K<sup>۱</sup>

میانگین‌گیری نزدیک‌ترین همسایه K، ارزش پاسخ برای یک مشاهده داده‌شده را با استفاده از پاسخ‌های مشاهداتی در همسایه‌های محلی پیش‌بینی می‌کند. می‌توان از این مدل با یک پاسخ قطعی برای طبقه‌بندی و یک پاسخ پیوسته برای پیش‌بینی استفاده کرد (نمس و همکاران، ۲۰۰۶). این روش، یک روش ناپارامتری است که بر اساس فاصله تا مشاهدات همسایه، پایه‌گذاری شده است. به این دلیل، میانگین‌گیری نزدیک‌ترین همسایه K قادر به طبقه‌بندی مشاهدات با استفاده از مرزهای ارزش پیش‌بینی نامنظم است. باین‌حال، الگوریتم، حساس به پیش‌بینی‌کننده‌های بی‌ربط است؛ بنابراین انتخاب پیش‌بینی‌کننده‌ها، قبل از اجرای مدل می‌تواند مفید باشد.

آنگویلی (۲۰۰۵) معتقد است، روش نزدیک‌ترین همسایگی K دارای یکی از ساده‌ترین و کارآمدترین الگوریتم‌های آموزش جهت پیش‌بینی داده‌ها است. بنا به نظر نمس و همکاران (۲۰۰۶) و ذوالفقاری و همکاران (۲۰۱۳)، در این تکنیک، تعداد بهینه‌ای از نمونه‌های موجود در یک مجموعه که واجد شبیه‌ترین ویژگی‌ها به نمونه هدف باشند در نظر گرفته می‌شود و سپس با توجه به فواصل اقلیدسی نمونه‌های هدف و هر یک از نمونه‌های بانک مرجع، وزن‌دهی عناصر آموزش دهنده 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

از مدل رگرسیون درختی توسعه‌یافته و متغیرهای کمکی (نقشه‌ی ژئومورفولوژی، اجزای سرزمین و شاخص‌های سنجش‌ازدور) احتمال حضور افق‌های مشخصه در منطقه‌ی زرد کرمان را تخمین زدند. نتایج نشان داد که نقشه‌ی ژئومورفولوژی، ابزاری مهم در روش‌های نقشه‌برداری رقومی خاک است که به افزایش دقت پیش‌بینی کمک شایانی می‌نماید.

### مدل آنالیز تشخیصی

تحلیل تشخیصی و رده‌بندی، تکنیک‌های چند متغیری هستند که با جدا کردن مجموعه‌های متمایز اشیاء (مشاهدات) و با تخصیص دادن اشیاء (مشاهدات) جدید به دسته‌های تعریف‌شده قبلی سروکار دارند. طبیعت تحلیل تشخیصی نسبتاً توضیحی است. در واقع تحلیل تشخیصی روشی است برای ترکیب کردن متغیرهای  $x_1, \dots, x_n$  برای ایجاد یک متغیر جدید  $y$ . این متغیر جدید که به نام تابع تشخیص شناخته شده، به گونه‌ای ساخته می‌شود که مقدار هر شرکت‌کننده در آن تا حد اکثر ممکن، افراد را در طبقه‌های مختلف تفکیک کند. برای بیان ریاضی مطلب فوق، اگر فرض شود که متغیرهای  $x_1, \dots, x_n$  متغیرهای تشریحی،  $y$  یک متغیر وابسته و از نوع کیفی چند سطحی و هدف از آنالیز تشخیصی یافتن تابعی خطی مثل  $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$  باشد، در حالتی که متغیر وابسته شامل  $k$  سطح باشد، هدف آن است که مشاهدات جدید  $(x_1, x_2, \dots, x_n)$  بر اساس  $y$  به یکی از  $k$  گروه نسبت داده شوند.

هدف تحلیل تشخیصی، کلاس‌بندی موارد به دو گروه یا بیشتر بر اساس مجموعه‌ای از خصوصیات است که مورد را توصیف می‌کنند. به‌طور کلی، یک مورد بر اساس مشاهداتی که از آن به دست آورده می‌شود، به گروه‌های از پیش تعیین‌شده تخصیص داده می‌شود  
تقی زاده مهرجردی و همکاران (۱۳۹۳) در مطالعه‌ای با عنوان نقشه برداری رقومی کلاس‌های خاک با استفاده از

<sup>۱</sup> K Nearest Neighbors averaging (KNN)

داده‌های پرت برخورد می‌کند. ج- توانایی درک و تحلیل روابط غیر خطی خصوصیات خاک را دارد د- هیچ فرضیه-ای برای توزیع داده‌ها نیاز ندارد. ه- بسیار راحت به روز می‌شود.

### مدل جنگل تصادفی

جنگل‌های تصادفی در حقیقت مجموعه‌هایی از درخت‌های پیش‌بینی‌کننده با احتمال و پراکندگی یکسان هستند. اساس این روش، انتخاب خطای داده‌های بزرگتر به عنوان خطای اصلی و ایجاد همبستگی بین خطاهای دیگر داده‌ها به ترتیب است. انتخاب خطای عمومی بر اساس بزرگترین داده‌ها، پیش‌بینی دقیق و قابل استنادی ارائه می‌دهد. زمانی این پیش‌بینی دقیق تر می‌شود که داده‌های ورودی به مدل بدون اعمال تغییرات باشد (بریمان، ۲۰۰۱). کامران و همکاران (۲۰۱۷) با استفاده از دو مدل رگرسیون لجستیک چند متغیره و جنگل‌های تصادفی اقدام به تهیه نقشه رقومی خاک کشور قبرس کردند که نتایج آنها نشان داد که مدل جنگل‌های تصادفی نسبت به مدل رگرسیون لجستیک چند متغیره از توانایی بالاتری برای پیش‌بینی کلاس‌های خاک منطقه برخوردار بوده و همچنین این پیش‌بینی با خطای کمتری نسبت به مدل دیگر همراه بوده است. سرینیواس و همکاران (۲۰۱۶) برای تهیه نقشه رقومی کربن آلی و غیر آلی در هندوستان اقدام به نمونه برداری ۱۱۹۸ نقطه به روش طبقه بندی تصادفی کردند و پس از انجام آزمایشات روی نمونه‌ها نتایج را به همراه برخی داده‌های کمکی از جمله شاخص پوشش گیاهی، داده‌های اقلیمی، پوشش زمین و ... در مدل جنگل‌های تصادفی برازش دادند. مقدار ۲۵ درصد از کل نقاط نمونه برداری به عنوان داده‌های اعتبار سنجی مدل انتخاب شد. نتایج آنها نشان داد که مقدار کربن آلی، کربن غیر آلی و مقدار مجموع ذخیره کربن خاک‌ها به ترتیب، ۲۲/۷۳،

فواصل اقلیدسی بین خاک مورد آزمون و تمامی خاک‌های مرجع محاسبه شده، سپس تعداد  $K$  عدد از نزدیک‌ترین خاک‌ها بر اساس فاصله به دست آمده، وزندهی می‌شوند. در انتها با توجه وزن هر یک از خاکها در مجموعه‌ای با  $K$  عدد خاک (نزدیک‌ترین همسایه)، بر آوردی از داده‌های هدف که مورد آزمون واقع شده‌اند، صورت می‌پذیرد. ایوبی و همکاران (۱۳۹۵) در مطالعه‌ای با عنوان مقایسه روش‌های  $K$  نزدیکترین همسایگی و شبکه عصبی مصنوعی برای پهنه‌بندی رقومی شوری خاک در منطقه‌ی چاه افضل اردکان، به این نتیجه دست یافتند که مدل  $K$  نزدیکترین همسایگی دارای دقت بیشتری نسبت به شبکه عصبی مصنوعی به منظور پیش‌بینی شوری خاک است، به طوری که این مدل به خوبی توانسته ارتباط قوی بین داده‌های شوری خاک و اطلاعات محیطی برقرار کند.

### مدل تحلیل درخت تصمیم<sup>۲</sup>

مدل درختی طبقه‌بندی و رگرسیون، یک روش ناپارامتری الگوریتمی است که قادر است متغیرهای کمی یا متغیرهای طبقه‌بندی شده را بر اساس مجموعه‌ای از متغیرهای پیش‌بینی کننده کمی و کیفی برآورد کند. در این روش، مجموعه‌ای از شرط‌های منطقی به صورت یک الگوریتم با ساختار درختی برای طبقه‌بندی یا پیش‌بینی کمی یک متغیر به کار می‌رود. ایجاد درخت تصمیم دو مرحله دارد. مرحله اول، مرحله ایجاد و رشد درخت است. این مرحله شامل پیوند<sup>۳</sup> و انشعاب<sup>۴</sup> است. مرحله دوم، مرحله توقف و هرس<sup>۵</sup> است. هدف این مرحله به حداقل رساندن خطای پیش‌بینی است (اسکال و همکاران، ۲۰۰۵).

بعضی از مزایای استفاده از درخت تصمیم که مور و همکاران (۱۹۹۱) در مقایسه با سایر روش‌های طبقه‌بندی بیان کردند عبارت‌اند از:

الف- وقتی داده‌ها کمی و کیفی هستند، تفسیر کردن آنها راحت‌تر است. ب- بسیار راحت تر با داده‌های گمشده یا

<sup>۴</sup> Splitting

<sup>۵</sup> Pruning

<sup>۲</sup> Decision Tree Analysis (DTA)

<sup>۳</sup> Merging

کردند که نتایج اعتبارسنجی مدل‌ها نشان داد مدل جنگل‌های تصادفی توانسته به طور معنی‌داری دقت بیشتری نسبت به مدل رگرسیون خطی چندگانه از خود نشان دهد. هیونگ و همکاران (۲۰۱۴) اقدام به تهیه نقشه مواد مادری خاک با استفاده از متغیرهای کمکی و مدل جنگل‌های تصادفی کردند که نتایج آن‌ها نشان داد زمانی که از متغیرهای کمکی مستخرج از تصاویر ماهواره‌ای و متغیرهای کمکی مدل رقومی ارتفاع به عنوان داده‌های ورودی به مدل استفاده شد، ضریب کاپای اعتبارسنجی مدل ۸۹/۶ درصد استخراج گردید.

### نتایج و بحث

متغیرهای کمکی مهم در پیش‌بینی شوری خاک، در منطقه مورد مطالعه به ترتیب اهمیت در جدول ۱ ارائه شده است.

۱۲/۸۳ و ۳۵/۵۵ پکا گرم می‌باشد. هیونگ و همکاران (۲۰۱۶) در مطالعه‌ای با عنوان مرور و مقایسه تکنیک‌های یادگیری ماشینی برای اهداف طبقه‌بندی در تهیه نقشه رقومی خاک، به این نکته اشاره کردند که مدل‌های جنگل‌های تصادفی و مدل‌های درختی به دلیل سرعت در پردازش پارامتری و سهولت تفسیر داده‌های خروجی، مدل‌های مناسبی جهت استفاده در اهداف طبقه‌بندی هستند.

پهلوان راد و همکاران (۲۰۱۶) داده‌های حاصل از نتایج آزمایشگاهی در مطالعات میدانی حدود ۸۵۰۰۰ هکتار از اراضی استان گلستان در ایران را به منظور تهیه نقشه رقومی خاک منطقه تحت دو مدل جنگل‌های تصادفی و رگرسیون لجستیک چند متغیره برآزش دادند که نتایج آن‌ها نشان داد مدل جنگل‌های تصادفی به طور معنی‌داری نسبت به مدل رگرسیون لجستیک چند متغیره توانایی پیش‌بینی سری‌های خاک منطقه را داشته است. هنگل و همکاران (۲۰۱۵) در بخشی از آفریقا اقدام به تهیه نقشه رقومی خصوصیات خاک با دو مدل رگرسیون چندگانه خطی و جنگل‌های تصادفی

جدول ۱- متغیرهای کمکی مهم در پیش‌بینی شوری خاک، به ترتیب اهمیت. برای هر ویژگی، از چپ به راست، از اهمیت متغیر کمکی کاسته می‌شود. درجه‌بندی اهمیت با به‌کارگیری الگوریتم ReliefAttributeEval نرم‌افزار WEKA مشخص گردید

ویژگی‌ها	متغیرهای کمکی مهم به ترتیب اهمیت
EC	Geomorphology, Valley depth, MRRTF, Topographic wetness index, Aspect, DEM, Catchment slope, Relative slope position, Slope, LS Factor

ترکیبی (MLR) صورت گرفته است. در میان مدل‌های استفاده شده برای پیش‌بینی شوری، مدل رگرسیون چند متغیره خطی ترکیبی (MLR) با ضریب تعیین ۰/۶۱۱ و ریشه دوم متوسط مربعات خطا ۰/۰۳۲ از بیشترین دقت برای پیش‌بینی برخوردار بوده است.

نتایج ارزیابی صحت پیش‌بینی شوری خاک‌های مورد مطالعه، با استفاده از روش اعتبارسنجی کافلد ۱۰ مکانی، در جدول ۲ ارائه شده است. این پیش‌بینی توسط مدل‌های نزدیک‌ترین همسایه K (KNN)، تحلیل درخت تصمیم (DTA)، شبکه عصبی مصنوعی (ANN)، جنگل تصادفی (R.F.) و رگرسیون چند متغیره خطی

جدول ۲- نتایج ارزیابی صحت پیش‌بینی شوری خاک، با استفاده از روش اعتبارسنجی کافلد ۱۰ مکانی. این پیش‌بینی توسط مدل‌های نزدیک‌ترین همسایه K (KNN)، تحلیل درخت تصمیم (DTA)، شبکه عصبی مصنوعی (ANN)، جنگل تصادفی (R.F.) و رگرسیون چند متغیره خطی ترکیبی (MLR) صورت گرفته است. میانگین داده‌های واقعی شوری خاک در منطقه مورد مطالعه برابر با ۰/۱۸ دسی سیمنس بر متر است

مدل	RMSE	MAE	R <sup>2</sup>
KNN	۰/۰۴۳	۰/۰۳۶	۰/۲۸۰
DTA	۰/۰۳۳	۰/۰۲۷	۰/۵۴۷
ANN	۰/۰۳۳	۰/۰۲۶	۰/۵۵۴
RF	۰/۰۳۲	۰/۰۲۵	۰/۶۰۹
MLR	۰/۰۳۲	۰/۰۲۵	۰/۶۱۱

تصمیم (Decision Tree Analysis, DTA)، رگرسیون چند متغیره خطی (Multi Linear Regression, MLR) و نزدیک‌ترین همسایه K (K Nearest Neighbor)، مدل نزدیک‌ترین همسایه K (KNN (Averaging)) با ضریب تعیین ۰/۹۴۰۱ و ریشه دوم متوسط مربعات خطا ۰/۰۱۱۶ از بیشترین دقت برای پیش‌بینی برخوردار بوده است.

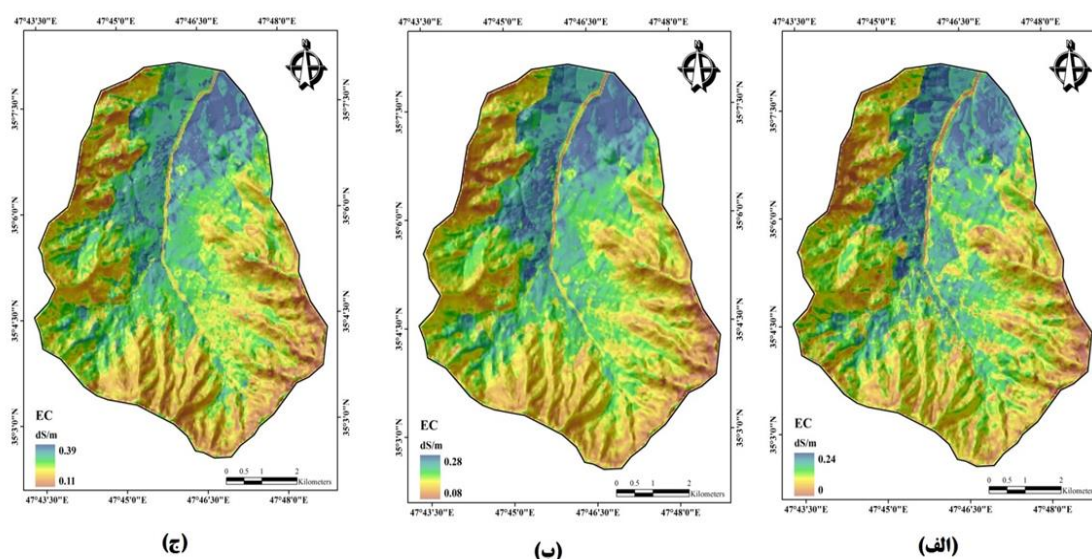
در جدول ۲، دقت پیش‌بینی شوری، از بالا به طرف پائین افزایش می‌یابد. کمترین دقت پیش‌بینی به مدل نزدیک‌ترین همسایه K (KNN)، اختصاص یافته است. این در حالی است که با استفاده از روش اعتبارسنجی کافلد ۵ تصادفی (جدول ۳)، از بین مدل‌های شبکه عصبی مصنوعی (Artificial Neural Network, ANN)، تحلیل درخت

جدول ۳- نتایج ارزیابی صحت پیش‌بینی شوری خاک، با استفاده از روش اعتبارسنجی کافلد ۵ تصادفی. این پیش‌بینی توسط مدل‌های شبکه عصبی مصنوعی (Artificial Neural Network, ANN)، تحلیل درخت تصمیم (Decision Tree Analysis, DTA)، رگرسیون چند متغیره خطی (Multi Linear Regression, MLR) و نزدیک‌ترین همسایه K (K Nearest Neighbor)، صورت گرفته است

مدل	RMSE	<sup>N</sup> RMSE=RMSE/Mean real	r <sup>C</sup>	R <sup>2</sup>	ME
ANN	۰/۰۱۴۰	۰/۰۷۷۳۸	۰/۹۰۹۳	۰/۸۶۲۹	-۰/۰۰۰۵۲
DTA	۰/۰۲۱۷	۰/۱۱۹۹۴	۰/۸۷۶۸	۰/۷۸۱	.
MLR	۰/۰۳۲۸	۰/۱۸۱۲۹	۰/۸۰۲۵	۰/۶۷۰۱	.
KNN	۰/۰۱۱۶	۰/۰۶۴۱۱	۰/۹۶۹۱	۰/۹۴۰۱	۰/۰۰۱۴۲

دارد (شکل ۴). در شکل ۴، نقشه ب، پیش‌بینی مقدار میانگین، نقشه الف، پیش‌بینی با دامنه پراکندگی ۵ درصد کمتر از میانگین و نقشه ج، پیش‌بینی با دامنه پراکندگی ۵ درصد بیشتر از میانگین را نشان می‌دهند.

مقدار هدایت الکتریکی خاک بین ۰/۱۱۴ و ۰/۳۸۴ دسی زیمنس بر متر در تغییر است. مقدار این ویژگی خاک از کوه به طرف اراضی پای کوهی و پائین‌ترین دشت‌های آبرفتی (QC) و از جنوب به شمال افزایش می‌یابد. بیشترین مقدار شوری در خاک‌های شمال شرق حوضه آبخیز وجود



شکل ۴- نقشه‌ی پیش‌بینی هدایت الکتریکی خاک با استفاده از مدل رگرسیون چند متغیره خطی ترکیبی. نقشه ب، پیش‌بینی مقدار میانگین، نقشه الف، پیش‌بینی با دامنه پراکندگی ۵ درصد کمتر از میانگین و نقشه ج، پیش‌بینی با دامنه پراکندگی ۵ درصد بیشتر از میانگین را نشان می‌دهند

افزایش و به دلایلی که قبلاً در مورد اثر شاخص خیسی گفته شد، شوری نیز افزایش می‌یابد. شیب حوضه از نظر تأثیرگذاری بر شوری در درجه هفتم اهمیت قرار دارد. در پشته‌های شیب که شیب بیشتر است، شوری نسبت به پای شیب که از شیب کمتری برخوردار است، کمتر است (اثر موقعیت نسبی شیب بر مقدار شوری). با افزایش شیب، شوری کاهش می‌یابد (اثر مقدار شیب بر شوری) (جدول ۳)

#### جمع‌بندی نتایج پیش‌بینی ویژگی‌های خاک توسط مدل‌های استفاده شده در نقشه‌برداری رقمی این ویژگی‌ها

جهت برآورد ویژگی‌های خاک، در مطالعه حاضر دو حالت مختلف مورد بررسی قرار گرفت. در حالت اول، مدل‌های شبکه عصبی مصنوعی، تحلیل درخت تصمیم و رگرسیون چند متغیره خطی برای پیش‌بینی مورد استفاده قرار گرفتند. همچنین جهت ترکیب نتایج مدل‌ها، از مدل نزدیک‌ترین همسایه استفاده گردید. نتایج این پیش‌بینی‌ها با استفاده از روش ارزیابی کافلد تصادفی ۵ مورد ارزیابی قرار گرفتند. در مرحله دوم، مدل‌های تحلیل درخت تصمیم، شبکه عصبی مصنوعی، نزدیک‌ترین همسایه و جنگل

متغیرهای کمکی مهم در پیش‌بینی شوری خاک به ترتیب اهمیت عبارت‌اند از: ژئومورفولوژی، عمق دره، شاخص همواری قله برآمدگی با درجه تفکیک بالا، شاخص خیسی، جهت شیب، مدل رقمی ارتفاع، شیب حوضه، موقعیت نسبی شیب، مقدار شیب و طول شیب (جدول ۱) ژئومورفولوژی مهم‌ترین متغیر کمکی در پیش‌بینی مقدار شوری خاک محسوب می‌شود. قبلاً گفته شد که از کوه به طرف اراضی پای کوهی مقدار شوری افزایش می‌یابد (شکل ۴). در این جهت، مقدار رس خاک افزایش یافته و در نتیجه آب بیشتری در خاک نگه‌داشته شده و شاخص خیسی نیز افزایش می‌یابد. با تبخیر این آب، غلظت آنیون‌ها و کاتیون‌های موجود در آن افزایش یافته و خاک شورتر می‌شود. از نظر اهمیت برای پیش‌بینی مقدار شوری خاک، شاخص خیسی در درجه چهارم قرار دارد (جدول ۳). اهمیت جهت شیب در سطح پنجم است. شیب‌های شمالی و غربی از شوری کمتری برخوردار می‌باشند. مقدار شوری بستگی به ارتفاع از سطح دریا نیز دارد (اثر مدل رقمی ارتفاع). با افزایش ارتفاع از سطح دریا، از میزان شوری کاسته شده است (شکل ۴). آب بیشتر در اراضی مسطح و گود جمع می‌شود. با کم شدن شیب حوضه، خیسی خاک

نزدیک‌ترین همسایه K، پائین‌ترین عملکرد را در پیش‌بینی همه ویژگی‌های خاک دارا بوده است.

(۲) بر اساس روش اعتبارسنجی کافلد ۵ تصادفی، دقت و صحت مدل نزدیک‌ترین همسایه بهترین عملکرد را در پیش‌بینی همه ویژگی‌های خاک دارا بوده است.

(۳) به طور کلی، بر اساس روش‌های ارزیابی کافلد ۱۰ مکانی و کافلد ۵ تصادفی، می‌توان بیان داشت که روش‌های ترکیبی دارای قابلیت بیشتری می‌باشند. این بدین معنی است که ترکیب نتایج پیش‌بینی سایر مدل‌ها می‌تواند نقشه‌هایی با دقت بالاتر را تولید کند.

تصادفی برای پیش‌بینی به‌کار رفتند. همچنین جهت ترکیب نتایج مدل‌ها در این حالت، از روش رگرسیون چند متغیره خطی ترکیبی استفاده گردید. با استفاده از روش ارزیابی کافلد مکانی ۱۰، نتایج پیش‌بینی مورد ارزیابی قرار گرفتند. نتایج به طور کلی نشان داد که:

(۱) بر اساس روش اعتبارسنجی کافلد ۱۰ مکانی، بهترین مدل برای پیش‌بینی تمام ویژگی‌های خاک، رگرسیون چند متغیره خطی ترکیبی است. برای همه ویژگی‌های خاک، مدل جنگل تصادفی، از نظر قدرت پیش‌بینی، رتبه دوم را به خود اختصاص می‌دهد. به‌طور کلی،

## فهرست منابع

۱. ایوبی، ش.، تقی زاده، ر.، نمازی، ر.، ذوالفقاری، ع.، و روستایی صدرآبادی، ف. ۱۳۹۵. مقایسه روش‌های K نزدیکترین همسایگی و شبکه عصبی مصنوعی برای پهنه‌بندی رقومی شوری خاک در منطقه‌ی چاه افضل اردکان. نشریه علوم آب و خاک (علوم و فنون کشاورزی و منابع طبیعی). سال بیستم. شماره ۷۶. صص. ۷۱-۵۹.
۲. تقی‌زاده مهرجردی، ر.، سرمیدیان، م.، امید، م.، تومانیان، ن.، روستا، م.، و رحیمیان، م.، ۱۳۹۳. نقشه برداری رقومی کلاس‌های خاک با استفاده از انواع روش‌های داده‌کاوی در منطقه‌ی اردکان استان یزد. مجله مهندسی زراعی (مجله علمی کشاورزی)، جلد ۳۷، شماره ۲. صص. ۱۱۵-۱۰۱.
۳. جعفری، ا.، خادمی، ح.، و ایوبی، ش. ۱۳۹۱. نقشه‌برداری رقومی افق‌های مشخصه و گروه‌های بزرگ خاک در منطقه زرنند کرمان. مجله علوم و فنون کشاورزی و منابع طبیعی، علوم آب و خاک، شماره ۶۲. صص. ۱۷۷ تا ۱۹۱.
۴. حریری، ع. ۱۳۷۴. نگرشی بر خاستگاه گروهی از سنگ‌های دگرگونه گستره قروه. پایان‌نامه کارشناسی ارشد، دانشگاه شهید بهشتی، تهران.
۵. حسینی، م. ۱۳۷۶. شرح نقشه زمین‌شناسی ۱:۱۰۰۰۰۰ چهار گوش قروه (پیوست نقشه)، سازمان زمین‌شناسی و اکتشاف معدنی کشور.
۶. صالحی، م. ح.، و خادمی، ح. ۱۳۸۷. مبانی نقشه‌برداری خاک. انتشارات جهاد دانشگاهی اصفهان. ۲۱۰ صفحه.
۷. گیوی، ج. ۱۳۷۶. ارزیابی کیفی تناسب اراضی برای نباتات زراعی و باغی، مؤسسه تحقیقات خاک و آب. نشریه فنی شماره ۱۰۱۵، ۱۰۰ صفحه.
8. Anguilli, F. 2005. Fast condensed nearest neighbor rule. *Proceedings of the 22nd International Conference on Machine Learning*, Bonn, Germany.
9. Asfaw, E.; Suryabhagavan, K.; Argaw, M. Soil salinity modeling and mapping using remote sensing and GIS: The case of Wonji sugar cane irrigation farm, Ethiopia. *J. Saudi Soc. Agric. Sci.* 2018, 17, pp. 250–258.

10. Barbouchi, M.; Abdelfattah, R.; Chokmani, K.; Aissa, N.B.; Lhissou, R.; El Harti, A. Soil salinity characterization using polarimetric InSAR coherence: Case studies in Tunisia and Morocco. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2014, 8, pp. 3823–3832.
11. Behrens, T. Forster, H. Scholten, T. Steninrucken, U. Spies, E. and Goldschmitt, M. 2005. Digital soil mapping using artificial neural networks. *Journal of Plant Nutrition and Soil Science.* 168: pp. 21-33.
12. Bouksila, F.; Bahri, A.; Berndtsson, R.; Persson, M.; Rozema, J.; Van der Zee, S.E. Assessment of soil salinization risks under irrigation with brackish water in semiarid Tunisia. *Environ. Exp. Bot.* 2013, 92, pp.176–185.
13. Breiman, L. 2001. "Random forests" Machine learning. 45, pp. 5-32.
14. Brovelli, M.A.; Sun, Y.; Yordanov, V. 2020. Monitoring forest change in the amazon using multi-temporal remote sensing data and machine learning classification on Google Earth Engine. *ISPRS Int. J. Geo-Inf.*, 9, 580.
15. Camera, C. Z. Zomeni, J.S. Noller, A.M. Zissimos, I.C. Christoforou, B. and A. Bruggeman. 2017. A high resolution map of soil types and physical properties for Cyprus: A digital soil mapping optimization. *Geoderma* 285, pp.35-49
16. Cañedo-Argüelles, M.; Kefford, B.J.; Piscart, C.; Prat, N.; Schäfer, R.B.; Schulz, C.-J. 2013. Salinisation of rivers: An urgent ecological issue. *Environ. Pollut.*, 173, pp.157–167.
17. Dehaan, R.; Taylor, G. Field-derived spectra of salinized soils and vegetation as indicators of irrigation-induced soil salinization. *Remote Sens. Environ.* 2002, 80, pp. 406–417.
18. Dent, D.; Young, A. Soil Survey and Land Evaluation; George Allen & Unwin: Sydney, NSW, Australia, 1981.
19. Ding, J.-L.; Wu, M.-C.; Liu, H.-X.; Li, Z.-G. Study on the soil salinization monitoring based on synthetical hyperspectral index. *Spectrosc. Spectr. Anal.* 2012, 32, pp. 1918–1922.
20. Dong, J.; Xiao, X.; Menarguez, M.A.; Zhang, G.; Qin, Y.; Thau, D.; Biradar, C.; Moore III, B. Mapping paddy rice planting area in northeastern Asia with Landsat 8 images, phenology-based algorithm and Google Earth Engine. *Remote Sens. Environ.* 2016, 185, pp. 142–154.
21. El Harti, A.; Lhissou, R.; Chokmani, K.; Ouzemou, J.-E.; Hassouna, M.; Bachaoui, E.M.; El Ghmari, A. Spatiotemporal monitoring of soil salinization in irrigated Tadla Plain (Morocco) using satellite spectral indices. *Int. J. Appl. Earth Obs. Geoinf.* 2016, 50, pp. 64–73.
22. Elshewy, M.A., Mohamed, M.H.A. & Refaat, M. 2024. Developing a Soil Salinity Model from Landsat 8 Satellite Bands Based on Advanced Machine Learning Algorithms. *J Indian Soc Remote Sens* 52, pp.617–632.
23. Gorji, T.; Sertel, E.; Tanik, A. 2017. Monitoring soil salinity via remote sensing technology under data scarce conditions: A case study from Turkey. *Ecol. Indic.*, 74, pp. 384–391.
24. Hengl, T. Rossiter, D.G. and Husnjak, S. 2002. Mapping soil properties from an existing national soil data set using freely available ancillary data. 17th World Congress of Soil Science. Thailand.
25. Hengl, T. Toomanian, N. Reuter, H. and Malakouti, M. J. 2007. Methods to interpolate soil categorical variables from profile observations: Lessons from Iran. *Geoderma* 140: pp. 417-427.
26. Hengl, T.; Heuvelink, G.B.M.; Kempen, B.; Leenaars, J.G.B.; Walsh, M.G.; Shepherd, K.D.; Sila, A.; MacMillan, R.A.; Jesus, J.; Tamene, L.; et al. 2015. Mapping Soil Properties of Africa at 250 m Resolution: Random Forests Significantly Improve Current Predictions. *PLoS ONE* , 10, 0125814.

27. Heung, B. Bulmer, C.E. and Schmidt, M.G. 2014. Predictive soil parent material mapping at a regional-scale: a random forest approach. *Geoderma*. 214, pp. 141-154.
28. Heung, B. H.C. Ho, J. Zhang, A. Knudby, C.E. Bulmer, and Schmidt, M.G. 2016. An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma*. 265, pp. 62-77
29. Hosmer, D.W. and Lemeshow, S. (2000). Applied logistic regression. John Wiley & Sons. New York. pp 392.
30. Jiang, H.; Shu, H. 2019. Optical remote-sensing data based research on detecting soil salinity at different depth in an arid-area oasis, Xinjiang, China. *Earth Sci. Inform.* 12, pp. 43–56.
31. Li, J.; Pu, L.; Han, M.; Zhu, M.; Zhang, R.; Xiang, Y. 2014. Soil salinization research in China: Advances and prospects. *J. Geogr. Sci.* 24, pp. 943–960.
32. Liu, J. Pattey, E. Nolin, M.C. Miller, J.R. and Ka, O. 2008. Mapping within-field soil drainage using remote sensing, DEM and apparent soil electrical conductivity. *Geoderma*, 143, pp. 261–272.
33. Ma, S.; He, B.; Ge, X.; Luo, X. 2023. Spatial prediction of soil salinity based on the Google Earth Engine platform with multitemporal synthetic remote sensing images. *Ecol. Inform.* 75, 102111.
34. Mehnatkesh, A. Ayoubi, S. Jalalian, A. and Sahrawat, K. 2013. Relationships between Soil depth and terrain attributes in a semi arid hilly region in western Iran. *Journal of Mountain Science* 10, pp.163-172.
35. Minasny, B. and McBratney, A. B. 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computers and Geosciences* 32, pp.1378-1388.
36. Moore, I. D. Grayson, R.B. and Ladson, A.R. 1991. Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. *Hydrological processes*. 5, pp. 3-30.
37. Mougnot, B.; Pouget, M.; Epema, G. 1993. Remote sensing of salt affected soils. *Remote Sens. Rev.* 7, pp. 241–259.
38. Nanni, M.R.; Demattê, J.A.M. 2006. Spectral reflectance methodology in comparison to traditional soil analysis. *Soil Sci. Soc. Am. J.* 70, pp. 393–407.
39. Nemes, A. Rawls, W.J. and Pachepsky, Y.A. 2006. Use of the nonparametric nearest neighbor approach to estimate soil hydraulic properties. *Soil Science Society of America Journal*, 70, pp. 327–336.
40. Pachepsky, Y.A. Timlin, D.J. and Rawls, W.J. 2001. Soil water retention as related to topographic variables. *Soil Science Society of America Journal* 65, pp. 1787–1795.
41. Pahlavan Rad, M. R. Khormali, F. Toomanian, N. Brungard, C.W. Kiani, F. Komaki, C.B. and Bogaert, P. 2016. Legacy soil maps as a covariate in digital soil mapping: A case study from Northern Iran. *Geoderma*. 279, pp.141-148.
42. Parastatidis, D.; Mitraka, Z.; Chrysoulakis, N.; Abrams, M. 2017. Online global land surface temperature estimation from Landsat. *Remote Sens.*, 9, 1208.
43. Schoeneberger, P.J. Wysocki, D.A. Benham, E.C. Soil Survey Staff. 2012. Field book for describing and sampling soils. 3rd Version. Natural Resources Conservation Service. *National Soil Survey Center*. Lincoln, NE.
44. Scudiero, E.; Skaggs, T.H.; Corwin, D.L. 2017. Simplifying field-scale assessment of spatiotemporal changes of soil salinity. *Sci. Total Environ.* 587, pp. 273–281.
45. Scull, P. Franklin, J. and Chadwick, O.A. 2005. The application of classification of tree analysis to soil type prediction in a desert landscape. *Eco. Model.* 181, pp. 1-15.
46. Shi, H.; Hellwich, O.; Luo, G.; Chen, C.; He, H.; Ochege, F.U.; Van de Voorde, T.; Kurban, A.; De Maeyer, P. A. 2021. global meta-analysis of soil



- salinity prediction integrating satellite remote sensing, soil sampling, and machine learning. *IEEE Trans. Geosci. Remote Sens.*, 60, 4505815.
47. Sreenivas, K. Dadhwal, V.K. Kumar, S. Harsha, G.S. Mitran, T. Sujatha, G. and Ravisankar, T. 2016. Digital mapping of soil organic and inorganic carbon status in India. *Geoderma*. 269, pp.160-173.
  48. Tian, A.; Fu, C.; Yau, H.-T.; Su, X.-Y.; Xiong, H. 2019. A new methodology of soil salinization degree classification by probability neural network model based on centroid of fractional lorenz chaos self-synchronization error dynamics. *IEEE Trans. Geosci. Remote Sens.*, 58, pp. 799–810.
  49. Triki Fourati, H.; Bouaziz, M.; Benzina, M.; Bouaziz, S. 2017. Detection of terrain indices related to soil salinity and mapping salt-affected soils using remote sensing and geostatistical techniques. *Environ. Monit. Assess.* 189, 177.
  50. Valavi, R. Elith, J. José, J. Lahoz, M. Gurutzeta, G. 2018. Block CV: an R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Biorxiv*.
  51. Wang, H.; Jia, G. 2012. Satellite-based monitoring of decadal soil salinization and climate effects in a semi-arid region of China. *Adv. Atmos. Sci.* 29, pp. 1089–1099.
  52. Wang, Q.; Li, P.; Chen, X. 2012. Retrieval of soil salt content from an integrated approach of combining inversed reflectance model and regressions: An experimental study. *IEEE Trans. Geosci. Remote Sens.* 50, pp. 3950–3957.
  53. Wang, Z.; Zhang, F.; Zhang, X.; Chan, N.W.; Ariken, M.; Zhou, X.; Wang, Y. 2021. Regional suitability prediction of soil salinization based on remote-sensing derivatives and optimal spectral index. *Sci. Total Environ.* 775, 145807.
  54. Wu, D.; Jia, K.; Zhang, X.; Zhang, J.; Abd El-Hamid, H.T. 2021. Remote sensing inversion for simulation of soil salinization based on hyperspectral data and ground analysis in Yinchuan, China. *Nat. Resour. Res.* 30, pp. 4641–4656.
  55. Xiong, J.; Thenkabail, P.S.; Tilton, J.C.; Gumma, M.K.; Teluguntla, P.; Oliphant, A.; Congalton, R.G.; Yadav, K.; Gorelick, N. 2017. Nominal 30-m cropland extent map of continental Africa by integrating pixel-based and object-based algorithms using Sentinel-2 and Landsat-8 data on Google Earth Engine. *Remote Sens.*, 9, 1065.
  56. Zhang, Q.; Li, L.; Sun, R.; Zhu, D.; Zhang, C.; Chen, Q. 2020. Retrieval of the soil salinity from Sentinel-1 Dual-Polarized SAR data based on deep neural network regression. *IEEE Geosci. Remote Sens. Lett.*, 19, 4006905.
  57. Zhang, Z.; Fan, Y.; Zhang, A.; Jiao, Z. 2022. Baseline-Based Soil Salinity Index (BSSI): A Novel Remote Sensing Monitoring Method of Soil Salinization. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 16, pp. 202–214.
  58. Zinck, J.K. 1989. Physiography and soil. Lecture notes for K6 course. Soil Division. *ITC. Enschede, The Netherlands*. 156 p.
  59. Zolfaghari, A. A. Tirgar Soltani, M. T. Dyck, M. and Weldeyohannes, A. 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.

## Digital Mapping of Soil Salinity Using Auxiliary Data and Machine Learning Models in Badr Watershed, Kurdistan Province

M. Zarinbahador \*and R. Sharifi

Agriculture and Natural Resources Research and Education Center of Tehran Province, Agriculture Research, Education and Extension Organization (AREEO), Tehran, Iran. [moslem.zarini@ymail.com](mailto:moslem.zarini@ymail.com)  
Faculty member of Agriculture and Natural Resources Research and Education Center of Tehran Province, Agricultural Research, Extension, and Education Organization (AREEO), Tehran, Iran. [rahmansharifi@yahoo.com](mailto:rahmansharifi@yahoo.com)

«Research Article»

Received: July 14, 2024 and Accepted: October 16, 2024

### Abstract

Use of remote sensing and machine learning techniques are increasingly recognized as cost-effective methods for displaying soil salinity maps. In this study, Landsat 8 satellite data and sophisticated machine learning techniques were used to map and evaluate soil salinity levels in the Badr Watershed. In this study, several Machine Learning techniques were used to predict salinity values in Badr Watershed. These algorithms included K-nearest neighbor (KNN), decision tree analysis (DTA), artificial neural network (ANN), random forest (RF) and mixed multivariate linear regression (MLR). In the first stage, auxiliary data such as Landsat 8 satellite images of the region and a digital elevation model with a spatial resolution of 10 meters were prepared from the country's Mapping Organization. The geological map of Qorveh was prepared from the geological site of the country, and the geological map of the Badr Watershed was extracted from it and digitized in the environment of the geographic information system. The geomorphological map was drawn and the location of the observation points was determined. Then, modeling was done, digital maps of soil classes and characteristics were prepared and the models were evaluated. Based on the Latin Supercube Technique, 125 outcrops were selected and excavated in the study area. After air-drying in the laboratory, the soil samples were pounded and passed through a 2 mm sieve. Then, soil salinity was measured. In order to estimate soil characteristics, two different conditions were investigated in this study. In the first case, ANN models, DTA and linear MLR were used for prediction. Also, to combine the results of the models, the nearest KNN was used. The results showed that the important auxiliary variables in predicting soil salinity, in order of importance, were geomorphology, depth of the valley, smoothness index of the ridge with a high degree of resolution, wetness index, slope direction, digital height model, basin slope, relative position of the slope, slope amount and slope length. Also, the results of the evaluation showed that among the models used to predict salinity, the combined MLR model with a coefficient of determination of 0.611 and a square root mean square error of 0.032 had the highest accuracy for prediction.

**Keyword:** Digital mapping, Decision Tree Analysis, Machine learning, Landsat 8, Multivariate Linear Regression

---

\*-Corresponding author's email :[moslem.zarini@ymail.com](mailto:moslem.zarini@ymail.com)

