پژوهش‌های آبخیزداری
شاپا: ۲۰۳۸-۲۹۸۱

پیش‌بینی و مدل‌سازی نرخ فرسایش خاک با استفاده از الگوریتم‌های یادگیری ماشین در شهرستان‌های مراوه‌تپه و کلاله استان گلستان

عبدالحسین بوعلی[۱]، نرگس کریمی‌نژاد*[۲]، مرتضی اکبری[۳]، مریم یعقوبی[۴]، محسن فراهی[۵]

۱- محقق پسادکتری گروه مدیریت مناطق خشک و بیابانی، دانشکده منابع‌طبیعی و محیط‌زیست، دانشگاه فردوسی مشهد، مشهد، ایران

۲- استادیار بخش مهندسی منابع‌طبیعی و محیط‌زیست، دانشکده کشاورزی، دانشگاه شیراز، شیراز، ایران

۳- دانشیار گروه مدیریت مناطق خشک و بیابانی، دانشکده منابع‌طبیعی و محیط‌زیست، دانشگاه فردوسی مشهد، مشهد، ایران

۴- دانشجوی دکتری گروه مهندسی آب، دانشکده کشاورزی، دانشگاه شهرکرد، شهرکرد، ایران

۵- استادیار گروه مهندسی طبیعت، دانشکده علوم محیطی و توسعه پایدار، دانشگاه سراوان، سراوان، ایران

چکیده مبسوط

مقدمه و هدف

فرسایش آبی یکی از چالش‌های مهم زیست‌محیطی در سراسر جهان است که منجر به نابودی خاک، کاهش حاصلخیزی و افزایش رسوب‌دهی در رودها می‌شود. استان گلستان، به‌ویژه مناطق شمالی آن با خاک‌های حساس لسی و شرایط اقلیمی خاص، یکی از کانون‌های اصلی فرسایش در ایران به‌شمار می‌آید. موضوع بررسی‌شده، ضعف روش‌های سنتی در ارائه پهنه‌بندی دقیق، سریع و قابل اطمینان برای فرسایش آبی در مناطق حساس استان گلستان است؛ جایی که روابط پیچیده و غیرخطی میان عامل‌های محیطی نیازمند رویکردهای پیشرفته‌تر است. الگوریتم‌های یادگیری ماشین به‌دلیل قابلیت مدل‌سازی روابط پیچیده و غیرخطی میان متغیرها، ابزار قدرتمندی برای پیش‌بینی مکانی پدیده‌هایی مانند فرسایش به‌شمار می‌آیند. هدف این پژوهش، مدل‌سازی مکانی فرسایش آبی در شهرستان‌های مراوه‌تپه و کلاله با استفاده از الگوریتم‌های پیشرفته یادگیری ماشین و مقایسه کارایی آن‌ها با یک مدل ترکیبی (Ensemble) بود.

مواد و روش‌ها

منطقه مطالعه‌شده زمین‌های شمالی استان گلستان با مساحت ۴۹۴۴٫۵ کیلومترمربع بود که به‌دلیل پستی‌بلندی شیب‌دار، بارندگی‌های فصلی و خاک‌های لسی، مستعد فرسایش آبی است.

نوع مقاله: پژوهشی
***مسئول مکاتبات:** narges.karimi991@gmail.com
استناد: بوعلی، ع.، کریمی‌نژاد، ن.، اکبری، م.، یعقوبی، م.، فراهی، م. ۱۴۰۵. پیش‌بینی و مدل‌سازی نرخ فرسایش خاک با استفاده از الگوریتم‌های یادگیری ماشین در شهرستان‌های مراوه‌تپه و کلاله استان گلستان. پژوهش‌های آبخیزداری. ۳۹(۲): ۹۵-۸۰.

شناسهٔ دیجیتال: 10.22092/WMRJ.2025.371189.1641

تاریخ دریافت: ۱۴۰۴/۸/۲۵، **تاریخ بازنگری**: ۱۴۰۴/۹/۱۲، **تاریخ پذیرش**: ۱۴۰۴/۹/۳۰، **تاریخ انتشار**: ۱۴۰۵/۴/۱

پژوهش‌های آبخیزداری، سال ۱۴۰۵، دورهٔ ۳۹، شمارهٔ ۲، شمارهٔ پیاپی ۱۵۱، تابستان ۱۴۰۵، صفحه‌های ۸۰ تا ۹۵.

ناشر: مرکز تحقیقات و آموزش کشاورزی و منابع‌طبیعی استان فارس

© نویسندگان

ازاین‌رو، ابتدا لایه‌های اطلاعات مربوط به عامل‌های مؤثر بر فرسایش تهیه شد. این عامل‌ها ۸ شاخص محیط زیستی و اقلیمی شامل شاخص بارش (داده‌های ایستگاهی)، شاخص پوشش گیاهی (NDVI)، شاخص بافت سطحی خاک (TGSI)، لایه کاربری زمین، شاخص رطوبت سطحی (NDMI)، مدل ارتفاعی رقومی (DEM)، شاخص خاک لخت (BI) و شاخص شوری خاک (NDSI) بودند که همگی از تصویرهای ماهواره‌ای لندست ۸ استخراج شدند. نقشه موقعیت فرسایش‌های مشاهده‌شده به‌عنوان لایه وابسته در مدل‌سازی استفاده شد. بر این اساس در بازدید میدانی ۱۰۰ نقطه (۵۰ نقطه از فرسایش آبی و ۵۰ نقطه از نبودن آن) برداشت شد. در این پژوهش، از ۷۰٪ داده‌ها برای آموزش و از ۳۰٪ آن برای آزمون استفاده شد. سه الگوریتم یادگیری ماشین شامل جنگل تصادفی (RF)، ماشین بردار پشتیبانی (SVM) و وایازی درختی تقویت‌شده (BRT) برای آموزش مدل به‌کار گرفته شدند. سرانجام، یک مدل ترکیبی (Ensemble) بر اساس میانگین وزنی خروجی از هر یک از این سه مدل ساخته شد تا دقت پیش‌بینی افزایش یابد. عملکرد مدل‌ها با استفاده از شاخص‌های آماری ضریب کاپا، منحنی مشخصه عملکرد گیرنده و شاخص آماری تسیل ارزیابی شد.

نتایج و بحث

نتایج ارزیابی نشان داد که عملکرد هر یک از سه مدل به‌طور جداگانه RF، SVMو BRT در مدل‌سازی فرسایش آبی قابل قبول بود، اما عملکرد مدل ترکیبی (Kappa=۰/۹۰، ROC=۰/۹۳ و TSS =۰/۸۹) با بهره‌گیری از نقاط قوت هر یک از سه مدل نامبرده، در تفکیک دقیق پهنه‌های مستعد به فرسایش از پهنه‌های پایدار و کاهش خطاهای طبقه‌بندی برتر بود. این برتری مؤید آن است که در مدل ترکیبی (Ensemble) با میانگین‌گیری وزنی، می‌توان نبودن قطعیت‌ها را مؤثرتر از هر یک از مدل‌ها به‌طور جداگانه مدیریت کرد و یک خروجی پایدار و قابل اطمینان به‌دست آورد. نقشه پتانسیل فرسایش نهایی تولیدشده به‌وسیلهٔ این مدل دقیق بود و به‌وضوح نشان داد که مناطق شمالی، غربی و جنوب‌غربی منطقه مطالعه‌شده کانون‌های اصلی فرسایش آبی هستند. این الگوی مکانی صرفاً تصادفی نبوده، بلکه ریشه در تعامل پیچیده عامل‌ها دارد؛ در این مناطق با پستی‌بلندی شیب‌دار، سرعت جریان آب و قدرت فرسایندگی آن افزایش یافته و هم‌زمان با دریافت بارندگی‌های بیشتر انرژی لازم برای فرآیند فرسایش فراهم می‌آید و پوشش گیاهی که به‌عنوان سد دفاعی اولیه خاک عمل می‌کند، ضعیف است. با تحلیل اهمیت متغیرها نیز مشخص شد بارش به‌عنوان عامل محرک و شاخص پوشش گیاهی (NDVI) به‌عنوان عامل مقاوم، به‌ترتیب مهم‌ترین مهارکنندگان فرسایش در این منطقه است. این یافته بیانگر آن است که با هرگونه اقدام حفاظتی که بتوان پوشش گیاهی را بهبود بخشید (مانند کاشت، مدیریت چرای دام و جلوگیری از نابودی)، می‌توان به‌طور مستقیم اثرات ویرانگر باران را خنثی کرد. این اقدام مؤثرترین راهکار برای مهار فرسایش در مناطق حساس است. ازاین‌رو، نتایج این پژوهش هم یک نقشه دقیق ارائه داد و هم یک چارچوب علمی برای اولویت‌بندی و اجرای اقدامات مدیریتی مبتنی بر شواهد فراهم آورد.

نتیجه‌گیری و پیشنهادها

برتری مدل ترکیبی در این پژوهش، بیانگر قابلیت اطمینان زیاد رویکردهای مبتنی بر میانگین‌گیری وزنی در کاهش خطا و افزایش دقت پیش‌بینی بود. استعداد مناطق پرخطر در بخش‌های شیب‌دار و با پوشش گیاهی کم، با اصول فیزیک فرسایش کاملاً سازگار است. اهمیت زیاد بارش و پوشش گیاهی مؤید آن است که مدیریت پایدار زمین، از جمله اجرای عملیات زیستی (مانند کاشت گونه‌های مقاوم و با ریشه‌های ژرف) مؤثرترین راهکار برای مهار فرسایش در منطقه بوده و بر پایهٔ گزارش‌های علمی ۳۰ تا ۵۰٪ از رواناب سطحی و فرسایش را کاهش می‌دهد. تصمیم‌گیرندگان، مدیران منابع‌طبیعی و برنامه‌ریزان شهری و روستایی از نقشه فرسایش تولیدشده در این پژوهش می‌توانند به‌عنوان یک ابزار علمی و کاربردی برای اولویت‌بندی اقدامات حفاظتی و تخصیص بهینه منابع، بهره‌گیرند. بر پایهٔ نتایج این پژوهش، پیشنهاد می‌شود در پژوهش‌های آینده با در نظر گرفتن محدودیت‌ها (مانند تفکیک مکانی برخی داده‌ها) از مدل‌های ترکیبی پیشرفته‌تر و داده‌هایی با دقت بیشتر استفاده شود.

واژگان کلیدی
روش‌های یادگیری ماشین، شمال استان گلستان، فرسایش آبی، مدل‌سازی

مقدمه

فرسایش خاک یکی از چالش‌های مهم محیط زیستی است که اثرات منفی قابل‌توجهی بر کشاورزی، منابع آب و بوم‌سازگان طبیعی دارد (ویجیتکوسوم۲۰۲۰). استان گلستان یکی از مناطق مهم تولید محصولات زراعی (مانند گندم، جو، برنج و سیب‌زمینی) است. این استان به‌دلیل داشتن خاک‌های حاصلخیز و شرایط اقلیمی مناسب، در تأمین امنیت غذایی کشور نقش مهمی دارد. بخش عمده‌ای از تولیدات کشاورزی استان گلستان از زمین‌های لسی تأمین می‌شود که به‌دلیل فرسایش‌پذیری زیاد، توجه ویژه به کیفیت خاک و مدیریت مؤثر این زمین‌ها ضروری است (نجفی‌نیا و همکاران ۲۰۱۸). بر پایهٔ نتایج بررسی‌ها اندازهٔ فرسایش خاک در این منطقه تحت تأثیر کاربری‌های مختلف، متفاوت است؛ به‌طوری که در شهرستان‌های کلاله و مراوه‌تپه، کاربری‌های زراعی (۱۳۷۹۵۶ هکتار) و مرتعی (۲۸۱۶۶۲ هکتار) بیشترین تأثیر را بر فرسایش دارند (جعفری اردکانی و همکاران ۲۰۰۹). این نهشته‌ها سبب تغذیه سرشاخه‌های گرگان‌رود و اترک شده و با تولید رسوب زیاد در پایین‌دست عواقب زیست‌محیطی و اقتصادی جدی به‌همراه دارد. به‌طور ویژه، نرخ تولید رسوب در این نهشته‌ها ممکن است به ۲۱ تا ۲۷ تن در هکتار برسد (عرب‌خدری ۲۰۲۱). مدل‌سازی و پیش‌بینی فرسایش در زمین‌های لسی، هم برای شناسایی و مدیریت این خطرها ضروری است و هم کشاورزان و مدیران منابع‌طبیعی می‌توانند با بهره‌گیری از آنها در تصمیم‌گیری‌ها و تعیین تدابیر مناسب، از کاهش کیفیت خاک (فرسایش و رسوب) و آسیب‌های تولیدات جلوگیری کنند (موسوی و همکاران ۲۰۲۱).
با توجه به پیچیدگی عامل‌های مؤثر بر فرسایش خاک، ممکن است استفاده از روش‌های سنتی برای ارزیابی و مدل‌سازی این پدیده ناکافی باشد (معماریان و اکبری ۲۰۲۱). مدل‌های سنتی مانند PSIAC[1] و IRIFR[2] به‌طور گسترده‌ای برای ارزیابی فرسایش آبی و بادی استفاده می‌شوند (جهانی و همکاران ۲۰۲۴). این مدل‌ها معمولاً بر اساس داده‌های میدانی و ویژگی‌های فیزیکی خاک، مانند بافت، شیب، پوشش گیاهی و بارش طراحی شده‌اند (بشری و همکاران ۲۰۲۳؛ معماریان و همکاران ۲۰۲۳). با این مدل‌ها معمولاً نمی‌توان تعاملات پیچیده میان عامل‌های مختلف را به‌خوبی شبیه‌سازی کرد. در نتیجه ممکن است دقت پیش‌بینی آن‌ها کم باشد (بوعلی و همکاران ۲۰۲۴). به این دلیل، در سال‌های پیشین به استفاده از روش‌های هوش مصنوعی مانند یادگیری ماشین توجه بیشتری شده است (اکبری و همکاران ۲۰۱۶). با این روش‌ها می‌توان حجم زیادی از داده‌های ویژگی‌های خاک، شرایط اقلیمی و پستی‌بلندی را پردازش کرد و با شناسایی الگوهای فرسایش پیش‌بینی دقیقی از اندازه آن، ارائه داد (بوعلی و همکاران ۲۰۲۵؛ القاضی و همکاران ۲۰۲۳).
نتایج پژوهشی در زمین‌های شرق دریاچه ارومیه نشان داد بهترین عملکرد در پیش‌بینی فرسایش‌پذیری بادی خاک مربوط به مدل شبکه عصبی هیبریدشده با الگوریتم بهینه‌سازی وال (WOA) بود (راعی و همکاران۲۰۲۰). جلالی‌فرد و همکاران (۲۰۱۸) فرسایش تونلی را در زمین‌های لسی استان گلستان بررسی کردند و با استفاده از وایازی لجستیک خطر تشکیل تونل را پیش‌بینی کردند. نتایج این پژوهش نشان داد متغیرهای مختلف مانند مقاومت خاک و فاصله از آبراهه تأثیر زیادی بر خطر تشکیل تونل دارند. نتایج پژوهش‌های انجام‌شده در آبخیز باختگان نشان داد عملکرد مدل جنگل تصادفی برای پیش‌بینی

1-Iran Research Institute of Forest and Rangelands
2- Pacific Southwest Inter- Agency Committee

فرسایش بادی و مدل ماشین بردار پشتیبانی برای پیش‌بینی فرسایش آبی، خوب بود. این نتایج نشان‌دهنده عملکرد زیاد مدل‌های هوش مصنوعی در شبیه‌سازی فرآیندهای پیچیده فرسایش خاک است (رن و همکاران ۲۰۲۴). پیری و مبارکی (۲۰۲۱) در زمین‌های قرقری شهرستان هیرمند به‌منظور پیش‌بینی شوری خاک از روش‌های زمین‌آمار و مدل‌های هوش مصنوعی شامل شبکه عصبی مصنوعی، درخت تصمیم و ماشین بردار پشتیبان استفاده کردند. نتایج این پژوهش نشان داد در برآورد شوری خاک دقت روش‌های هوش مصنوعی به‌طور قابل‌توجهی بیشتر از روش‌های زمین‌آمار بود.

اگرچه در پژوهش‌های پیشین از دقت زیاد الگوریتم‌های یادگیری ماشین در پهنه‌بندی فرسایش گزارش شده است، اما اغلب این پژوهش‌ها بر اساس کاربرد جداگانهٔ این مدل‌ها بوده و از قابلیت مدل‌های ترکیبی برای کاهش نبودن قطعیت در آنها بهره‌گیری نشده است. افزون بر این، در بیشتر این پژوهش‌ها به تلفیق جامع شاخص‌های سنجش از دور برای فرسایش آبی کمتر توجه شده است. از این رو، سعی شد در این پژوهش این خلأ جبران شود. مروری بر نتایج پژوهش‌های پیشین بیانگر آن است که پژوهشگران برای غلبه بر محدودیت‌های مدل‌های سنتی، به استفاده از مدل‌های هوش مصنوعی و داده‌کاوی روی آورده‌اند. در این راستا، کارایی زیاد مدل‌هایی مانند شبکه‌های عصبی، جنگل تصادفی و ماشین بردار پشتیبان در پیش‌بینی انواع فرسایش (آبی و بادی) در مناطق مختلف ایران از جمله استان گلستان به اثبات رسیده است. با این حال، هدف این پژوهش ارزیابی و مدل‌سازی فرسایش خاک در نهشته‌های لسی شمال شرق استان گلستان (شهرستان‌های مراوه‌تپه و کلاله)، با استفاده از روش‌های داده‌کاوی بود. نوآوری این پژوهش در به‌کارگیری یک مدل ترکیبی وزنی مبتنی بر سه الگوریتم یادگیری ماشین و ادغام شاخص‌های ماهواره‌ای برای ارائه دقیق‌ترین پهنه‌بندی مکانی فرسایش آبی در شمال استان گلستان بود. با بهره‌گیری از نتایج این پژوهش می‌توان عامل‌های مؤثر بر فرسایش خاک را دقیق‌تر شناسایی کرد و راهکارهای مدیریتی مناسبی برای کاهش اثرات منفی آن ارائه داد. با توجه به اهمیت موضوع (تهدید و فرصت نهشته‌های لسی) و نیاز به استفاده از فناوری‌های نوین در مدیریت منابع‌طبیعی، انجام این پژوهش می‌تواند گامی مؤثر برای حفاظت از خاک و توسعه پایدار در منطقه به‌شمار آید.

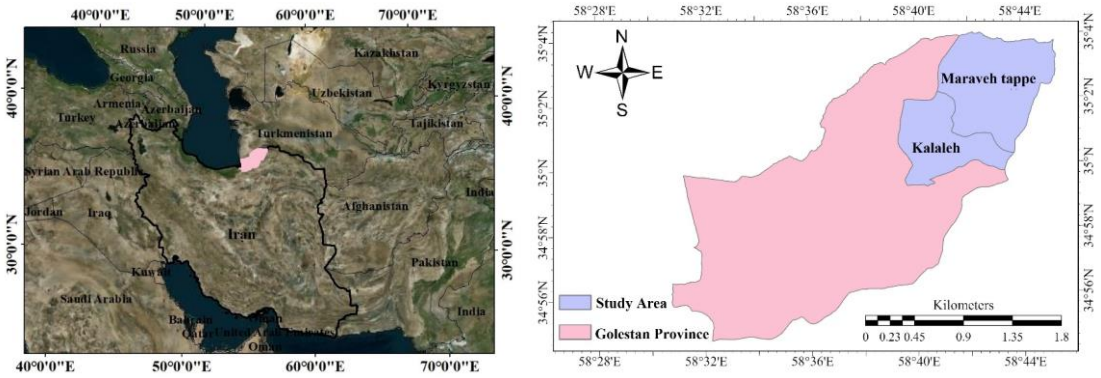
مواد و روش‌ها

معرفی منطقه مطالعه‌شده

منطقه پژوهشی شمال‌شرق استان گلستان شامل شهرستان‌های کلاله و مراوه تپه با مساحت ۴۹۴۴٫۵ کیلومتر مربع بود (شکل ۱). بیش از دو سوم استان گلستان آب و هوای خشک و نیمه‌خشک دارد که با نزدیک شدن به سوی شمال و شهرستان‌های مراوه‌تپه و کلاله، خشکی آن افزایش می‌یابد. میانگین سالانهٔ بارش و دمای منطقه در مدت دوره ۳۰ ساله (۱۴۰۰-۱۳۷۰) به‌ترتیب ۲۷۵/۴ میلی‌متر و ۱۶/۵ درجه سانتی‌گراد برآورد شد.

لایه‌های ورودی

ازجمله عامل‌های مؤثر بر فرسایش آبی می‌توان به عامل‌های اقلیمی، خاک، پستی بلندی و پوشش‌گیاهی اشاره کرد (یانگ و همکاران ۲۰۲۳). اولین گام برای مدل‌سازی مکانی فرسایش با مدل‌های یادگیری ماشین، تعیین و انتخاب عامل‌های مؤثر بر فرسایش است (محمدی‌فرد و غلامی ۲۰۲۱). ازاین‌رو، در این طرح ابتدا با بررسی منابع و ارزیابی مدل‌های موجود، عامل‌های تأثیرگذار بر فرسایش آبی تعیین شد. دومین ورودی به مدل‌های یادگیری ماشین در مدل‌سازی مکانی، نقاط فرسایشی است. در این پژوهش نقاط فرسایش آبی، در بازدیدهای میدانی تعیین شد. بر این اساس ۱۰۰ نقطه فرسایش آبی در منطقه شناسایی شد.



(الف) (ب)

(a) (b)

شکل ۱ -موقعیت منطقه مطالعه شده: الف- ایران، ب- شهرستان‌های کلاله و مراوه‌تپه در استان گلستان.

Figure 1- Location of the study area: a) Iran, b) Kalaleh and Maraveh Tappeh counties in Golestan Province.

تعیین شاخص‌های سنجش از دوری و پستی‌بلندی

برای تهیه شاخص‌های سنجش از دوری نیز از تصویر ماهواره‌ی لندست استفاده شد. به‌منظور سهولت در استفاده از باندها، این تصویر از سامانه موتور گوگل‌ارث تهیه شد. در این پژوهش از تصویر لندست ۸ استفاده شد. تصویرها در تاریخ تیر ماه ۱۴۰۳ انتخاب شدند تا اثر تغییرات فصلی به حداقل برسد. این داده‌ها به‌طور پیش‌فرض با تصحیحات رادیومتریک و جوی بودند و متغیرها و شاخص‌ها بر اساس همین داده‌های تصحیح‌شده استخراج شدند.

در این پژوهش، به‌منظور مدل‌سازی فرسایش آبی از مجموعه‌ای از شاخص‌های سنجش از دوری استفاده شد که به‌طور ویژه برای ارزیابی ویژگی‌های مختلف فرسایش خاک طراحی شده‌اند. این شاخص‌ها شامل NDVI (شاخص تفاضلی بهنجارشده پوشش گیاهی)، NDMI (شاخص رطوبت بهنجارشده)، NDSI (شاخص شوری بهنجارشده)، TGSI (اندازه ذرات بالای سطح خاک) و BSI (شاخص خاک لخت)، CHRIPS (بارش)، کاربری زمین و مدل رقومی ارتفاعی (تهیه‌شده از سازمان نقشه‌برداری ایران با دقت ۳۰ متر)، هستند. با استفاده از این شاخص‌ها، می‌توان به تحلیل دقیق‌تری از وضعیت رطوبت، مواد آلی و دیگر ویژگی‌های فیزیکی و شیمیایی خاک دست یافت. افزون بر این، نقاط آموزشی برای اجرای مدل‌ها بر اساس بودن (۵۰ نمونه) و نبودن (۵۰ نمونه) فرسایش آبی که در مدت بازدید میدانی برداشت‌شده بودند، اجرا شد. ازاین‌رو، از مناطقی هم که فرسایش آبی وجود نداشت، نمونه‌گیری شد. به‌منظور یکپارچه‌سازی لایه‌های ورودی برای مدل‌سازی، تمام لایه‌های محیطی با استفاده از روش‌های استاندارد نزدیکترین همسایه و درونیابی دوخطی هم‌مقیاس-سازی شدند. هدف این پژوهش، ایجاد یک مدل جامع و کارآمد برای پیش‌بینی فرسایش آبی بود که بتوان به مدیریت بهتر منابع‌طبیعی و کشاورزی کمک کرد.

مدل‌های یادگیری ماشین

در این پژوهش به‌منظور پیش‌بینی فرسایش خاک، ۳ روش یادگیری ماشین با ویژگی‌ها و مزایای ویژه خود به‌کار گرفته شد. تمام این مدل‌ها در محیط برنامه‌نویسی R اجرا شدند. سنجه‌های مدل‌ها بر اساس تنظیمات بهینه‌سازی‌شده در فرآیند مدل‌سازی تعیین شدند. در مدل SVM از هسته RBF به‌دلیل عملکرد خوب در مدل‌سازی روابط غیرخطی برای داده‌های

محیطی، استفاده شد. در مدل RF تعداد ۵۰۰ درخت بر اساس آزمون اندازه‌های مختلف با بهترین عملکرد به‌کار گرفته شد. در مدل BRT نرخ یادگیری برابر با ۰/۰۱ در نظر گرفته شد تا از بیش‌برازش جلوگیری شده و فرایند یادگیری به‌طور پایدار انجام شود. مدل جنگل تصادفی[3] یکی از روش‌های خوب در یادگیری ماشین است که برای طبقه‌بندی و وایازی استفاده می‌شود. این مدل با ایجاد مجموعه‌ای از درختان تصمیم و ترکیب نتایج آن‌ها دقت پیش‌بینی را افزایش می‌دهد. جنگل تصادفی به‌دلیل کاهش خطر بیش‌برازش و قابلیت پردازش داده‌های بزرگ و پیچیده، به‌عنوان یک روش بسیار مؤثر شناخته‌شده است (بلگیو و دَراگو ۲۰۱۶). مدل ماشین بردار پشتیبان[4] یک الگوریتم قدرتمند برای شناسایی مرزهای میان طبقه‌ها است. این مدل با استفاده از هسته‌های مختلف، به‌ویژه هسته خطی، داده‌ها را در فضاهای چندبعدی تفکیک می‌کند. SVM به‌دلیل به‌کارگیری داده‌های غیرخطی و پیچیده، برای تحلیل فرسایش خاک بسیار مناسب است (تانگ و همکاران۲۰۲۰). مدل وایازی درختی پیشرفته[5] ترکیبی از چندین درخت تصمیم است که برای بهبود دقت پیش‌بینی طراحی‌شده است. این مدل با تمرکز بر الگوهای پیچیده داده‌ها، روابط غیرخطی میان متغیرها را شناسایی می‌کند. عملکرد BRT با استفاده از روش‌های تقویت به‌طور قابل‌توجهی افزایش یافته و به‌عنوان یک ابزار مؤثر در تحلیل فرسایش خاک شناخته‌شده است (پارک و همکاران ۲۰۱۶).

ارزیابی عملکرد مدل‌ها

در این پژوهش از سه شاخص ضریب کاپا، منحنی تشخیص عملکرد[6] و آمار واقعی مهارت مرتبط با آستانه[7] به‌عنوان روش‌های ارزیابی دقت مدل‌ها استفاده شد. در این پژوهش داده‌ها با روش۷۰٪ آموزش و ۳۰٪ آزمون تقسیم شدند. ضریب کاپا، که اندازه توافق میان پیش‌بینی‌های مدل و داده‌های واقعی را می‌سنجد، به‌عنوان معیاری برای ارزیابی عملکرد کلی مدل در دسته‌بندی صحیح نمونه‌ها به‌کار گرفته شد. دامنه تغییرات این شاخص از ۱+ تا ۱- متغیر است (بوفانا و همکاران۲۰۲۰). منحنی مشخصه عملکرد نیز، با نمایش تصویری از تعادل میان نرخ مثبت واقعی و نرخ مثبت کاذب در آستانه‌های مختلف، امکان ارزیابی جامع‌تری از عملکرد مدل در تفکیک طبقه‌ها را فراهم آورد. در این شاخص طبقه‌های عالی (۱- ۰/۹)، بسیارخوب (۰/۸ – ۰/۹)، خوب (۰/۷ – ۰/۸)، متوسط (۰/۶ – ۰/۷) و ضعیف (۰/۵ – ۰/۶) برای همبستگی کیفی و کمی سطح زیر منحنی ارائه‌شده است (دوکوتا و همکاران۲۰۱۳). سرانجام، آمار مهارت واقعی مبتنی بر آستانه ، به‌عنوان معیاری برای ارزیابی حساسیت مدل به تغییرات در آستانه تصمیم‌گیری، به‌کار گرفته شد تا از ارائه نتایج قابل‌قبول در شرایط مختلف به‌وسیلهٔ مدل اطمینان حاصل شود. طبقه‌های در شاخص TSS به‌ترتیب ضعیف (کمتر از ۰/۲) متوسط (۰/۲ تا ۰/۶) و خوب (بزرگ‌تر از ۰/۶) در نظر گرفته شده است (الوش و همکاران ۲۰۰۶). استفاده از این سه شاخص، امکان ارزیابی چندوجهی و دقیق‌تری از عملکرد مدل‌ها را فراهم آورد.

نتایج و بحث

شاخص‌های سنجش از دوری

در مدل‌سازی فرسایش آبی منطقهٔ شمال استان گلستان (شهرستان‌های مراوه‌تپه و کلاله)، هر یک از شاخص‌های تعیین‌شده، عامل‌های تعیین‌کننده شدت فرسایش به‌شمار می‌آیند. شاخص BI بیانگر مناطقی با پوشش گیاهی کم یا بدون پوشش گیاهی بود که اندازه‌های بیشتر آن در جنوب و جنوب‌شرق منطقه، نشان‌دهندهٔ آسیب‌پذیری این نواحی در برابر فرسایش آبی بود. شاخص Chirpsبیانگر اندازه و پراکندگی بارش بود. اندازه‌های بیشتر این شاخص در نواحی مرکزی و جنوبی بیانگر توان بیشتر برای ایجاد رواناب و فرسایش بود. مدل رقومی ارتفاع، نمایانگر شیب و

3- Random Forest (RF)
4- Support Vector Machine (SVM)
5- Boosted Regression Trees (BRT)
6- Receiver Operating Characteristic (ROC)
7- True Skill Statistic (TSS)

بلندی زمین بود که شیب زیاد در نواحی جنوبی و شرقی، عامل مهمی در جابجایی ذرات خاک و افزایش فرسایش بود. شاخص کاربری زمین نشان‌دهندهٔ پراکندگی مناطق کشاورزی و منابع‌طبیعی بود و مناطق با پوشش گیاهی ضعیف‌تر به‌شدت در معرض تهدید فرسایش بودند. شاخص NDMI بیانگر وضعیت رطوبت خاک بود و اندازه‌های کمتر آن در جنوب، نشان‌دهنده خشکی و افزایش حساسیت خاک به فرسایش بود.

شاخص NDSI بیانگر اندازه شوری خاک بود. خاک‌های شور بیشتر در مناطق جنوبی و شرقی مشاهده شد و ساختار فیزیکی ضعیف مسبب کاهش پایداری آن‌ها در برابر جریان آب و افزایش استعداد آن‌ها به فرسایش بود. شاخص NDVI بیانگر تراکم و سلامت پوشش گیاهی بود. شاخص TGSI نیز به‌طور ویژه، نشان‌دهندهٔ بافت خاک سطحی و خصوصیات مرتبط با آن بود. اندازه‌های کم شاخص‌های NDVI و TGSI در جنوب منطقه، نمایانگر اهمیت پوشش گیاهی به‌عنوان عاملی برای کاهش فرسایش بود. در مجموع، با بررسی شاخص‌های مختلف ارائه‌شده، می‌توان نتیجه گرفت که با حرکت در منطقه به‌سوی نواحی بلند و با شیب بیشتر، خاک لخت‌تر، شوری بیشتر، و پوشش گیاهی تنک‌تر شده و این موضوع نشان‌دهنده افزایش خطر و شدت فرسایش آبی در منطقه است.

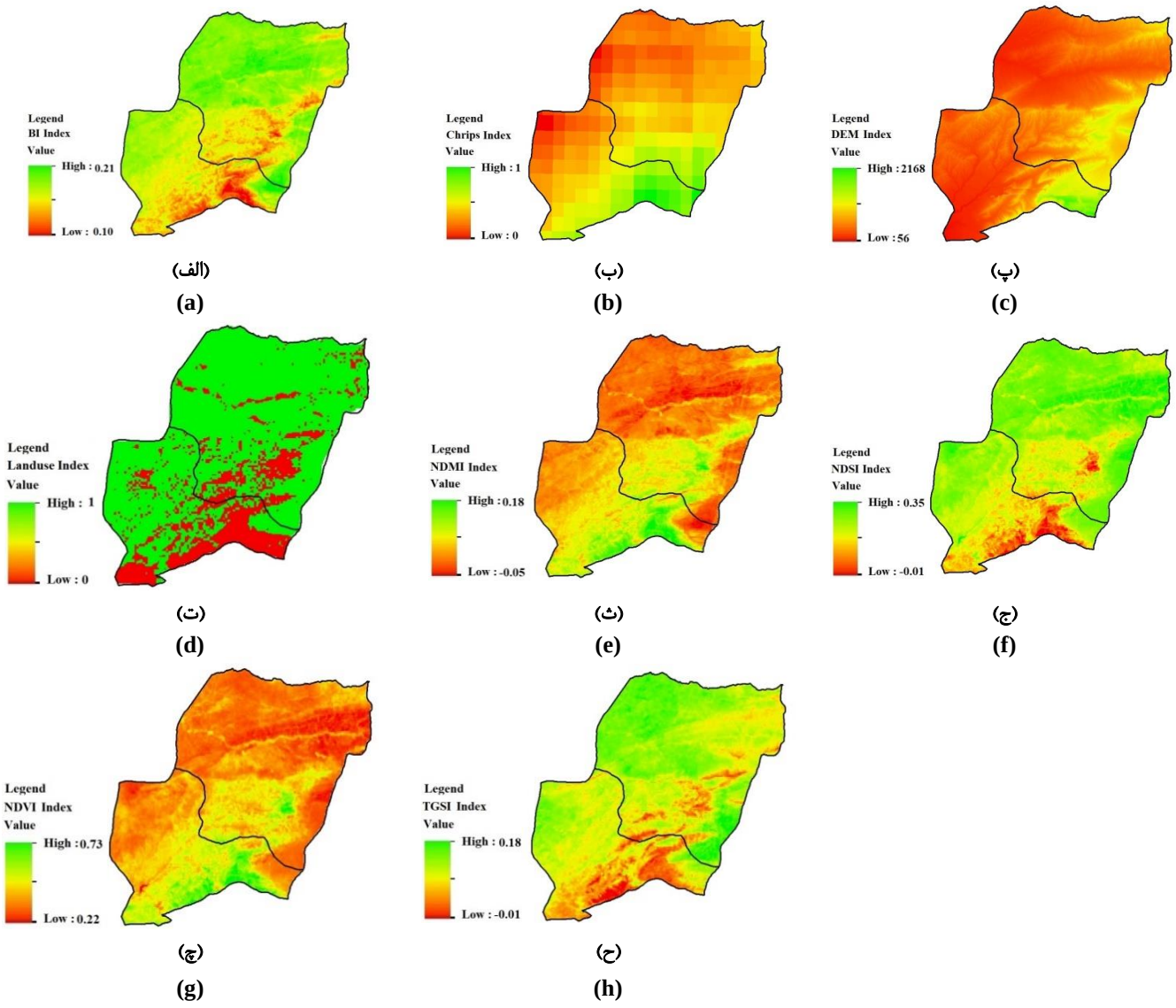
مجموعه این شاخص‌ها، امکان تحلیل جامع و مدل‌سازی دقیق فرسایش آبی را در شهرستان‌های مراوه‌تپه و کلاله فراهم آورد. این نتایج با یافته‌های اکبری و همکاران (۲۰۲۰) و پژوهش‌های مشابه در مناطق خشک ایران همخوانی دارد.

مدل‌سازی فرسایش آبی

پس از تهیه نقشه‌های شاخص‌های سنجش از دوری و تعیین نقاط بودن و نبودن فرسایش آبی، مدل‌سازی انجام شد. بر اساس نتایج به‌دست‌آمده، سه الگوریتم پیشرفته یادگیری ماشین برای مدل‌سازی مکانی فرسایش آبی در شمال استان گلستان به‌کار گرفته شد (شکل۳). با مقایسه نقشه‌های به‌دست آمده و با استفاده از مدل BRT عمدتاً نواحی شمالی و مرکزی منطقه به‌عنوان پهنه‌های با توان فرسایشی زیاد شناسایی شد.

بر اساس اندازهٔ شاخص OOB Error برابر با ۰/۱۲، مدل RF (بیشترین اندازهٔ ۰/۹۸)، مناطق گسترده‌تر و پراکنده‌تری در شمال و مرکز را در طبقه‌های پرخطر شناسایی کرد و مرزهای خطر را گسترده‌تر رسم کرد. مدل SVM نیز شبیه دیگر مدل‌ها بود. برخی تفاوت‌های محلی بیانگر آن بود که با اندازهٔ بیشتر شاخص عملکرد (ROC = ۰/۸۴، KAPPA = ۰/۸۴ و TSS=۰/۸۷) مناطق خطر با دقت مکانی بیشتری شناسایی شد. دلیل عملکرد زیاد و الگوریتم پیشرفته مدل SVM آن است که بر اساس شناسایی شرایط متفاوت و تعیین توابع گوناگون ناسنجه‌ای، امکان بیشترین اندازهٔ تفکیک در میان داده‌ها را فراهم می‌آورد (واپنیک ۱۹۹۵). بر پایهٔ نتایج جدول ۱، عملکرد هر یک از سه مدل به‌طور جداگانه در ارزیابی فرسایش خاک نسبتاً مطلوب بود، اما عملکرد مدل ترکیبی (Kappa=۰/۹۰، ROC=۰/۹۳ و TSS ۰/۸۹ =)، به‌مراتب برتر بود. این مدل بر پایه میانگین وزنی خروجی سه الگوریتم و متناسب با معیارهای کارایی هر مدل ساخته شد، به‌گونه‌ای که سهم مدل‌های دقیق‌تر در ترکیب بیشتر بود.

نقشه خروجی مدل ترکیبی نشان داد که این رویکرد ضمن ترکیب نقاط قوت هر یک از مدل‌های استفاده‌شده به‌طور جداگانه، مرزهای مکانی فرسایش را با وضوح و جزئیات بیشتری در مقایسه با کاربرد هر یک از آنها نشان داد و با بهره‌گیری از این رویکرد توزیع مکانی نواحی پر خطر به‌طور واقع‌بینانه‌تر رسم شد. مقایسه سه مدل یادگیری ماشین نشان داد که دقت مکانی مدل SVM بیشتر بود، اما عملکرد مدل ترکیبی برتر بود و با بهره‌گیری از آن مرزهای فرسایش واقع‌بینانه‌تر رسم شد. این یافته با نتایج پژوهش‌های اخیر مبنی بر استفاده از مدل‌های ترکیبی برای پیش‌بینی فرسایش در مناطق خشک و نیمه‌خشک هم‌راستا است (بوعلی و همکاران ۲۰۲۴).



شکل ۲- شاخص‌های سنجش از دور انتخاب‌شده به‌منظور مدل‌سازی فرسایش آبی. الف- شاخص خاک لخت، ب- شاخص بارش، پ- شاخص بلندی، ت- شاخص کاربری زمین، ث- شاخص رطوبت خاک، ج- شاخص شوری خاک، چ- شاخص پوشش گیاهی و ح-شاخص بافت سطحی خاک.

Figure 2- Selected remote sensing indices for water erosion modeling. a) Bare Soil Index, b) Precipitation Index, c) Elevation Index, d) Land Use Index, e) Soil Moisture Index, f) Soil Salinity Index, g) Vegetation Index and Surface, h) Soil Texture Index.

چهار طبقهٔ فرسایش آبی بر اساس اندازه‌های خروجی مدل ترکیبی و با روش Natural Breaks (Jenks) آستانه‌بندی شدند تا تغییرات طبیعی داده‌ها و توزیع مکانی شدت فرسایش به‌خوبی نمایان شود. طبقه‌بندی انجام‌شده با داده‌های میدانی بودن و نبودن فرسایش اعتبارسنجی شد و هم‌خوانی مناسبی میان طبقه‌های پیش‌بینی‌شده و مشاهده‌ای واقعی مشاهده شد. بر اساس نقشه ترکیبی، خطر فرسایش آبی در مناطق شمالی، غربی و جنوب‌غربی منطقه بیشتر بود. بر اساس نقشه ترکیبی ۲۳٪ از منطقه (۱۳۵۷۲,۳۱ هکتار) در طبقه فرسایش آبی کم، ۱۵٪ (۷۳۷۰۷ هکتار) در طبقه متوسط، ۲۲ ٪ (۱۱۰۴۳۲ هکتار) در طبقه شدید و ۴۰٪ (۱۹۶۷۳۸ هکتار) در طبقه بسیارشدید فرسایش آبی طبقه‌بندی شدند (شکل۳).

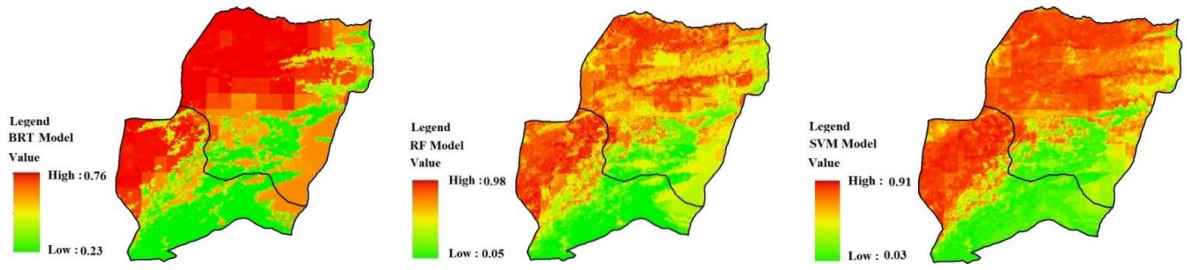
مهم‌ترین متغیرها

شناسایی متغیرهای تأثیرگذار بر فرسایش آبی اهمیت زیادی در پژوهش‌های محیط‌زیست، حفاظت از منابع‌طبیعی و توسعه پایدار دارد. با شناخت این متغیرها، می‌توان برنامه‌های مناسبی برای مهار و کاهش فرسایش آبی ارائه داد و سبب بهبود سامانه‌های مدیریت خاک شد. بر این اساس، اندازهٔ اهمیت شاخص‌های بررسی‌شده در مدل ترکیبی تعیین شد. نتایج نشان داد که به‌ترتیب اهمیت بارش، پوشش گیاهی (NDVI)، بافت خاک سطحی (TGSI)، کاربری زمین، رطوبت سطحی (NDMI)، تغییرات بلندی (DEM)، خاک لخت (BI) و شاخص شوری خاک (NDSI)، بیشتر بود (شکل۴). تحلیل نتایج بیانگر همخوانی اهمیت نسبی هر یک از متغیرها با الگوی مکانی شاخص مربوطه در منطقه بود. به بیان دیگر، بیشترین اثرگذاری در نقشه منطقه مربوط به متغیرهایی بود که بیشترین وزن را داشتند و پراکنش مکانی آنها با نمودار اهمیت متغیرها تطابق قابل‌توجهی داشت. این موضوع نشان‌دهنده رابطه مستقیم میان ارزش آماری متغیرها و اثر آن‌ها در منطقه بود. با بهره‌گیری از این اطلاعات در تصمیم‌گیری‌های مدیریتی و برنامه‌ریزی می‌توان از خطرات ناشی از فرسایش آبی جلوگیری کرد.

جدول۱- ارزیابی عملکرد مدل‌ها بر اساس شاخص‌های مختلف.

Table 1- Performance evaluation of models based on different indicators.

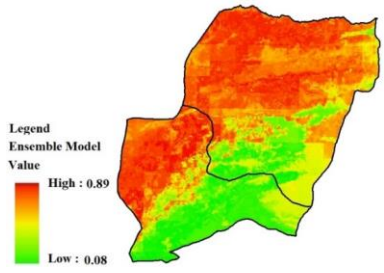
Models	KAPPA	ROC	TSS	Weighted average
BRT	0.77	0.81	0.59	2.17
SVM	0.84	0.84	0.87	2.55
RF	0.81	0.80	0.61	2.22
Ensemble	0.90	0.93	0.89	-



(الف) (a)

(ب) (b)

(پ) (c)

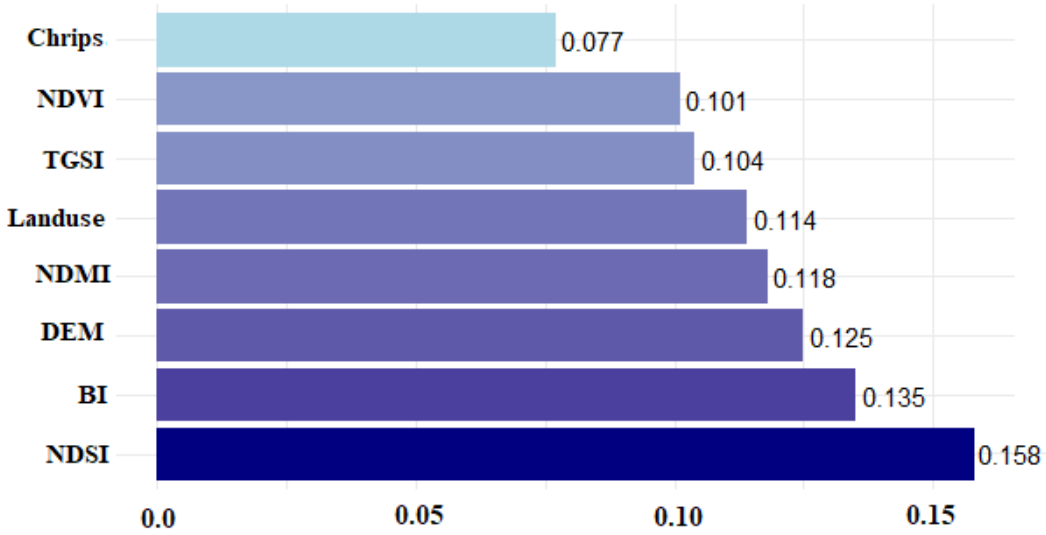


(ت)

(d)

شکل۳- مدل‌سازی فرسایش آبی با استفاده از مدل‌های الف-BRT، ب- RF، ج- SVM و ت- ترکیبی.

Figure 3- Water erosion modeling using a) BRT, b) RF, c) SVM, and d) the hybrid model.



شکل ۴- اهمیت متغیرهای اثرگذار بر فرسایش آبی در مدل ترکیبی.

Figure 4- Importance of variables affecting water erosion in a hybrid model.

نتیجه‌گیری و پیشنهادها

بر پایهٔ نتایج این پژوهش، مدل‌سازی فرسایش آبی در شمال استان گلستان تحت تأثیر مجموعه‌ای از شاخص‌های محیطی و بوم‌شناختی است. از مهم‌ترین آن‌ها می‌توان به بارش، پوشش گیاهی (NDVI)، بافت خاک سطحی (TGSI)، کاربری زمین، رطوبت سطحی (NDMI)، تغییرات بلندی (DEM)، خاک لخت (BI) و شوری خاک (NDSI) اشاره کرد.

جدول ارزیابی عملکرد مدل‌ها نیز بیانگر برتری مدل ترکیبی (Ensemble) در مقایسه با کاربرد هر یک از سه مدل BRT، RF و SVM به‌طور جداگانه بود. در مدل ترکیبی با تلفیق خروجی این الگوریتم‌ها و وزن‌دهی مبتنی بر عملکرد هرکدام، دقت و وضوح مکانی پیش‌بینی فرسایش آبی به‌طور قابل‌توجهی افزایش یافت. بر اساس نقشه ترکیبی، بیشترین احتمال فرسایش آبی مربوط به محدوده‌های شمالی، غربی و جنوب‌غربی منطقه بود که دلیل آن می‌تواند

شرایط پستی‌بلندی، پوشش گیاهی ضعیف و وجود خاک‌های حساس به فرسایش در این مناطق باشد. اهمیت زیاد متغیرهای بارش و پوشش گیاهی در تعیین شدت فرسایش، مؤید نقش مهم این عامل‌ها در تحریک رواناب سطحی و تثبیت خاک است.

پیشنهاد می‌شود در پژوهش‌های آتی اثرات تغییرات اقلیمی بر فرسایش آبی، به‌کارگیری شاخص‌های دقیق‌تر رطوبت خاک و پوشش گیاهی بررسی شوند و بازدیدهای میدانی به منظور افزایش دقت مدل‌های یادگیری ماشین گسترش یابند. با این اقدامات می‌توان دقت پیش‌بینی، وضوح مکانی و قابلیت تعمیم مدل‌ها را افزایش داد و مبنای تصمیم‌گیری علمی و مدیریت پایدار خاک در مناطق حساس به فرسایش آبی را برنامه‌ریزی کرد. در مجموع نتایج این پژوهش نشان داد کاربرد مدل‌های ترکیبی یادگیری ماشین در شناخت دقیق‌تر و مدیریت بهینه فرسایش آبی، رویکردی مؤثر است و برای برنامه‌ریزی‌های پایدار حفاظت خاک در مناطق حساس مانند شمال استان گلستان می‌تواند مبنای تصمیم‌گیری‌های علمی و مدیریتی برای مقابله با تهدیدات فرسایش آبی باشد.

تضاد منافع نویسندگان

نویسندگان این مقاله اعلام می‌دارند که هیچ‌گونه تضاد منافعی در راستای نگارش و انتشار مطالب و نتایج این پژوهش ندارند.

دسترسی به داده‌ها

داده‌ها و نتایج استفاده‌شده در این پژوهش، با مکاتبه با نویسندهٔ مسئول در اختیار مخاطب قرار خواهد گرفت.

مشارکت نویسندگان

نویسنده اول: مفهوم‌سازی، انجام تحلیل‌های نرم‌افزاری، نگارش نسخهٔ اولیهٔ مقاله

نویسنده دوم: راهنمایی، ویرایش و بازبینی مقاله، بررسی نتایج

نویسنده سوم: مفهوم‌سازی، مشاوره، بازبینی متن مقاله، تحلیل‌های نرم افزاری

نویسنده چهارم: بازدید میدانی جمع‌آوری و آماده‌سازی داده‌ها

نویسنده پنجم: جمع‌آوری و آماده‌سازی داده‌ها، ویرایش و بازبینی مقاله

فهرست منابع

Arab Khadri H. 2021. The status of water erosion and sediment yield in Iran: A statistical and comparative analysis. Strategic Research in Agricultural Sciences and Natural Resources, 6(2): 139–156. (In Persian).

Allouche O, Tsoar A, Kadmon R. 2006. Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). Journal of Applied Ecology. 43(6):1223–1232. https://doi.org/10.1111/j.1365-2664.2006.01214.x

Bashari H, Boali A, Soltani S. 2023. Accommodating uncertainty in soil erosion risk assessment: Integration of Bayesian belief networks and MPSIAC model. Natural Hazards Research. 1(4): 134 – 147. https://doi.org/10.1016/j.nhres.2023.09.009

Belgiu M, Drăgu L. 2016. Random forest in remote sensing: A review of applications and

Akbari M, Ownegh M, Asgari HR, Sadoddin A, Khosravi H. 2016. Desertification risk assessment and management program. Global Journal of Environmental Science and Management. 2(4):365–380. https://doi.org/10.22034/gjesm.2016.02.04.006

Akbari M, Modarres R, Alizadeh Noughani M. 2020. Assessing early warning for desertification hazard based on E-SMART indicators in arid regions of northeastern Iran, Journal of Arid Environments. Elsevier, 174:104086. https://doi.org/10.1016/j.jaridenv.2019.104086.

Alqadhi S, Mallick J, Talukdar S, Alkahtani M. 2023. An artificial intelligence-based assessment of soil erosion probability indices and contributing factors in the Abha-Khamis watershed, Saudi Arabia. Frontiers in Ecology and Evolution. 11: p. 1189184. https://doi: 10.3389/fevo.2023.1189184

future directions. ISPRS Journal of Photogrammetry and Remote Sensing. 114: 24-31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Boali A, Asgari H. R, Mohammadian Behbahani A, Salmanmahiny A, Naimi B. 2024. Remote sensing applications: Society and environment remotely sensed desertification modeling using ensemble of machine learning algorithms. Remote Sensing Applications: Society and Environment. 34(14): 1–16. https://doi.org/ 10.1016/j.rsase.2024.101149

Boali A, Kariminejad N, Hosseinalizadeh M. 2024. Enhancing wind erosion risk assessment through remote sensing techniques. Plos one. 19(10): e0308854. https://doi.org/10.1371/journal.pone.0308854

Boali A, Hosseinalizadeh M, Kariminejad N, Asgari HR, Behbahani AM, Naimi B, Shafaie V, Rad MM. 2025. Evaluation of early warning signals for soil erosion using remote sensing indices in northeastern Iran. Scientific Reports. 15 (9742):1–13. https://doi.org/10.1038/s41598-025-94926-x

Bofana J, Zhang M, Nabil M, Wu B, Tian F, Liu W, Moyo C. 2020. Comparison of different cropland classification methods under diversified agroecological conditions in the Zambezi River Basin. Remote Sensing. 12(13): p. 2096. https://doi.org/10.3390/rs12132096

Devkota KC, Regmi AD, Pourghasemi HR, Yoshida K, Pradhan B, Ryu IC, Althuwaynee OF. 2013. Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in GIS and their comparison at Mugling-Narayanghat road section in Nepal Himalaya. Natural Hazards. 65(1): 135–165. https://doi.org/10.1007/s11069-012-0347-6

Jalali Fard A, Hosseinalizadeh M, Kamaki Choghi B, Azim Mohseni M. 2018. Modeling of piping erosion in loess lands. Environmental Erosion Research. 8(4): 1–18. (In Persian).

Jafar Ardakani A, Bayat R, Piran HR, Shariat Jafari M. 2009. Investigation of erosion and sedimentation status in the loess deposits of Golestan province. Paper presented at the 6th Iranian Conference on Engineering Geology and Environment, Tehran, Iran. (In Persian).

Jahani M, Taleghani S, Akbari, M. 2024. Quantitative evaluation of soil erodibility potential using SLEMSA model (Case study: Karkheh watershed, Lorestan province, Environmental Erosion Research Journal. 14 (2):161-179. https://doi.org/10.61186/jeer.14.2.10

Memarian H, Abdi Bastami S, Akbari M, Tajbakhsh SM, Azamirad M. 2023. An integrative approach of the physical-based stability index mapping with the maximum entropy stochastic model for risk analysis of mass movements. Environment, Development and Sustainability. 25: 2808–2830. https://doi.org/10.1007/s10668-022-02165-1

Memarian H. Akbari M. 2021. Prediction of combined effect of climate and land use changes on soil erosion in Iran using GloSEM data. Journal of Ecohydrology. 8(2): 513-534. https://doi.org/10.22059/ije.2021.320754.1482. (In Persian).

Mohamadifar A, Gholami H. 2021. Assessment of the interpretability of data mining for the spatial modelling of water erosion using game theory. Catena, 20(10): 51-78). https://doi.org/10.1016/j.catena.2021.105178

Najafi Nia M, Khoramali F, Kiani F, Motlagh MB. 2018. Micromorphology of early Pleistocene loess soils and their comparison with modern loesses. Journal of Agronomy Engineering. 41(4):67–82. (In Persian). https://doi.org/10.22055/agen.2018.22565.1360

Park S, Jungho I, Jang E, Rhee J. 2016. Drought assessment and monitoring through blending of multi-sensor indices using machine learning approaches for different climate regions. Agricultural and Forest Meteorology. 216: 157–169. https://doi.org/10.1016/j.agrformet.2015.10.011

Piri H, Mobaraki M. 2022. Comparison of artificial intelligence and geostatistical methods in soil surface salinity prediction in Ghorghori, Hirmand. Journal of Environmental and Water Engineering. 8(3):622–635. (In Persian). https://doi.org/10.22034/jewe.2022.312194.1656

Raee B, Ahmadi A, Neyshabouri M, Ghorbani M, Asadzadeh F. 2020. Assessment of the efficiency of some artificial intelligence methods in modeling soil wind erodibility in a part of the eastern lands of Urmia Lake. Iranian Journal of Water and Soil Research. 51(1): 61–76. (In Persian). https://doi.org/10.22059/ijswr.2019.283359.668233

Ren Z, Liu C, Zhao X, Jin Y, Ou Y, Liu R, Cheng H. 2024. A large-scale riverbank erosion risk assessment model integrating multi-source data and explainable artificial intelligence (XAI). Ecological Indicators. 166: p. 112575. https://doi.org/10.1016/j.ecolind.2024.112575

Tang X, Li J, Liu M, Liu W, Hong H. 2020. Flood susceptibility assessment based on a novel random Naïve Bayes method: A comparison between different factor discretization methods. Catena, 190. p. 104536.

International Soil and Water Conservation Research. 9(2):217-228. https://doi.org/10.1016/j.iswcr.2020.10.005

Yang X, Yang Q, Zhu H, Wang L, Wang C, Pang G, Hussain S. 2023. Quantitative Evaluation of Soil Water and Wind Erosion Rates in Pakistan. Remote Sensing. 15(9): 1–23. https://doi.org/10.3390/rs15092404

https://doi.org/10.1016/j.catena.2020.104536

Vapnik VN. 1995. The Nature of Statistical Learning Theory. The Nature of Statistical Learning Theory. https://doi.org/10.1007/978-1-4757-2440-0

Wijitkosum S. 2020. Factor influencing land degradation sensitivity and desertification in a drought prone watershed in Thailand.

Fars Agricultural and Natural Resources and Education Center

Watershed Management Research

ISSN: 2981- 2038

Agricultural Research, Education and Extension Organization

Prediction and Modeling of Soil Erosion Rates Using Machine Learning Algorithms in Maraveh Tappeh and Kalaleh Counties, Golestan Province

Abdolhossein Boali[1], Narges Kariminejad[*2], Morteza Akbari[3], Maryam Yaghoobi[4], Mohsen Farahi[5]

1- Postdoctoral Researcher, Department of Desert and Arid Zones Management, Faculty of Natural Resources and Environment, Ferdowsi University of Mashhad, Mashhad, Iran

2- Assistant Professor, Department of Natural Resources and Environmental Engineering, College of Agriculture, Shiraz University, Shiraz, Iran

3- Associate Professor, Department of Desert and Arid Zones Management, Faculty of Natural Resources and Environment, Ferdowsi University of Mashhad, Mashhad, Iran

4- PhD Student, Department of Water Engineering, Faculty of Agriculture, Shahrekord University, Shahrekord, Iran

5- Assistant Professor, Department of Natural Engineering, Faculty of Environmental Sciences and Sustainable Development, University of Saravan, Saravan, Iran

Extended Abstract

Introduction and Goal

Water erosion is one of the major environmental hazards worldwide, leading to soil degradation, reduced fertility, and increased sedimentation in rivers. Golestan Province, particularly its northern regions with sensitive loess soils and specific climatic conditions, is considered one of the main erosion hotspots in Iran. The investigated issue is the limitation of traditional methods in providing accurate, rapid, and reliable zoning of water erosion in the sensitive regions of Golestan Province, where complex and nonlinear relationships among environmental factors require more advanced approaches. Machine learning (ML) algorithms, due to their ability to model complex and non-linear relationships between variables, are powerful tools for the spatial prediction of phenomena such as erosion. The objective of this study was to model the spatial potential of water erosion in the Maraveh Tappeh and Kalaleh counties using advanced machine learning algorithms and to compare their performance with an ensemble model.

Materials and Methods

The study area covers the northern lands of Golestan Province, with an area of approximately 4944.5 km², which is susceptible to water erosion due to its steep topography, seasonal rainfall, and loess soils.

Article Type: Research Article
***Corresponding Authors' E-mail:** narges.karimi991@gmail.com
Citation: Boali, A., Kariminejad, N., Akbari, M., Yaghoobi, M., Farahi, M. 2026. Prediction and Modeling of Soil Erosion Rates Using Machine Learning Algorithms in Maraveh Tappeh and Kalaleh Counties, Golestan Province. Watershed Management Research. 39(2): 80-95.
DOI: 10.22092/WMRJ.2025.371189.1641
Received: 16 November 2025, **Received in revised form:** 03 December 2025, **Accepted:** 21 December 2025
Published online: 22 June 2026
Watershed Management Research, Vol. 39, No. 2, Ser. No. 151, Summer 2026, pp.80-95.
Publisher: Fars Agricultural and Natural Resources and Education Center
©Author(s)

Therefore, first, layers of informational of factors affecting erosion were first prepared.
These factors included eight environmental and climatic indices: Precipitation Index (station data), Vegetation Index (NDVI), Topographic Grain Size Index (TGSI), Land Use layer, Normalized Difference Moisture Index (NDMI), Digital Elevation Model (DEM), Bare Soil Index (BI), and Normalized Difference Salinity Index (NDSI), all of which were extracted from Landsat 8 satellite imagery. A map of observed erosion locations was used as the dependent variable layer in the modeling. Accordingly, 100 points (50 points of water erosion and 50 points of its absence) were taken during the field visit. In this study, 70% of the data was used for training and 30% for testing. Three machine learning algorithms Random Forest (RF), Support Vector Machine (SVM), and Boosted Regression Trees (BRT) were employed to train the models. Finally, an ensemble model was developed based on the weighted average of the outputs of these three individual models to enhance prediction accuracy. The performance of the models was evaluated using statistical indices, including the Kappa coefficient, Receiver Operating Characteristic (ROC) curve, and True Skill Statistic (TSS).

Results and Discussion

The evaluation results indicated that the performance of each of the three individual models (RF, SVM, and BRT) in modeling water erosion was acceptable, however, the ensemble model (Kappa=0.90, ROC=0.93, and TSS=0.89), by leveraging the strengths of all three models, demonstrated superior performance in accurately distinguish erosion-prone from stable zones and in reducing classification errors. This superiority confirms that the ensemble approach, using weighted averaging, uncertainties can be managed more effectively than in any individual model, resulting in a more stable, reliable output. The final erosion potential map, generated by this model was clearly identified that the northern, western, and southwestern parts of the study area are the main hotspots for water erosion. This spatial pattern is not random but is rooted in the complex interaction of factors; in these areas have steep topography, which increases water flow velocity and its erosive power, receive higher rainfall, providing the necessary energy for the erosion process, and possess weaker vegetation cover, which acts as the soil's primary defense. Analysis of variable importance also revealed that precipitation as a driving factor and the Vegetation Index (NDVI) as the resisting factor, are respectively the most influential controls on erosion in the region. This finding indicates that any conservation measure that improve vegetation cover (such as planting, grazing management, and preventing degradation) can directly neutralize the destructive effects of rainfall. This approach is the most effective strategy for controlling erosion in sensitive areas. Therefore, the results of this study not only provide an accurate map but also offer a scientific framework for prioritizing and implementing evidence-based management actions.

Conclusion and Suggestions

The superiority of the ensemble model in this research demonstrates the high reliability of weighted averaging approaches in reducing error and enhancing prediction accuracy. The concentration of high-risk areas in steep slopes with low vegetation cover is fully consistent with the principles of erosion physics. The high importance of precipitation and vegetation cover emphasizes that sustainable land management, such as implementing biological operations (such as planting resistant and deep-rooted species) is the most effective strategy for erosion control in the region, and according to scientific reports, can reduce surface runoff and erosion by 30–50%. Decision-makers, natural resource managers, and urban and rural planners can use the erosion potential map produced in this study as a practical scientific tool for prioritizing conservation measures and optimize resource allocation. Based on the results of this study, it is recommended that future research, while considering existing limitations (such as the spatial resolution of some datasets), make use of more advanced ensemble models and higher-accuracy data.

Keywords: Modeling, machine learning methods, northern Golestan Province, water erosion

Article Type: Research Article

Conflicts of Interest
The authors of this article declare that there are no beneficial conflicts of interest in the writing and publication of the materials and results of this research.

Data Availability Statement
The data and results used in this study will be made available by the corresponding author upon request.

Authors' Contribution
Author 1: Conceptualization, Software, Writing, original draft
Author 2: Supervision, writing, review, editing, Validation
Author 3: Conceptualization, Consultation, Writing, review and editing, Software
Author 4: Investigation, Data Curation
Author 5: Data Curation, Writing, review and editing