

مدلسازی هدررفت خاک ناشی از فرسایش خندقی در مناطق فاقد آمار

چکیده

فرسایش خندقی به‌عنوان یکی از مخرب‌ترین شکل تخریب زمین و هدررفت خاک در سطح جهانی مطرح می‌باشد. باتوجه به زمان‌بر و هزینه‌بر بودن پایش میدانی، این پژوهش به دنبال مدلسازی و برآورد حجم خاک از دست رفته به‌وسیله آن در حوزه آبخیز چوپانلو در استان آذربایجان غربی بود. به این منظور، ابتدا پایش میدانی جهت شناسایی خندق‌ها انجام شد و سپس به منظور خوشه‌بندی خندق‌ها و تعیین خندق‌های منتخب، لایه‌های رقوم‌ی عوامل تأثیرگذار بر گسترش خندق‌ها از جمله عوامل توپوگرافی (ارتفاع، شیب، جهت، انحنای سطح و شاخص موقعیت شیب نسبی)، پوشش گیاهی، کاربری اراضی، سنگ‌شناسی و هیدرواقليم (فاصله از جریان، تراکم زهکشی، شاخص رطوبت توپوگرافی، بارش سالیانه و فراوانی بارش‌های سنگین) تهیه شدند. سپس حجم خاک از دست رفته ناشی از فرسایش خندقی در طی سه سال ۱۴۰۰-۱۴۰۲ برای خندق‌های منتخب به‌عنوان متغیر وابسته در عرصه اندازه‌گیری شد. مدلسازی در این پژوهش با استفاده از سه مدل جنگل تصادفی، ماشین بردار پشتیبان و شبکه عصبی مصنوعی و با رویکرد اعتبارسنجی متقاطع صورت پذیرفت. نتایج فرمول کوکران نشان داد که از بین ۶۷ مورد خندق شناسایی شده در عرصه تعداد ۵۸ مورد حداقل نمونه لازم می‌باشند که این تعداد خندق منتخب پس از خوشه‌بندی از بین سه خوشه شناسایی شده انتخاب شدند. مقدار فرسایش خالص سالانه خاک ناشی از خندق‌های منتخب (۵۸ مورد) به‌ترتیب برابر با ۱۷۲، ۱۹۶ و ۲۰۸ تن در طی سال‌های ۱۴۰۰، ۱۴۰۱ و ۱۴۰۲ می‌باشد. نتایج نشان داد که مدل جنگل تصادفی عملکرد خوب، مدل ماشین بردار پشتیبان عملکرد متوسط و مدل شبکه عصبی عملکرد ضعیفی در مدلسازی داشتند.

کلیدواژه‌ها: حوزه آبخیز چوپانلو، فرسایش خندقی، مدلسازی، هدررفت خاک، یادگیری ماشین

Modeling soil loss due to gully erosion in the data-scarce regions

ABSTRACT

Gully erosion is recognized as a detrimental form of land degradation and soil loss worldwide. Considering the time-consuming and costly nature of field monitoring, this research aimed to develop models for estimating the volume of soil lost due to gully erosion in the Choopanlu watershed, located in West Azerbaijan province, Iran. The study commenced with field monitoring to identify gullies in the area. Following this, digital layers of factors influencing gully erosion were prepared to facilitate gully clustering and selection. These factors included topographical characteristics (elevation, slope, aspect, surface curvature, and relative slope position index), vegetation, land use, soil, lithology, and hydroclimate indicators (distance from stream, drainage density, topographic wetness index, annual precipitation, and frequency of heavy rainfall events). Subsequently, the volume of soil lost due to gully erosion during the three-year period (2021-2023) was measured as the dependent variable for the selected gullies through field observations. In this study, three machine learning models including random forest (RF), support vector machine (SVM), and artificial neural network (ANN) were employed using a cross-validation approach. Cochran's formula results indicated that among the 67 identified gullies in the field, a minimum sample size of 58 gullies was required. Following clustering, this number of selected gullies was chosen from the three identified clusters. The annual soil erosion caused by the selected gullies (i.e., 58 gullies) was estimated to be 172 tons in 2021, 196 tons in 2022, and 208 tons in 2023. According to the modeling results, it can be inferred that the RF model demonstrated the best performance, followed by the SVM model with moderate performance, and the ANN model exhibiting the poorest performance in modeling soil loss due to gully erosion.

Keywords: Choopanlu watershed, Gully erosion, Machine learning, Modeling, Soil loss

مقدمه

هدررفت خاک به عنوان عامل مهم و تهدید کننده در تحقق توسعه اقتصادی، اجتماعی و امنیت غذایی کشور است. این پدیده دارای اثرات منفی همچون کاهش تولیدات کشاورزی، آلودگی منابع آبی، تخریب زیستگاه‌ها، زیست‌بوم‌ها و زیرساخت‌ها و سرمایه‌های تولیدی (ساختمان‌ها، جاده‌ها و دارایی‌های انسانی) می‌باشد (Jie et al., 2002; Pereira et al., 2018). میزان فرسایش خاک در اراضی کشور سالانه ۱۶/۵ تن در هکتار است. در حالی که این میزان در اروپا و آسیا در هر هکتار سالانه حدود شش دهم تن است. بنابراین، سالانه ۲/۷ میلیارد تن از خاک حاصلخیز کشور به دلیل فرسایش از بین می‌رود و سرعت هدررفت خاک ایران در مقایسه با اروپا و سایر کشورهای آسیایی به میزان سه برابر بیشتر است (Mohammadi et al., 2021). این سرعت بیش از حد فرسایش خاک در ایران، تهدیدی برای امنیت غذایی ملی به شمار می‌رود.

از بین شکل‌های مختلف فرسایش خاک و تخریب سرزمین، فرسایش خندقی یکی از مهمترین و شدیدترین نوع آن می‌باشد (Valentin et al., 2005)؛ بطوریکه در مناطق خندقی هدررفت خاک ناشی از آن بسته به شرایط مختلف حوزه‌های آبخیز می‌تواند بین ۱۰ تا ۹۴ درصد از کل فرسایش را بخود اختصاص دهد (Poesen, 2003). این نوع هدررفت خاک تهدید جدی برای منابع طبیعی، اراضی کشاورزی و زیرساخت‌هایی اساسی مثل خطوط ارتباطی و انرژی (برق و گاز) می‌باشد (Martins et al., 2024).

تمرکز مطالعات پیشین در زمینه فرسایش خندقی را می‌توان به چهار بخش اصلی تقسیم کرد. این چهار بخش شامل (۱) تعیین سهم فرسایش خندقی نسبت به فرسایش کل، (۲) ارزیابی متغیرهای اثرگذار بر ایجاد و گسترش فرسایش خندقی، (۳) مطالعه پهنه‌بندی حساسیت مناطق نسبت به وقوع فرسایش خندقی و نهایتاً (۴) برآورد میزان هدررفت و رشد فرسایش خندقی می‌باشند. در زمینه تعیین سهم فرسایش خندقی نسبت به فرسایش کل، می‌توان به پژوهش‌های Nysse et al. (2008) و Tebebu et al. (2010) اشاره داشت که ایشان نشان داد ۲۸ تا ۹۷ درصد از کل هدررفت خاک در کشور اتیوپی ناشی از فرسایش خندقی است. در همین راستا، Zhang et al. (2018) با بررسی سهم فرسایش خندقی در تولید کل رسوب در یک حوزه آبخیز کوچک واقع در جنوب‌غربی چین نشان دادند که در شرایط تحویل ۱۰۰ درصدی رسوب، فرسایش خندقی بیش از ۲۵ درصد کل رسوب را به خود اختصاص می‌دهد. در زمینه ارزیابی متغیرهای اثرگذار بر ایجاد و گسترش فرسایش خندقی، Dube et al. (2014) نشان دادند که وقوع خندق عمدتاً تحت تأثیر پارامترهای ژئومتری و فیزیوگرافی (شامل پوشش زمین، نوع خاک، فاصله از رودخانه، فاصله از جاده، شاخص حمل رسوب، شاخص توان آبراهه و شاخص رطوبت توپوگرافی) می‌باشد. همچنین، Dong et al. (2019) با بررسی ارتباط میان متغیرهای مختلف هواشناسی و میزان رشد فرسایش‌های خندقی در کشور چین (پس از اندازه‌گیری پنج خندق بزرگ در طول شش سال متوالی)، به این نتیجه رسیدند که تغییرات هندسه فرسایش‌های خندقی تحت تأثیر تغییرات عوامل هواشناسی بوده و در سال‌هایی که شدت بارندگی زیاد است، رشد فرسایش خندقی نیز بیشتر بوده است. همچنین، Wang et al. (2021) در مطالعه‌ای در چین گروه‌های مختلفی از خندق‌های تثبیت شده و تثبیت نشده را با استفاده از پهپاد در سال‌های مختلف مورد پایش قرار دادند. خصوصیات محیطی مختلف مانند تراکم زهکشی، شیب زمین، نرخ رواناب سالانه و مساحت محدوده بالادست پیشانی خندق‌ها محاسبه شدند. براساس نتایج حاصله، تغییرات مورفولوژی و اقدامات حفاظت خاک مانند تراس‌ها در بالادست پیشانی خندق‌ها موجب کاهش رشد فرسایش‌های خندقی شده است. همچنین، نرخ رواناب منطقه بالادست ارتباط معنی‌داری با میزان تغییرات سالانه فرسایش‌های خندقی داشت. در همین راستا، Yermolaev et al. (2022) به بررسی پویایی فرسایش‌های خندقی و تأثیرپذیری آن‌ها از تغییرات محیطی در کشور پرداختند. براساس نتایج ایشان، تغییر کاربری اراضی مهمترین عامل محرک در پیشروی فرسایش‌های خندقی بوده که از طریق افزایش رواناب در محدوده بالادست بر افزایش رشد فرسایش‌های خندقی اثر می‌گذارد. در ارتباط با پهنه‌بندی حساسیت مناطق نسبت به وقوع فرسایش خندقی مطالعات متعددی انجام شده است؛ که از بین‌های روش‌های مورد استفاده می‌توان به روش فرآیند تحلیل شبکه‌ای (گرنامی و شادفر، ۱۳۹۷)، شبکه عصبی مصنوعی (Chuma et al., 2023)، جنگل تصادفی (Hasanuzzaman et al., 2024)؛ Ait Naceur et al., 2024)، درخت رگ‌سیون تقویت شده (Garosi et al., 2018; Chuma et al., 2023)، الگوریتم K نزدیکترین همسایه (Avand et al., 2019; Rouhani et al., 2021)، ماشین بردار پشتیبان (Mohebzadeh et al., 2022) و درختان طبقه‌بندی و رگ‌سیون (Mfondoum et al., 2023) اشاره کرد. در ارتباط با برآورد میزان هدررفت خاک و رشد فرسایش خندقی عمدتاً از روش‌های تجربی و

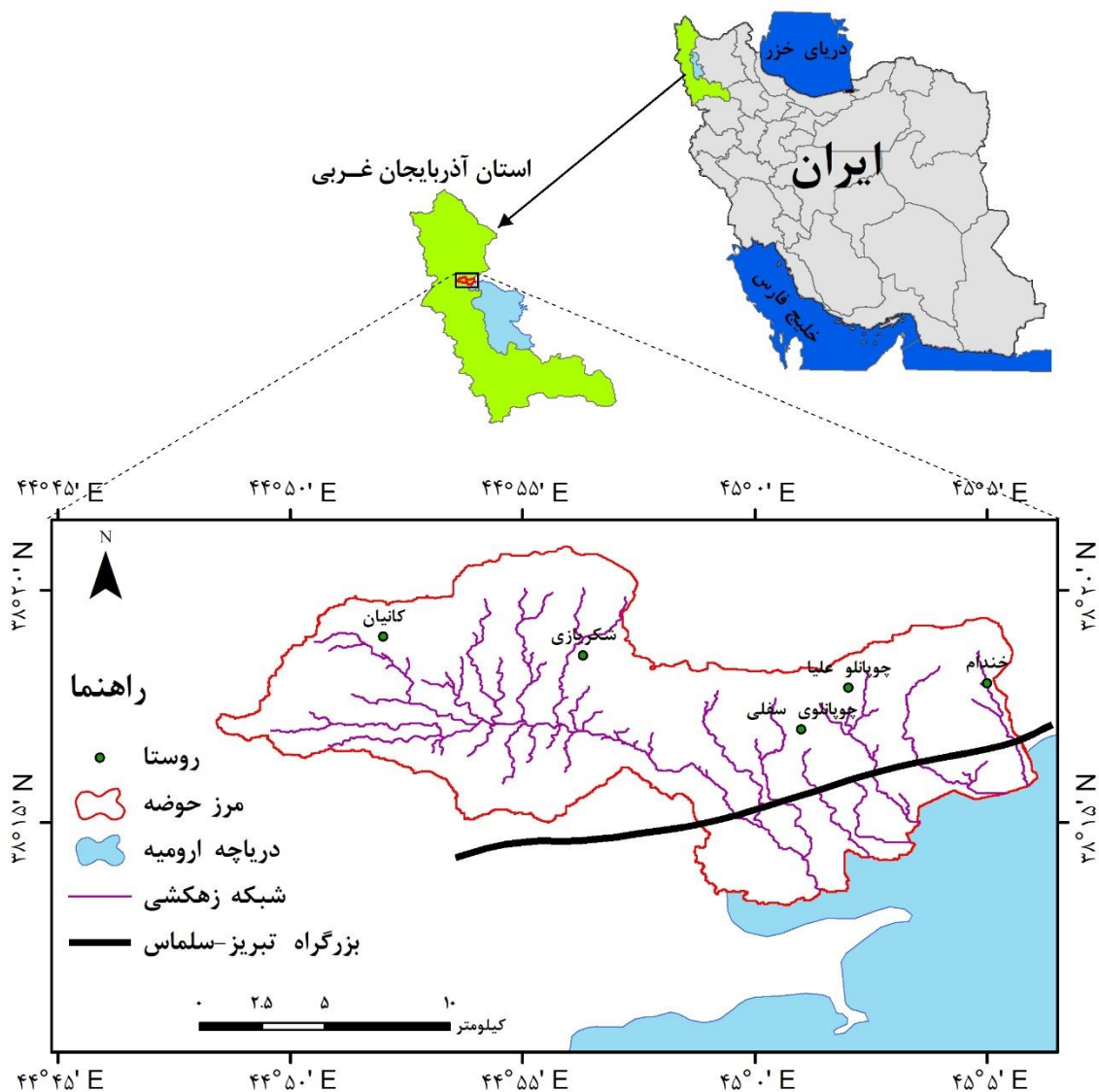
اندازه‌گیری‌های میدانی استفاده شده است. برای مثال (Kheradmand and Soleimanpour (2014) در حوزه آبخیز آبدان استان بوشهر، با استفاده از مدل فائو به پیش‌بینی رشد خندق‌های موجود در منطقه پرداختند. نتایج ایشان نشان داد که مدل فائو تا حدودی قادر به پیش‌بینی رشد خندق‌ها است اما به دلیل سادگی ساختار مدل، نتایج پیش‌بینی انطباق کاملی با واقعیت‌های منطقه ندارد. همچنین (Mukai (2017 میزان هدررفت خاک ناشی از فرسایش‌های خندقی را با استفاده از پایش عکس‌های هوایی و اندازه‌گیری‌های میدانی در اتیوپی تعیین نمود. مطالعات ایشان نشان داد که میانگین هدررفت خاک خندق‌های منطقه مورد مطالعه ۱۶/۲ تن در هکتار در سال بوده است و عوامل توپوگرافی، بارندگی و خصوصیات خاکشناسی تأثیر به‌سزایی در رشد خندق‌ها دارند.

مرور منابع پیشین حاکی از آن است که مدلسازی در زمینه فرسایش خندقی تاکنون معطوف به پهنه‌بندی حساسیت و استعداد اراضی نسبت رخداد فرسایش خندقی در داخل و خارج از کشور بوده است. اما در زمینه برآورد و پیش‌بینی میزان هدررفت خاک ناشی از فرسایش‌های خندقی عمدتاً به روش‌های اندازه‌گیری میدانی و سنتی اتکا شده است. با توجه به اینکه اطلاعات و داده‌های دقیق مربوط به میزان هدررفت خاک ناشی از فرسایش‌های خندقی در بسیاری از حوضه‌های کشور موجود نیست، تهیه و جمع‌آوری این اطلاعات با بازدیدهای مکرر میدانی (جهت اندازه‌گیری و پایش هندسه خندق‌ها) نیازمند هزینه و زمان زیاد می‌باشد. بنابراین، ارائه یک رویکرد نوین برای پیش‌بینی میزان هدررفت خاک ناشی از فرسایش‌های خندقی از اهمیت بسیار بالایی برخوردار است تا با دقت قابل قبولی میزان رشد فرسایش‌های خندقی را مشخص نموده و در زمان و هزینه نیز صرفه‌جویی شود. پیشینه پژوهش‌های انجام شده در زمینه فرسایش خندقی حاکی از آن است که تاکنون رویکردی نوین جهت برآورد یا پیش‌بینی میزان هدررفت خاک ناشی از فرسایش‌های خندقی ارائه نشده است. بر این اساس، این پژوهش به دنبال پر کردن خلاء مطالعاتی در این زمینه بوده و با بهره‌گیری از روش‌های یادگیری ماشین، سعی در ارائه رویکردی به‌جهت پیشبرد این زمینه تحقیقاتی دارد.

روش‌شناسی پژوهش

معرفی منطقه مورد مطالعه

حوزه آبخیز چوپانلو در شمال غربی دریاچه ارومیه واقع شده است و جزء شهرستان سلماس می‌باشد (شکل ۱). این منطقه از نظر شرایط اقلیمی، نیمه‌خشک محسوب شده و از نظر زمین‌شناسی شامل سازندهای زمین‌شناسی با واحدهای سنگی مارن قرمز، مارن گچی، ماسه سنگ و سنگ آهک می‌باشد. شواهد حاکی از نقش عمده چرای بی‌رویه و تغییر کاربری اراضی مرتعی به کشاورزی در گسترش فرسایش خندقی می‌باشد. در برخی از قسمت‌ها، افزایش رواناب سطحی در اثر اقدامات غیر صحیح جاده‌سازی و آبگذرها موجب ایجاد خندق در این منطقه شده است. فرسایش‌های خندقی این حوضه موجب تخریب سرزمین و تولید رسوب زیاد برای مناطق پایین دست شده است. قرار گرفتن دریاچه ارومیه در قسمت خروجی جریان این حوضه اهمیت مطالعات فرسایش و رسوب را چند برابر نموده است، زیرا رسوبات ناشی از این حوضه حیات قسمت شمالی دریاچه ارومیه را نیز تهدید می‌کند. همچنین شواهد حاکی از آن است که بسیاری از خطوط انتقال نیرو (برق و گاز) نیز در این حوضه تحت تهدید قرار دارند.



شکل ۱. موقعیت حوزه آبخیز چوپانلو

بازدید میدانی و ثبت موقعیت تمام خندق‌های منطقه مورد مطالعه

به منظور شناسایی موقعیت خندق‌ها، از کارشناسان اداره منابع طبیعی شهرستان سلماس، افراد بومی روستاها و همچنین تصاویر گوگل ارث استفاده شد. بر این اساس، پس از بازدید میدانی و مشاهده مستقیم خندق‌های حوزه آبخیز چوپانلو، موقعیت تمامی خندق‌ها با استفاده از دستگاه GPS ثبت شد.

تعیین تعداد خندق‌ها برای اندازه‌گیری میدانی

تعداد خندق‌های مورد نیاز برای اندازه‌گیری میدانی با استفاده از فرمول کوکران تعیین گردید (Cochran, 1977) (رابطه ۱):

$$n = \frac{\frac{Z^2 pq}{d^2}}{1 + \frac{1}{N} \left(\frac{Z^2 pq}{d^2} - 1 \right)} \quad \text{(رابطه ۱)}$$

که در آن n حجم نمونه آماری و N حجم جامعه آماری هستند. پارامتر d اشتباه مجاز است که معمولاً $0/05$ در نظر گرفته می‌شود. پارامتر Z مقدار متغیر نرمال با سطح اطمینان $1-a$ است. در آزمون دو دامنه مقدار Z برای سطح اطمینان 95 درصد برابر $1/96$ و برای سطح 99 درصد برابر $2/58$ است. پارامتر p نسبت برخورداری از صفت مورد نظر و پارامتر q نسبت عدم برخورداری از صفت مورد نظر است. معمولاً مقدار p و q را $0/5$ در نظر گرفته می‌شود. در این مرحله کفایت داده‌ها از لحاظ آماری مدنظر قرار گرفت.

تعیین خندق‌های معرف با استفاده از روش خوشه‌بندی

پس از تعیین تعداد حداقل نمونه لازم، انتخاب خندق‌های معرف از طریق خوشه‌بندی انجام شد. در واقع خوشه‌بندی بدین منظور انجام گرفت تا خندق‌های انتخابی معرف خندق‌های کل منطقه باشند و از تمامی خوشه‌ها نمونه وجود داشته باشد. بنابراین بر اساس خصوصیات محیطی خندق‌ها (شامل ارتفاع، شیب زمین، جهت دامنه، انحنای زمین، شاخص موقعیت شیب نسبی، سنگ‌شناسی، فاصله از آبراهه، تراکم زهکشی، فاصله از جاده، شاخص رطوبت توپوگرافی، کاربری اراضی و بارش) در گروه‌های مختلف خوشه‌بندی شده تا از هر خوشه تعدادی خندق به عنوان خندق‌های معرف انتخاب گردد. در این پژوهش، روش خوشه‌بندی سلسله‌مراتبی با بهره‌گیری از تابع مربع فاصله اقلیدسی (جهت اندازه‌گیری فاصله بین مشاهدات) جهت شناسایی خوشه‌ها استفاده شد.

نمونه‌برداری و اندازه‌گیری میدانی

به منظور پایش رشد خندق‌ها و اندازه‌گیری میزان هدررفت خاک، ابعاد هندسی مقاطع مدنظر در هر خندق شامل عرض بالا، عرض پایین، طول و عمق خندق در طی چهار سال (۱۳۹۹-۱۴۰۲) و هر سال در شروع سال آبی (مهرماه هر سال) صورت پذیرفت. اندازه‌گیری پارامترهای ابعادی خندق‌ها با مینا قرار دادن پژوهش (Casalí et al., 2006). در فواصل ۱۰ متری از سر هر خندق تا انتهای آن انجام شد. محاسبه‌ی سطح مقطع عرضی خندق، از رابطه ۲ (Tebebu et al., 2010) در چهار مقطع زمانی مختلف (سال اول، سال دوم، سال سوم و سال چهارم) انجام شد:

$$CA = \sum_{i=1}^n \left(\left(\frac{W_1 + W_2}{2} \right) D \right) \quad \text{رابطه ۲}$$

در این رابطه؛ CA : سطح مقطع عرضی خندق (متر مربع)، W_1 : عرض بالای خندق (متر)، W_2 : عرض پایین خندق (متر)، و D : عمق خندق (متر) می‌باشد.

محاسبه حجم خاک از دست رفته ناشی از فرسایش خندقی نیز طبق رابطه ۳ (Tebebu et al., 2010) محاسبه شد:

$$V = \sum_{i=1}^n L \left[\left(\left(\frac{W_1 + W_2}{2} \right) D \right) \right] \quad \text{رابطه ۳}$$

در این رابطه؛ V : حجم خاک از دست رفته ناشی از فرسایش خندقی (متر مکعب)، L : طول خندق (متر)، W_1 : عرض بالای خندق (متر)، W_2 : عرض پایین خندق (متر)، و D : عمق خندق (متر) است.

نهایتاً محاسبه وزن هدررفت خاک در هر خندق از طریق ضرب مقدار حجم خاک از دست‌رفته (V) در جرم مخصوص ظاهری خاک محاسبه شد و با کم کردن از وزن هدررفت خاک در سال پیش میزان هدررفت خالص سالیانه محاسبه شد. بنابراین با برداشت چهار سال داده میدانی (از مهر ۱۳۹۹ تا مهر ۱۴۰۲)، میزان هدررفت خاک برای سه سال آبی ۱۳۹۹-۱۴۰۰، ۱۴۰۰-۱۴۰۱ و ۱۴۰۱-۱۴۰۲ محاسبه شد.

استخراج متغیرهای پیش‌بینی کننده هدررفت خاک ناشی از فرسایش خندقی

عوامل پیش‌بینی کننده در این پژوهش شامل مجموعه‌ای از فاکتورهای توپوگرافی (ارتفاع، شیب، جهت، انحنای سطح و شاخص موقعیت

شیب نسبی)، پوشش گیاهی، کاربری اراضی، بافت خاک، سنگ شناسی و هیدرواقليم (فاصله از جریان، تراکم زهکشی، شاخص رطوبت توپوگرافی، بارش سالیانه و فراوانی بارش های سنگین) می باشند. علت انتخاب این عوامل به این سبب است که عوامل توپوگرافی تأثیر به سزایی در فرایندهای هیدرولوژیکی و فرسایش دارند و از این رو به عنوان متغیرهای پیش بینی کننده مورد استفاده قرار گرفتند. خاک شناسی یکی از عوامل مهم در وقوع و گسترش فرسایش خندقی است به گونه ای که سرعت رشد خندقها در خاکهای حساس بسیار بالاتر است. عامل سنگ شناسی همواره میزان حساسیت واحد سنگ شناسی به فرسایش را مشخص می کند و می تواند اطلاعات مفیدی برای تخمین فرسایش خندقی را ارائه نماید. همچنین، از آنجایی که آبراهه ها (زهکش ها)، جریان های سطحی و زیرسطحی محدوده های اطراف و بالادست را جمع آوری می کنند، همگرایی جریان ها می تواند پتانسیل بالایی برای وقوع و رشد فرسایش های خندقی داشته باشد، از این رو عامل فاصله از جریان به عنوان یکی از متغیرهای پیش بینی کننده انتخاب شده است.

تراکم زهکشی نیز نشان دهنده ی میزان تراکم جریان ها در واحد سطح منطقه است و بر پتانسیل فرسایش خاک و فرایندهای هیدرولوژیکی تأثیرگذار است. شاخص رطوبت توپوگرافی نیز به عنوان یک شاخص ترکیبی توپوگرافی-هیدرولوژیکی، میزان جریان بالادست پیکسل و وضعیت شیب پیکسل را همزمان در نظر می گیرد تا میزان جریان بالقوه و شرایط نفوذ را نشان دهد. بارش و فراوانی بارش های سنگین (بیشتر از ۵ میلی متر) نیز از عوامل اصلی وقوع فرسایش آبی است که به عنوان یکی از عوامل مهم پیش بینی کننده در نظر گرفته شد. لایه های اطلاعاتی این متغیرها در سیستم اطلاعات جغرافیایی در مقیاس ۱:۵۰۰۰۰ تا ۱:۱۰۰۰۰۰ تهیه شدند. لازم به توضیح می باشد که برخی با زمان تغییر می کنند، بنابراین جهت برآورد هدررفت خاک در هر سال از متغیرهای پوشش گیاهی، بارش سالیانه و فراوانی بارش های سنگین همان سال استفاده شده است.

مدلسازی و برآورد هدررفت خاک ناشی از فرسایش خندقی

مدلسازی در این پژوهش با رویکرد اعتبارسنجی متقاطع انجام شد و داده ها به تعداد سه دسته ($k=3$) تقسیم شدند. با توجه به این که میزان هدررفت خالص خاک در سه سال اندازه گیری شده بود، هر بار یکی از سال ها به عنوان داده های اعتبارسنجی کنار گذاشته شد و آموزش مدل ها با دو سال باقیمانده صورت پذیرفت. جدول ۱ گروه داده های در نظر گرفته شده برای ساخت و اعتبارسنجی مدل ها را نشان می دهد.

جدول ۱. رویکرد اعتبارسنجی متقاطع جهت تفکیک داده ها

گروه	داده های آموزش	داده های اعتبارسنجی
اول	۱۴۰۰ - ۱۴۰۱	۱۴۰۲
دوم	۱۴۰۰ و ۱۴۰۲	۱۴۰۱
سوم	۱۴۰۱ - ۱۴۰۲	۱۴۰۰

در فرآیند مدل سازی، عوامل محیطی به عنوان متغیرهای مستقل و وزن هدررفت خاک خندقها به عنوان متغیر وابسته در نظر گرفته شد. مدل ها براساس داده های گروه آموزش، آموزش داده شد و با استفاده از داده های کنارگذاشته شده مورد صحتسنجی قرار گرفتند (جدول ۱). در این پژوهش، سه مدل مختلف یادگیری ماشین شامل ماشین بردار پشتیبان^۱ (SVM)، جنگل تصادفی^۲ (RF) و شبکه ی عصبی مصنوعی^۳ (ANN) برای مدل سازی در نظر گرفته شد و مدل سازی در محیط نرم افزار R و از طریق کتابخانه Caret (Kuhn, 2015) انجام شد. دقت مدل ها

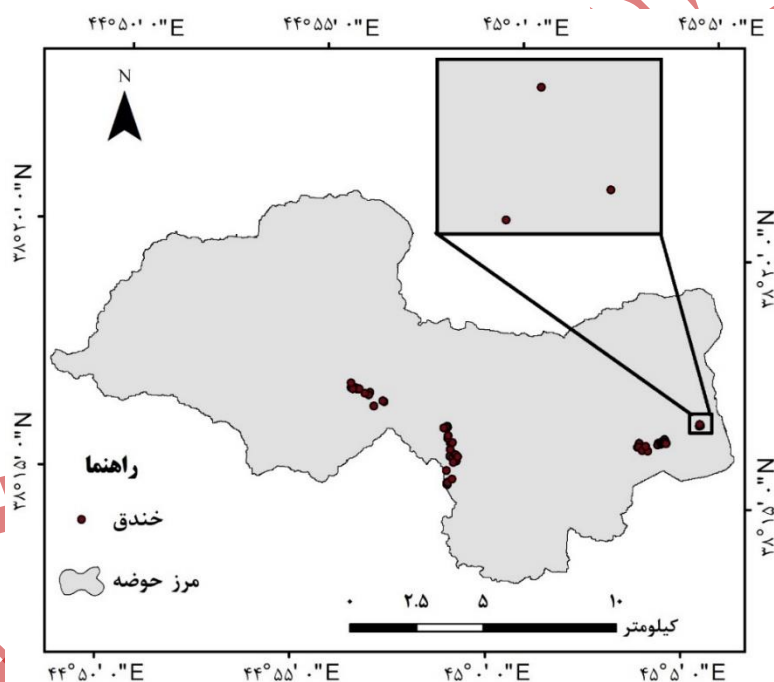
1. Support vector machine
2. Random forest
3. Artificial neural network

در این مرحله نیز با استفاده از معیارهای کمی مانند خطای جذر میانگین مربعات (RMSE)، شاخص تبیین (R^2)، شاخص خطای RSR (نسبت RMSE به انحراف معیار داده‌های مشاهداتی) و شاخص تطابق (d) (Willmott, 1981) مورد بررسی قرار گرفت. مقادیر کمتر معیارهای RMSE و RSR و مقادیر بیشتر شاخص‌های تبیین و تطابق به منزله عملکرد بهتر مدل است.

یافته‌های پژوهش

شناسایی مناطق خندقی

بازدیدهای میدانی متعددی در حوزه‌ی آبخیز چوپانلو انجام گرفت و خندق‌های موجود شناسایی شدند (شکل ۲). همانطور که در شکل مشخص است، خندق‌های شناسایی شده را در سه منطقه مرکزی، جنوبی و جنوب شرقی حوزه آبخیز چوپانلو قرار گرفته‌اند. از نظر کاربری اراضی، خندق‌های واقع در مناطق مرکزی دارای کاربری دیم بوده، خندق‌های واقع در جنوب عمدتاً کاربری مرتعی داشته و خندق‌های واقع در جنوب شرقی حوضه در اراضی شورزار واقع شده‌اند. از نظر زمین شناسی، خندق‌ها عمدتاً در واحدهای سنگ شناسی OMrb (کنگلومرای قرمز، ماسه سنگ، مارن و مارن گچی) و Mur (مارن قرمز، مارن گچی، ماسه سنگ و کنگلومرا - سازند قرمز بالایی) واقع شده‌اند.



شکل ۲. نقشه موقعیت کل خندق‌های حوزه آبخیز چوپانلو استان آذربایجان غربی

انتخاب خندق‌های معرف

پس از تعیین دقیق کل خندق‌های موجود در حوضه، به منظور تعیین تعداد خندق‌های مورد نیاز (منتخب) برای اندازه‌گیری‌های میدانی، از فرمول کوکران و در نظر گرفتن خطای ۵ درصد استفاده شد. در واقع، حجم نمونه‌ای که با فرمول کوکران محاسبه می‌شود، حداقل حجم قابل قبول در تجزیه و تحلیل‌های آماری است. بر این اساس، از تعداد ۶۷ خندق موجود در حوزه آبخیز چوپانلو ۵۸ خندق حداقل نمونه لازم می‌باشد.

نتایج خوشه‌بندی خندق‌ها در جدول ۲ ارائه شده است. همانطور که مشخص است، خندق‌ها در سه خوشه همگن با در نظر گرفتن درصد مشابهت درون گروهی بالا واقع شده‌اند. بطوریکه در خوشه اول ۳۲ خندق، در خوشه دوم ۱۶ خندق و در خوشه سوم ۱۹ خندق قرار گرفته است. میزان

مشابهت خندق‌های خوشه اول ۸۶/۱ درصد، میزان مشابهت خندق‌های خوشه دوم ۷۶/۱ درصد و میزان مشابهت خندق‌های خوشه سوم ۸۳/۳ درصد است. بر این اساس، تعداد ۵۸ خندق از بین تمامی خوشه‌ها انتخاب شد که مشخصات آنها در جدول ۳ ارائه شده است.

جدول ۲. تقسیم‌بندی خندق‌ها در خوشه‌های مختلف

خوشه	شماره خندق	درصد شباهت
۱	۱, ۲, ۳, ۴, ۵, ۶, ۷, ۸, ۹, ۱۰, ۱۱, ۱۲, ۱۳, ۱۴, ۱۵, ۱۶, ۱۷, ۱۸, ۱۹, ۲۰, ۲۱, ۲۲, ۲۳, ۲۴, ۲۵, ۲۶, ۲۷, ۲۸, ۲۹, ۳۰, ۳۱, ۳۲	۸۶/۱
۲	۳۳, ۳۴, ۳۵, ۳۶, ۳۷, ۳۸, ۳۹, ۴۰, ۴۱, ۴۲, ۴۳, ۴۴, ۴۵, ۵۷, ۵۸, ۶۷	۷۶/۱
۳	۴۶, ۴۷, ۴۸, ۴۹, ۵۰, ۵۱, ۵۲, ۵۳, ۵۴, ۵۵, ۵۶, ۵۹, ۶۰, ۶۱, ۶۲, ۶۳, ۶۴, ۶۵, ۶۶	۸۳/۳

جدول ۳. مشخصات خندق‌های منتخب در زمان شروع نمونه‌برداری (سال ۱۳۹۹)

شماره خندق منتخب	طول جغرافیایی	عرض جغرافیایی	طول (متر)	عرض جغرافیایی	عرض جغرافیایی	شماره خندق منتخب	میانگین عمق (متر)	میانگین عرض پایین (متر)	میانگین عرض بالا (متر)	میانگین عرض عمق (متر)	طول (متر)	عرض جغرافیایی	طول جغرافیایی
۱	۵۰۵۲۶۸	۴۲۳۵۷۹۷	۲۴	۵/۷۸	۲/۱۰	۱/۴۵	۲۸	۲/۱۰	۵/۷۸	۲/۱۰	۲۴	۴۲۳۵۷۹۷	۵۰۵۲۶۸
۲	۵۰۵۲۲۴	۴۲۳۵۶۴۹	۱۶۸	۱۰/۲۵	۶/۵۴	۱/۳۹	۳۹	۶/۵۴	۱۰/۲۵	۶/۵۴	۱۶۸	۴۲۳۵۶۴۹	۵۰۵۲۲۴
۶	۵۰۵۴۰۰	۴۲۳۵۵۳۱	۳۰	۱/۵۴	۰/۷۸	۰/۵۱	۴۰	۰/۷۸	۱/۵۴	۰/۷۸	۳۰	۴۲۳۵۵۳۱	۵۰۵۴۰۰
۷	۵۰۵۶۰۷	۴۲۳۵۵۳۰	۳۶	۶/۹۳	۴/۲۳	۰/۹۸	۴۱	۴/۲۳	۶/۹۳	۴/۲۳	۳۶	۴۲۳۵۵۳۰	۵۰۵۶۰۷
۸	۵۰۵۵۱۷	۴۲۳۵۶۹۸	۱۸	۵/۵۳	۲/۴۳	۰/۹۵	۴۲	۲/۴۳	۵/۵۳	۲/۴۳	۱۸	۴۲۳۵۶۹۸	۵۰۵۵۱۷
۹	۵۰۵۹۷۵	۴۲۳۵۸۳۹	۱۸	۶/۳۰	۴/۹۳	۱/۲۳	۴۳	۴/۹۳	۶/۳۰	۴/۹۳	۱۸	۴۲۳۵۸۳۹	۵۰۵۹۷۵
۱۰	۵۰۵۹۷۳	۴۲۳۵۷۶۹	۶۶	۶/۲۵	۴/۳۶	۰/۶۲	۴۴	۴/۳۶	۶/۲۵	۴/۳۶	۶۶	۴۲۳۵۷۶۹	۵۰۵۹۷۳
۱۱	۵۰۶۰۸۹	۴۲۳۵۸۰۸	۱۲	۲/۲۰	۱/۰۰	۰/۷۵	۴۵	۱/۰۰	۲/۲۰	۱/۰۰	۱۲	۴۲۳۵۸۰۸	۵۰۶۰۸۹
۱۲	۵۰۶۰۸۷	۴۲۳۵۸۲۰	۵۰	۲/۸۴	۱/۶۸	۰/۹۲	۴۶	۱/۶۸	۲/۸۴	۱/۶۸	۵۰	۴۲۳۵۸۲۰	۵۰۶۰۸۷
۱۳	۵۰۶۱۰۲	۴۲۳۵۸۴۵	۴۴	۴/۹۹	۲/۳۳	۰/۷۷	۴۷	۲/۳۳	۴/۹۹	۲/۳۳	۴۴	۴۲۳۵۸۴۵	۵۰۶۱۰۲
۱۴	۵۰۶۱۳۳	۴۲۳۵۸۹۷	۴۲	۳/۳۷	۰/۹۳	۱/۱۰	۴۸	۰/۹۳	۳/۳۷	۰/۹۳	۴۲	۴۲۳۵۸۹۷	۵۰۶۱۳۳
۱۵	۵۰۶۱۴۶	۴۲۳۵۹۱۱	۲۴	۴/۲۰	۱/۲۳	۰/۷۸	۴۹	۱/۲۳	۴/۲۰	۱/۲۳	۲۴	۴۲۳۵۹۱۱	۵۰۶۱۴۶
۱۶	۵۰۶۱۵۶	۴۲۳۵۹۲۰	۱۸	۳/۲۳	۰/۶۷	۱/۰۷	۵۰	۰/۶۷	۳/۲۳	۰/۶۷	۱۸	۴۲۳۵۹۲۰	۵۰۶۱۵۶
۱۷	۵۰۶۱۵۳	۴۲۳۵۹۳۵	۳۶	۲/۴۵	۰/۸۰	۰/۹۳	۵۱	۰/۸۰	۲/۴۵	۰/۸۰	۳۶	۴۲۳۵۹۳۵	۵۰۶۱۵۳
۱۸	۵۰۶۱۸۹	۴۲۳۵۹۲۵	۳۰	۳/۰۰	۰/۶۴	۱/۰۰	۵۲	۰/۶۴	۳/۰۰	۰/۶۴	۳۰	۴۲۳۵۹۲۵	۵۰۶۱۸۹
۲۰	۵۰۶۲۵۲	۴۲۳۵۹۱۳	۲۴	۳/۱۵	۱/۳۴	۰/۸۵	۵۳	۱/۳۴	۳/۱۵	۱/۳۴	۲۴	۴۲۳۵۹۱۳	۵۰۶۲۵۲
۲۱	۵۰۶۲۸۳	۴۲۳۵۸۵۵	۱۸	۴/۵۳	۲/۱۷	۰/۸۷	۵۴	۲/۱۷	۴/۵۳	۲/۱۷	۱۸	۴۲۳۵۸۵۵	۵۰۶۲۸۳
۲۲	۵۰۶۲۶۷	۴۲۳۵۸۳۶	۱۲	۲/۹۵	۱/۵۰	۰/۶۵	۵۵	۱/۵۰	۲/۹۵	۱/۵۰	۱۲	۴۲۳۵۸۳۶	۵۰۶۲۶۷
۲۳	۵۰۷۴۸۴	۴۲۳۶۵۸۲	۲۴	۲/۴۳	۱/۱۰	۰/۶۵	۵۶	۱/۱۰	۲/۴۳	۱/۱۰	۲۴	۴۲۳۶۵۸۲	۵۰۷۴۸۴
۲۴	۵۰۷۵۳۱	۴۲۳۶۵۹۹	۳۶	۲/۶۰	۱/۰۸	۰/۹۸	۵۷	۱/۰۸	۲/۶۰	۱/۰۸	۳۶	۴۲۳۶۵۹۹	۵۰۷۵۳۱
۲۶	۴۹۸۲۱۸	۴۲۳۳۷۸۷	۵۸	۱/۳۱	۰/۶۹	۰/۲۴	۵۸	۰/۶۹	۱/۳۱	۰/۶۹	۵۸	۴۲۳۳۷۸۷	۴۹۸۲۱۸
۲۸	۴۹۸۲۲۰	۴۲۳۳۸۸۶	۱۳۸	۲/۶۸	۱/۶۷	۰/۶۵	۵۹	۱/۶۷	۲/۶۸	۱/۶۷	۱۳۸	۴۲۳۳۸۸۶	۴۹۸۲۲۰
۳۱	۴۹۸۳۶۹	۴۲۳۴۰۰۱	۱۲	۱/۳۰	۰/۵۰	۰/۵۰	۶۱	۰/۵۰	۱/۳۰	۰/۵۰	۱۲	۴۲۳۴۰۰۱	۴۹۸۳۶۹
۳۲	۴۹۸۱۴۶	۴۲۳۴۳۱۳	۱۸	۱/۴۰	۰/۵۰	۰/۴۲	۶۲	۰/۵۰	۱/۴۰	۰/۵۰	۱۸	۴۲۳۴۳۱۳	۴۹۸۱۴۶
۳۳	۴۹۸۲۳۵	۴۲۳۴۴۸۱	۸۴	۱/۵۶	۰/۶۴	۰/۶۹	۶۳	۰/۶۴	۱/۵۶	۰/۶۴	۸۴	۴۲۳۴۴۸۱	۴۹۸۲۳۵
۳۴	۴۹۸۳۴۶	۴۲۳۴۸۰۰	۶	۱/۳۰	۰/۳۰	۰/۳۰	۶۴	۰/۳۰	۱/۳۰	۰/۳۰	۶	۴۲۳۴۸۰۰	۴۹۸۳۴۶
۳۵	۴۹۸۴۹۵	۴۲۳۴۸۹	۱۴	۱/۲۷	۰/۴۰	۰/۵۵	۶۵	۰/۴۰	۱/۲۷	۰/۴۰	۱۴	۴۲۳۴۸۹	۴۹۸۴۹۵
۳۶	۴۹۸۳۸۱	۴۲۳۴۳۷	۱۵	۲/۰۰	۱/۱۰	۰/۵۳	۶۶	۱/۱۰	۲/۰۰	۱/۱۰	۱۵	۴۲۳۴۳۷	۴۹۸۳۸۱
۳۷	۴۹۸۴۵۲	۴۲۳۴۹۱۴	۱۵	۱/۵۷	۰/۶۰	۰/۴۲	۶۷	۰/۶۰	۱/۵۷	۰/۴۲	۱۵	۴۲۳۴۹۱۴	۴۹۸۴۵۲

شماره خندق منتخب در این جدول متناسب با شماره خندق‌های ارائه شده در جدول ۲ می‌باشد.

هدررفت سالیانه خاک ناشی از فرسایش خندقی

از طریق عملیات میدانی اندازه‌گیری ابعاد هندسی خندق‌های منتخب در حوزه آبخیز چوپانلو در سال‌های ۱۳۹۹، ۱۴۰۰، ۱۴۰۱ و ۱۴۰۲ شامل طول خندق، میانگین عرض بالا، میانگین عرض پایین و میانگین عمق خندق جهت محاسبه وزن خالص هدررفت خاک در هر سال انجام شد. به‌عنوان نمونه، نتایج اندازه‌گیری‌های سال ۱۳۹۹ در جدول ۲ ارائه شده است. همچنین در شکل ۳ نمونه‌ای از تصاویر برداشت‌های میدانی مشاهده می‌شود.

لازم به ذکر است که مقدار وزن هدررفت خاک به‌دست آمده در سال اول اندازه‌گیری (۱۳۹۹)، مربوط به تاریخچه عمر خندق است. به عبارتی دیگر، وزن هدررفت خاک از ابتدای شکل‌گیری خندق تا سال ۱۳۹۹ محاسبه شد و اطلاعات هندسی اندازه‌گیری شده خندق‌ها در سال ۱۳۹۹ مبنای محاسبه‌ی میزان هدررفت خاک در سال‌های بعدی (۱۴۰۰، ۱۴۰۱ و ۱۴۰۲) بود و به این صورت، امکان تعیین میزان هدررفت خالص سالانه فراهم شد؛ به‌طوری‌که مقدار خالص میزان هدررفت خاک در سال ۱۴۰۰ با استفاده از تفاضل وزن هدررفت خاک در سال ۱۴۰۰ از سال قبل (۱۳۹۹) بدست آمد و برای سال‌های بعدی نیز به همین ترتیب انجام شد. در جدول ۴ نتایج هدررفت خالص سالانه خاک (تن در سال) برای هر یک از خندق‌ها ارائه شده است. مقدار هدررفت سالانه خاک ناشی از این خندق‌های منتخب به‌ترتیب برابر با ۱۷۲، ۱۹۶ و ۲۰۸ تن در سال‌های ۱۴۰۰، ۱۴۰۱ و ۱۴۰۲ می‌باشد.



شکل ۳. نمونه‌ای از تصاویر برداشت‌های میدانی از خندق‌ها در حوزه آبخیز چوپانلو استان آذربایجان غربی (ثبت شده توسط نویسندگان)

جدول ۴. هدررفت خالص سالانه خاک (تن در سال)

شماره خندق	۱۴۰۰	۱۴۰۱	۱۴۰۲	شماره خندق	۱۴۰۰	۱۴۰۱	۱۴۰۲
۱	۲/۱۲	۲/۶۹	۳/۷۹	۳۰	۱/۰۲	۰/۷۷	۰/۷۷
۲	۴۰/۲۵	۴۷/۲۳	۲۸/۴۵	۳۱	۱/۳۶	۷/۸۳	۱/۰۲
۳	-/۹۴	-/۸۸	-/۷۱	۳۲	۱۸/۵۲	۳۷/۷۱	۱۴/۰۸
۴	۴/۷۸	۳/۵۹	۷/۰۲	۳۳	۱/۱۲	-/۸۴	۱/۲۲
۵	۱/۷۸	۱/۳۳	۱/۳۳	۳۴	-/۵۷	-/۴۳	۱۰/۱۹
۶	۲/۵۵	۳/۷۱	۱/۲۰	۳۵	۱۹/۴۳	۹/۹۳	۲۷/۰۷
۷	۸/۶۹	۶/۵۲	۸/۲۸	۳۶	-/۵۶	-/۴۲	-/۶۸
۸	-/۴۸	-/۳۶	-/۳۶	۳۷	۲/۳۱	۱/۷۳	۲/۲۵
۹	۲/۸۵	۲/۱۴	۲/۱۴	۳۸	-/۵۳	۱/۶۳	۱۶/۷۸
۱۰	۴/۰۵	۳/۰۴	۳/۰۴	۳۹	-/۱۷	-/۱۳	-/۲۵
۱۱	۲/۲۶	۲/۸۲	۲/۶۳	۴۰	-/۲۴	-/۱۸	-/۵۴
۱۲	۱/۶۱	۱/۲۱	۲/۰۴	۴۱	-/۶۱	۱/۰۶	۱/۲۰
۱۳	-/۸۷	-/۶۵	۱/۰۷	۴۲	-/۳۴	-/۲۵	۱/۶۴
۱۴	۱/۴۶	۱/۱۰	۱/۵۰	۴۳	-/۱۴	-/۱۱	-/۲۱
۱۵	۱/۳۴	۱/۰۱	۱/۰۱	۴۴	-/۱۶	-/۱۲	-/۴۷
۱۶	۱/۳۱	-/۹۹	-/۹۹	۴۵	-/۲۳	-/۱۷	-/۳۹
۱۷	۱/۵۲	۱/۱۴	۱/۵۲	۴۶	۱/۵۲	۱/۱۴	۱/۳۶
۱۸	-/۶۷	-/۵۰	-/۹۰	۴۷	-/۹۰	-/۶۷	-/۸۷
۱۹	۱/۰۹	۱/۲۹	-/۸۳	۴۸	۲/۳۲	۱/۷۴	۱/۷۴
۲۰	۲/۱۷	۱/۵۲	۱/۲۸	۴۹	۱/۹۵	۱/۴۶	۱/۶۲
۲۱	۱/۳۵	۱/۰۲	۱/۰۲	۵۰	۲/۲۵	۱/۶۸	۴/۰۳
۲۲	۷/۰۹	۵/۵۹	۶/۵۹	۵۱	۲/۹۷	۳/۹۸	۴/۰۳
۲۳	-/۲۶	-/۱۹	-/۷۸	۵۲	۶/۶۲	۸/۲۸	۶/۱۳
۲۴	-/۴۲	-/۳۲	-/۵۹	۵۳	۲/۷۹	۲/۰۹	۴/۴۳
۲۵	۲/۱۸	۱۲/۵۴	۱/۶۴	۵۴	۱/۴۸	۱/۱۱	۳/۴۰
۲۶	-/۱۱	-/۶۸	۱/۴۱	۵۵	-/۱۶	-/۶۴	۱/۸۱
۲۷	-/۲۸	-/۳۹	-/۶۳	۵۶	-/۵۱	-/۳۸	۱/۲۰
۲۸	-/۵۶	-/۸۳	۳/۲۰	۵۷	-/۳۹	-/۹۹	-/۸۹
۲۹	-/۳۹	-/۲۹	-/۲۹	۵۸	۴/۵۵	۳/۴۱	۳/۴۱

مدلسازی هدررفت خاک ناشی از فرسایش خندقی

بررسی همخطی چندگانه بین متغیرهای مستقل

قبل از شروع مدلسازی، همخطی چندگانه بین متغیرهای مستقل بررسی شد. نتایج نشان داد که دو فاکتور فاصله از جاده و ارتفاع از سطح دریا مقادیر همخطی بالایی را بخود اختصاص داده‌اند (مقادیر شاخص تورم واریانس به ترتیب برابر با ۳۹/۳۶ و ۳۶/۳۶ بودند). با توجه به اینکه مناطق خندقی تغییرات ارتفاعی زیادی نداشتند و عمده خندق‌ها در ارتفاعات پایین قرار داشتند و از طرفی برخی از خندق‌ها متأثر از جاده ایجاد شده و یا خطوط جاده‌ای را تهدید می‌کنند، بجای حذف هر دوی متغیرها ابتدا متغیر ارتفاع حذف شد و نتایج همخطی مجدداً

بررسی شد. نتایج نشان داد که با حذف متغیر ارتفاع، میزان همخطی بین متغیرها کاهش پیدا می‌کند و نیازی به حذف متغیر فاصله از جاده که اهمیت بیشتری در سطح عرصه دارد، نیست. بنابراین مدلسازی میزان هدررفت خندق با حذف تنها یک متغیر (ارتفاع) انجام شد (جدول ۵).

جدول ۵. نتایج بررسی همخطی چندگانه بین متغیرها

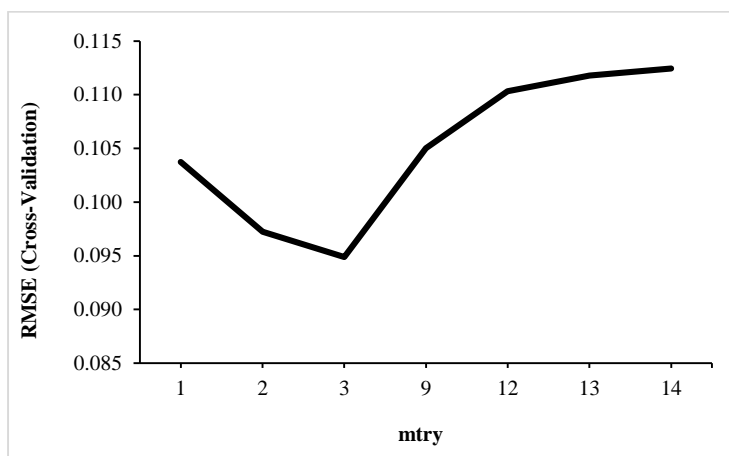
شاخص تورم واریانس		متغیر
نهایی	اولیه	
۷/۲۷	۳۹/۳۶	فاصله از جاده
متغیر حذف شده	۳۶/۳۶	ارتفاع از سطح دریا
۷/۶۷	۷/۶۷	تراکم زهکشی
۲/۸۳	۳/۳۴	کاربری اراضی
۳/۱۲	۳/۲۱	فاصله از آبراهه
۲/۹۵	۲/۹۷	سنگ‌شناسی
۲/۱۹	۲/۱۹	شیب زمین
۲/۱۵	۲/۱۷	فراوانی بارش‌های سنگین
۲/۰۲	۲/۰۴	پوشش گیاهی
۱/۹۹	۲/۰۱	انحنای زمین
۱/۵۴	۱/۸۷	شاخص موقعیت شیب نسبی
۱/۸۵	۱/۸۷	شاخص رطوبت توپوگرافی
۱/۶۴	۱/۷۱	یافت خاک
۱/۱۹	۱/۱۹	جهت شیب
۱/۱۹	۱/۱۹	بارش سالانه

انتخاب پارامترهای بهینه

مقادیر بهینه پارامترها به بهره‌گیری از آرگومان tuneLength از طریق کتابخانه Caret تعیین شدند. مقادیر بهینه هر یک پارامترهای مدل‌ها در جدول ۶ ارائه شده است. به عنوان مثال، برای مدل جنگل تصادفی ابتدا با افزایش مقادیر پارامتر mtry (تعداد متغیرهایی که به صورت تصادفی به عنوان کاندید در هر تقسیم، نمونه‌گیری می‌شوند)، میزان خطا کاهش می‌یابد ولی بعد از مقدار ۳ به میزان خطای مدل افزوده شده است. بنابراین مقدار بهینه این پارامتر برابر با ۳ می‌باشد (شکل ۴).

جدول ۶. مقادیر بهینه پارامترهای مدل‌ها

مقدار بهینه	پارامتر	مدل
۰/۱	نرخ یادگیری	
۳	تعداد لایه پنهان	شبکه عصبی
۰/۸۵	سیگما (Sigma)	
۱۴/۴۳	هزینه (C)	ماشین بردار پشتیبان
۳	تعداد متغیرهای تصادفی در هر تقسیم (mtry)	جنگل تصادفی



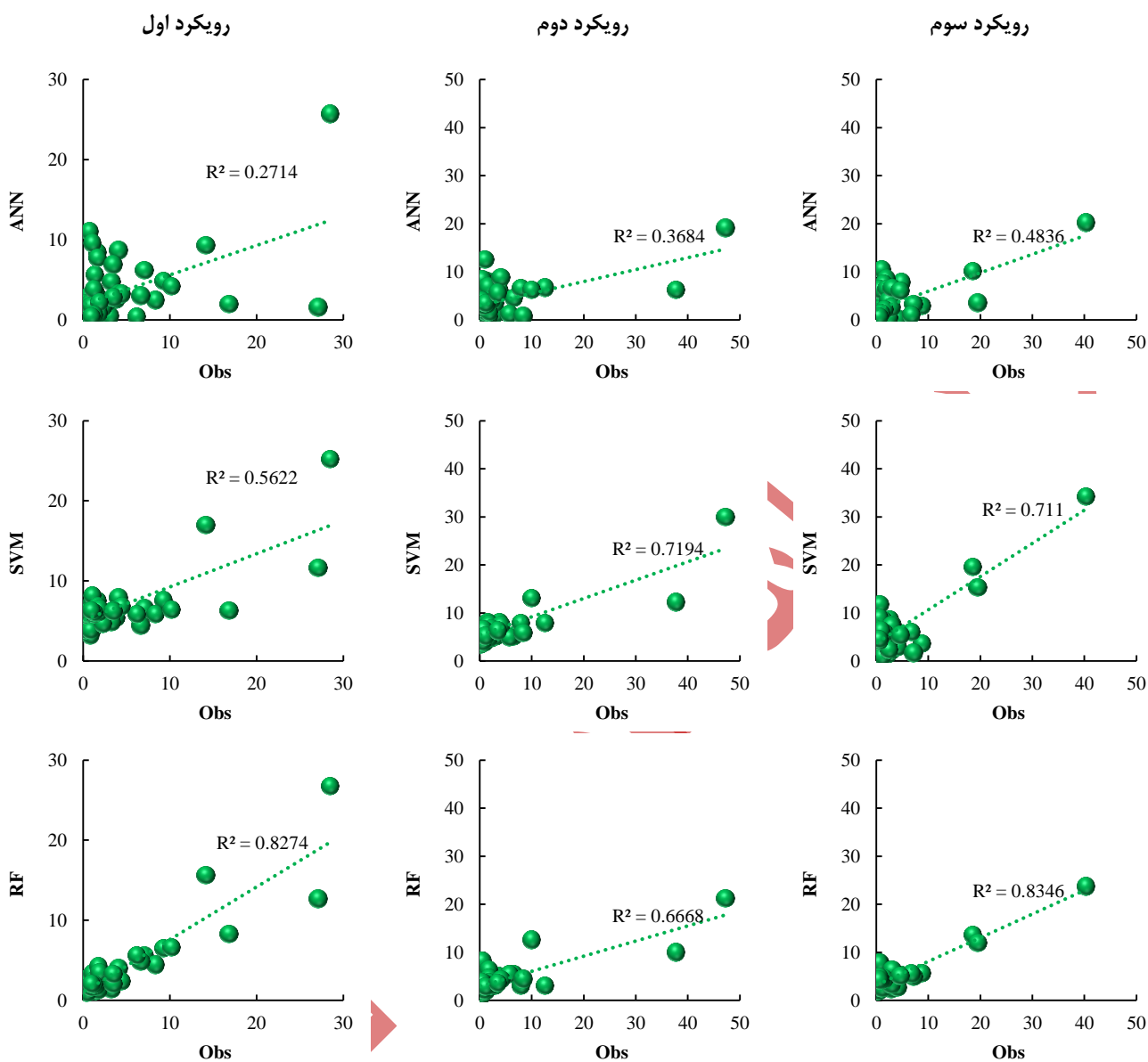
شکل ۴. تغییرات خطای مدل جنگل تصادفی در برابر مقادیر مختلف پارامتر mtry

ارزیابی دقت پیش‌بینی مدل‌ها

نتایج ارزیابی دقت پیش‌بینی مدل‌ها با استفاده از داده‌های اعتبار‌سنجی در جدول ۷ ارائه شده است. تشریح نتایج اعتبار‌سنجی مدل‌ها به شرح زیر می‌باشد:

- بررسی مقادیر RMSE حاکی از آن است که مدل جنگل تصادفی بهترین عملکرد را ارائه کرده است. بطوریکه در رویکرد تقسیم داده اول مقدار ۲/۵۷ را بخود اختصاص داده است که کمتر از سایر مدل‌ها است. همچنین این مدل در گروه‌های دوم و سوم نیز کمترین مقدار RMSE را نسبت به سایر مدل‌ها داشته است (به ترتیب برابر با ۵/۶۸ و ۳/۷۷). بعد از مدل RF، مدل SVM از نظر میزان RMSE قرار دارد و مدل ANN بیشترین مقادیر خطا را در تمامی سناریوهای تفکیک داده‌ای بخود اختصاص داده است.
- از نظر مقادیر R^2 ، مدل جنگل تصادفی در رویکرد تقسیم داده اول و سوم بیشترین مقدار را بخود اختصاص داده‌اند (۸۳٪)، اما در رویکرد تقسیم داده دوم میزان R^2 مدل SVM بیشتر از بقیه می‌باشد. مدل ANN بمانند مقادیر RMSE، R^2 کمتری را نسبت به سایر مدل‌ها دارا می‌باشد.
- از نظر معیار ارزیابی شاخص خطای RSR، مدل جنگل تصادفی دارای کمترین مقدار این شاخص بوده و بالاترین دقت را به خود اختصاص داده است. مدل ماشین بردار پشتیبان نیز از نظر شاخص خطای RSR، در جایگاه دوم و مدل شبکه عصبی در جایگاه سوم قرار گرفته است.
- از نظر شاخص ارزیابی تطابق (d)، مدل جنگل تصادفی در گروه تقسیم اول دارای بیشترین تطابق میان داده‌های مشاهده‌ای و پیش‌بینی بوده (۹۳٪) و بیشترین مقدار این شاخص را به خود اختصاص داده است. اما در رویکردهای تقسیم دوم و سوم مدل ماشین بردار پشتیبان به ترتیب با مقدار شاخص تطابق ۰/۷۵ و ۰/۸۶ بهتر از دو مدل دیگر می‌باشد. مقادیر شاخص تطابق در مدل شبکه عصبی مصنوعی کمتر از بقیه مدل‌ها می‌باشد و این مدل از نظر این معیار نیز در رتبه سوم قرار می‌گیرد.
- با در نظر گرفتن معیارهای ارزیابی دقت مدل‌ها می‌توان نتیجه گرفت که مدل RF عملکرد خوب، مدل SVM عملکرد متوسط و مدل ANN عملکرد ضعیفی را در برآورد میزان هدررفت خاک ناشی از فرسایش خندقی داشته‌اند (جدول ۷).

نمودارهای پراکنش داده‌های مشاهده‌ای و برآوردی هدررفت خاک ناشی از فرسایش‌های خندقی در مرحله اعتبار‌سنجی مدل‌ها با رویکردهای تقسیم داده‌ای مختلف در شکل ۵ ارائه شده است. این نمودارهای پراکنش نیز حاکی از عملکرد بهتر مدل جنگل تصادفی می‌باشد (شکل ۵).



شکل ۵. مقادیر مشاهداتی (Obs) هدررفت خاک ناشی از فرسایش خندقی در برابر مقادیر مدل‌سازی شده (داده‌های اعتبارسنجی)

جدول ۷. نتایج ارزیابی دقت مدل‌ها در مدل‌سازی میزان هدررفت خاک ناشی از فرسایش خندقی

مدل مورد استفاده			معیار ارزیابی	رویکرد تقسیم داده
شبکه عصبی مصنوعی (ANN)	ماشین بردار پشتیبان (SVM)	جنگل تصادفی (RF)		
۴/۹۳	۴/۹۰	۲/۵۷	اول	RMSE
۶/۴۴	۶/۱۶	۵/۶۸	دوم	
۴/۴۸	۴/۳۲	۳/۷۷	سوم	
۰/۲۷	۰/۵۶	۰/۸۲	اول	R ²
۰/۳۷	۰/۷۲	۰/۶۷	دوم	

سوم	۰/۸۳	۰/۷۱	۰/۴۸
اول	۰/۴۵	۰/۸۶	۰/۸۷
دوم	۰/۷۲	۰/۷۸	۰/۸۱
سوم	۰/۶۱	۰/۷۰	۰/۷۳
اول	۰/۹۳	۰/۷۴	۰/۷۰
دوم	۰/۷۱	۰/۷۵	۰/۶۲
سوم	۰/۸۵	۰/۸۶	۰/۷۵

بحث

نتایج بررسی همخطی چندگانه نشان داد که دو فاکتور فاصله از جاده و ارتفاع از سطح دریا هم‌سو بوده و وجود آنها در مدل‌سازی باعث می‌شود که ضرایب مدل‌سازی معتبر نباشد، زیرا تأثیر هر متغیر توصیفی بر «متغیر پاسخ» شامل تأثیر سایر متغیرهای مدل نیز می‌شود. بنابراین، واریانس برآوردگرهای ضرایب مدل‌ها افزایش یافته و در عمل، پیش‌بینی توسط مدل‌ها با خطای بزرگی همراه خواهد بود (Shrestha, 2020).

مدل‌های یادگیری ماشین برای مدل‌سازی حساسیت و استعداد شکل‌گیری فرسایش خندقی در گذشته مورد استفاده قرار گرفته‌اند (Garosi et al., 2018; Gayen et al., 2019)، اما کارایی آن‌ها در زمینه مدل‌سازی رشد فرسایش خندقی که ماهیت زمانی-مکانی دارد، تاکنون مشخص نشده بود. در این مطالعه، نتایج مدل‌سازی میزان هدررفت خاک ناشی از فرسایش خندقی نشان داد که مدل جنگل تصادفی عملکرد بهتری از مدل ماشین بردار پشتیبان داشته است. همچنین مدل شبکه عصبی عملکرد ضعیفی را بخود اختصاص داده است. با توجه به اینکه مدل‌سازی هدررفت خاک ناشی از فرسایش خندقی با استفاده از مدل‌های یادگیری ماشین تا به کنون انجام نشده است، امکان مقایسه مستقیم وجود ندارد. با این حال در مطالعات پیشین (Gayen et al., 2019) عملکرد بهتر مدل جنگل تصادفی نسبت به ماشین بردار پشتیبان در شناسایی مناطق مستعد فرسایش خندقی تایید شده است. برتری مدل جنگل تصادفی نسبت به مدل‌های ماشین بردار پشتیبان و شبکه عصبی مصنوعی در مدل‌سازی و شبیه‌سازی را می‌توان به عوامل متعددی نسبت داد که در مقالات تحقیقاتی مختلف مورد بحث قرار گرفته‌اند. به گفته Genuer et al. (2010)، مدل جنگل تصادفی به دلیل توانایی در مدیریت داده‌ها با ابعاد بالا و مقاومت در برابر بیش‌برازش، عملکرد بهتری را نشان می‌دهد. علاوه بر این، این مدل می‌تواند به طور موثر روابط غیرخطی بین متغیرها را مدیریت کنند. از سوی دیگر، مدل‌های SVM و ANN با داده‌های با ابعاد زیاد مشکل داشته و همچنین ممکن است دچار بیش‌برازش شوند. مدل SVM از نظر محاسباتی می‌تواند دشوار باشند، در حالی که ANN ممکن است از مشکل گرادیان ناپدید شدن در طول آموزش رنج ببرند (Hornik et al., 1989). به طور کلی، مدل جنگل تصادفی رویکردی کارآمدتر و مؤثرتر برای مدل‌سازی و شبیه‌سازی ارائه می‌کند، زیرا می‌تواند روابط پیچیده را مدیریت کند، بیش‌برازش را کاهش دهد، متغیرهای کم‌اهمیت‌تر را حذف کرده و در سناریوهای مختلف عملکرد خوبی را داشته باشد.

در نظر گرفتن فاکتورهایی که در زمان تغییرپذیر می‌باشند شرایط پیش‌بینی در آینده را فراهم می‌کند و در سال‌های آتی نیازی به اندازه‌گیری مجدد صحرائی نیست. به این صورت که در سال‌های بعد، نیاز می‌باشد که تنها مقدار بارندگی سالانه، تعداد بارندگی‌های بیشتر از ۵ میلی‌متر و نقشه پوشش گیاهی را برای پیش‌بینی هدررفت خاک خندق‌ها تهیه شود. عوامل مربوط به بارندگی از ایستگاه‌های هواشناسی و عامل پوشش گیاهی از تصاویر ماهواره‌ای لندست و یا سنتینل تهیه می‌شوند. سایر متغیرهای مستقل نیازی به اندازه‌گیری مجدد بعد از اجرای پروژه ندارند زیرا در گذر زمان تغییرات زیادی ندارند (مانند ارتفاع، شیب و غیره). همچنین، میزان هدررفت خاک نیز نیازی به اندازه‌گیری مجدد صحرائی ندارد. در حقیقت، مدل‌های آموزش دیده شده، در سال‌های بعد می‌تواند بر اساس یادگیری‌های کسب شده در مرحله آموزش، با در نظر گرفتن اطلاعات بارندگی و پوشش گیاهی جدید، میزان هدررفت خاک ناشی از خندق‌ها را پیش‌بینی نماید. این

مسأله مهمترین وجه تمایز این پروژه‌ی تحقیقاتی محسوب می‌شود و نشان می‌دهد که مدل‌سازی می‌تواند در صرفه‌جویی وقت و هزینه، خدمات ارزنده‌ای به مدیریت حفاظت آب و خاک کشور ارائه دهد.

محدودیت‌های اصلی این پژوهش به شرح زیر می‌باشد: (۱) کم بودن تعداد سال‌های نمونه‌برداری یکی از اصلی‌ترین محدودیت‌های این پژوهش است. چرا که اندازه‌گیری‌ها در درزآمدت می‌تواند گویای بهتر برآورد هدررفت متوسط خاک ناشی از فرسایش خندقی باشد. اندازه‌گیری‌های کوتاه‌مدت تحت تأثیر دوره‌های خشکسالی و یا ترسالی قرار گرفته و نمی‌تواند گویای هدررفت واقعی عرصه باشد که این امر نتایج مدل‌سازی را نیز تحت تأثیر قرار می‌دهد. همچنین مدل‌سازی با داده‌های طولانی مدت می‌تواند کمک به ایجاد مدل‌های قابل اعتمادتری کند. (۲) نبود داده در مقیاس مناسب از جمله محدودیت‌های دیگر این پژوهش بود. به‌عنوان مثال ایستگاه هواشناسی جهت ثبت داده‌های بارش به اندازه کافی در حوضه مورد مطالعه وجود نداشت. بنابراین تغییرات بارندگی برای تک‌تک خندق‌ها می‌توانست موجب بهبود مدل‌سازی شود.

نتیجه‌گیری و پیشنهادها

نتایج این پژوهش نشان داد که می‌توان خندق‌های منطقه مورد مطالعه را در سه دسته همگن قرار داد. اندازه‌گیری میدانی از تعداد ۵۸ خندق منتخب نشان داد که مقدار هدررفت خالص سالانه خاک ناشی از این خندق‌ها به ترتیب برابر با ۱۷۲، ۱۹۶ و ۲۰۸ تن در طی سال‌های ۱۴۰۰، ۱۴۰۱ و ۱۴۰۲ بوده است. نتایج مدل‌سازی هدررفت خاک ناشی از فرسایش خندقی حاکی از آن بود که مدل جنگل تصادفی بهترین مدل بود، در صورتی که مدل‌های ماشین بردار پشتیبان و شبکه عصبی به ترتیب عملکرد متوسط و ضعیفی را ارائه کردند. بر اساس نتایج حاصل شده و محدودیت‌های پژوهش، پیشنهادات پژوهشی و مدیریتی زیر ارائه می‌شود:

- کم بودن تعداد سال‌های نمونه‌برداری یکی از اصلی‌ترین محدودیت‌های این پژوهش است. چرا که اندازه‌گیری‌ها در درزآمدت گویای بهتر برآورد هدررفت متوسط خاک ناشی از فرسایش خندقی می‌باشند. از طرف دیگر، اندازه‌گیری‌های کوتاه‌مدت تحت تأثیر خشکسالی‌ها و یا ترسالی‌ها قرار می‌گیرند و همچنین نتایج مدل‌سازی را تحت تأثیر قرار می‌دهد. بر این اساس پیشنهاد می‌شود اندازه‌گیری‌ها در دوره طولانی‌مدت‌تر انجام شود تا نتایج قابل اعتمادتری حاصل شود.
- با توجه به اینکه بارش یکی از فاکتورهای اصلی در گسترش فرسایش خندقی می‌باشد، کم بودن ایستگاه‌های هواشناسی در سطح عرصه از جمله محدودیت‌های دیگر تحقیق می‌باشد. جهت رفع این محدودیت، استقرار باران‌سنج‌های زیاد در سطح عرصه و یا انتخاب منطقه‌ای وسیع که چندین ایستگاه هواشناسی را شامل شود توصیه می‌شود.
- نتایج مدل‌سازی نشان داد که مدل شبکه عصبی عملکرد ضعیفی را داشته است. با توجه به پیشرفت مدل‌های شبکه عصبی و ظهور مدل‌های یادگیری عمیق، در صورتیکه داده‌های اندازه‌گیری شده زیاد باشد امکان استفاده از شبکه‌های عصبی عمیق می‌تواند نتایج بهتری را حاصل کند.
- برخی از خندق‌های موجود در این منطقه، خطوط جاده‌ای، انتقال برق و گاز را بطور جدی تهدید می‌کنند. پیشنهاد می‌شود حداقل در این مناطق بحرانی عملیات اصلاحی ترکیبی شامل جلوگیری از ورود آب به خندق‌ها و کاشت پوشش گیاهی بومی در محدوده‌های خندقی جهت تثبیت انجام شود.

سپاس‌گذاری

این پژوهش مستخرج از نتایج پروژه تحقیقاتی با کد مصوب «۵۱۴-۹۹۰۲۴-۲۹-۵۰-۰» در سازمان تحقیقات، آموزش و ترویج کشاورزی می‌باشد. از حمایت‌های مادی و معنوی این سازمان، پژوهش‌کننده حفاظت خاک و آب‌خیزداری و مرکز تحقیقات و آموزش

"هیچگونه تعارض منافع بین نویسندگان وجود ندارد"

منابع

گرنامی، رضا و شادفر، صمد (۱۳۹۷). کاربرد GIS در تعیین مناطق مستعد فرسایش خندقی با روش فرایند تحلیل شبکه. پژوهش‌های آبخیزداری، ۴(۳۱)، ۵۸-۶۸.

REFERENCES

- Ait Naceur, H., Abdo, H. G., Igmoullan, B., Namous, M., Alshehri, F., & Albanai, J. (2024). Implementation of random forest, adaptive boosting, and gradient boosting decision trees algorithms for gully erosion susceptibility mapping using remote sensing and GIS. *Environmental Earth Sciences*, 83(3), 121. <https://doi.org/10.1007/s12665-024-11424-5>.
- Avand, M., Janizadeh, S., Naghibi, S. A., Pourghasemi, H. R., Khosrobeigi Bozchaloei, S., & Blaschke, T. (2019). A comparative assessment of random forest and k-nearest neighbor classifiers for gully erosion susceptibility mapping. *Water*, 11(10), 2076. <https://doi.org/10.3390/w11102076>.
- Casalí, J., Loizu, J., Campo, M. A., De Santisteban, L. M., & Álvarez-Mozos, J. (2006). Accuracy of methods for field assessment of rill and ephemeral gully erosion. *Catena*, 67(2), 128-138. <https://doi.org/10.1016/j.catena.2006.03.005>.
- Chuma, G. B., Mugumaarahama, Y., Mond, J. M., Bagula, E. M., Ndeko, A. B., Lucungu, P. B., ... & Schmitz, S. (2023). Gully erosion susceptibility mapping using four machine learning methods in Luzinzi watershed, eastern Democratic Republic of Congo. *Physics and Chemistry of the Earth, Parts A/B/C*, 129, 103295. <https://doi.org/10.1016/j.pce.2022.103295>.
- Cochran, W.G., 1977. *Sampling techniques*. 3rd edn. (New York: John Wiley & Sons).
- Dong, Y., Wu, Y., Qin, W., Guo, Q., Yin, Z., & Duan, X. (2019). The gully erosion rates in the black soil region of northeastern China: Induced by different processes and indicated by different indexes. *Catena*, 182, 104146. <https://doi.org/10.1016/j.catena.2019.104146>.
- Dube, F., Nhapi, I., Murwira, A., Gumindoga, W., Goldin, J., & Mashauri, D. A. (2014). Potential of weight of evidence modelling for gully erosion hazard assessment in Mbire District-Zimbabwe. *Physics and chemistry of the Earth, Parts A/B/C*, 67, 145-152. <https://doi.org/10.1016/j.pce.2014.02.002>.
- Garosi, Y., Sheklabadi, M., Pourghasemi, H. R., Besalatpour, A. A., Conoscenti, C., & Van Oost, K. (2018). Comparison of differences in resolution and sources of controlling factors for gully erosion susceptibility mapping. *Geoderma*, 330, 65-78. <https://doi.org/10.1016/j.geoderma.2018.05.027>.
- Gayen, A., Pourghasemi, H. R., Saha, S., Keesstra, S., & Bai, S. (2019). Gully erosion susceptibility assessment and management of hazard-prone areas in India using different machine learning algorithms. *Science of the total environment*, 668, 124-138. <https://doi.org/10.1016/j.scitotenv.2019.02.436>.
- Genuer, R., Poggi, J. M., & Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern recognition letters*, 31(14), 2225-2236. <https://doi.org/10.1016/j.patrec.2010.03.014>.
- Gornami, R., & Shadfar, S. (2018). Application of the GIS in the Determination of Susceptible Areas to Gully Erosion Using the Analytic Network Process (ANP). *Watershed Management Research Journal*, 31(4), 58-68. <https://doi.org/10.22092/wmej.2018.121633.1112>. (inPersian)
- Hasanuzzaman, M., Adhikary, P. P., & Shit, P. K. (2024). Gully erosion susceptibility mapping and prioritization of gully-dominant sub-watersheds using machine learning algorithms: Evidence from the Silabati River (tropical river, India). *Advances in Space Research*, 73(3), 1653-1666. <https://doi.org/10.1016/j.asr.2023.10.051>.
- Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5), 359-366. [https://doi.org/10.1016/0893-6080\(89\)90020-8](https://doi.org/10.1016/0893-6080(89)90020-8).
- Jie, C., Jing-Zhang, C., Man-Zhi, T., & Zi-tong, G. (2002). Soil degradation: a global problem endangering sustainable development. *Journal of Geographical Sciences*, 12, 243-252. <https://doi.org/10.1007/BF02837480>.
- Kheradmand, H. R., & Soleimanpour, S. M. (2016). Longitudinal progress prediction of gully erosion with FAO model (Case study: a part of Abdan watershed, Bushehr province, South of Iran). *International Journal of Biology Pharmacy and Allied Sciences*, 5(2), 151-155.
- Kuhn, M. (2015). Caret: classification and regression training. *Astrophysics Source Code Library*, ascl-1505.

- Martins, B., Pinheiro, C., Nunes, A., Bento-Gonçalves, A., & Hermenegildo, C. (2024). Geo-environmental factors controlling gully distribution at the local scale in a Mediterranean environment. *Catena*, 236, 107712. <https://doi.org/10.1016/j.catena.2023.107712>.
- Mfondoum, A. H. N., Nguet, P. W., Seuui, D. T., Mfondoum, J. V. M., Ngenyam, H. B., Diba, I., ... & Beni, L. M. (2023). Stepwise integration of analytical hierarchy process with machine learning algorithms for landslide, gully erosion and flash flood susceptibility mapping over the North-Moungo perimeter, Cameroon. *Geoenvironmental Disasters*, 10(1), 22. <https://doi.org/10.1186/s40677-023-00254-5>.
- Mohammadi, S., Balouei, F., Haji, K., Khaledi Darvishan, A., & Karydas, C. G. (2021). Country-scale spatio-temporal monitoring of soil erosion in Iran using the G2 model. *International Journal of Digital Earth*, 14(8), 1019-1039. <https://doi.org/10.1080/17538947.2021.1919230>.
- Mohebzadeh, H., Biswas, A., Rudra, R., & Daggupati, P. (2022). Machine learning techniques for gully erosion susceptibility mapping: a review. *Geosciences*, 12(12), 429. <https://doi.org/10.3390/geosciences12120429>.
- Mukai, S. (2017). Gully erosion rates and analysis of determining factors: a case study from the semi-arid main Ethiopian rift valley. *Land degradation & development*, 28(2), 602-615. <https://doi.org/10.1002/ldr.2532>.
- Nyssen, J., Poesen, J., Moeyersons, J., Haile, M., & Deckers, J. (2008). Dynamics of soil erosion rates and controlling factors in the Northern Ethiopian Highlands—towards a sediment budget. *Earth surface processes and landforms*, 33(5), 695-711. <https://doi.org/10.1002/esp.1569>.
- Pereira, P., Bogunovic, I., Muñoz-Rojas, M., & Brevik, E. C. (2018). Soil ecosystem services, sustainability, valuation and management. *Current Opinion in Environmental Science & Health*, 5, 7-13. <https://doi.org/10.1016/j.coesh.2017.12.003>.
- Poesen, J., Nachtergaele, J., Verstraeten, G., & Valentin, C. (2003). Gully erosion and environmental change: importance and research needs. *Catena*, 50(2-4), 91-133. [https://doi.org/10.1016/S0341-8162\(02\)00143-1](https://doi.org/10.1016/S0341-8162(02)00143-1).
- Rouhani, H., Fathabadi, A., & Baartman, J. (2021). A wrapper feature selection approach for efficient modelling of gully erosion susceptibility mapping. *Progress in Physical Geography: Earth and Environment*, 45(4), 580-599. <https://doi.org/10.1177/0309133320979897>.
- Shrestha, N. (2020). Detecting multicollinearity in regression analysis. *American Journal of Applied Mathematics and Statistics*, 8(2), 39-42. doi: 10.12691/ajams-8-2-1.
- Tebebu, T. Y., Abiy, A. Z., Zegeye, A. D., Dahlke, H. E., Easton, Z. M., Tilahun, S. A., ... & Steenhuis, T. S. (2010). Surface and subsurface flow effect on permanent gully formation and upland erosion near Lake Tana in the northern highlands of Ethiopia. *Hydrology and Earth System Sciences*, 14(11), 2207-2217.
- Wang, J., Zhang, Y., Deng, J., Yu, S., & Zhao, Y. (2021). Long-term gully erosion and its response to human intervention in the tableland region of the Chinese loess plateau. *Remote Sensing*, 13(24), 5053. <https://doi.org/10.3390/rs13245053>.
- Willmott, C. J. (1981). On the validation of models. *Physical geography*, 2(2), 184-194. <https://doi.org/10.1080/02723646.1981.10642213>.
- Yermolaev, O., Medvedeva, R., & Poesen, J. (2022). Spatial and temporal dynamics of gully erosion in anthropogenically modified forest and forest-steppe landscapes of the European part of Russia. *Earth Surface Processes and Landforms*, 47(12), 2926-2940. <https://doi.org/10.1002/esp.5433>.
- Zhang, X., Fan, J., Liu, Q., & Xiong, D. (2018). The contribution of gully erosion to total sediment production in a small watershed in Southwest China. *Physical Geography*, 39(3), 246-263. <https://doi.org/10.1080/02723646.2017.1356114>.

Modeling soil loss due to gully erosion in the data-scarce regions

EXTENDED ABSTRACT

Introduction

Gully erosion is a critical and detrimental form of land degradation and soil loss affecting various regions worldwide. Its significant impact necessitates comprehensive assessments and monitoring efforts to mitigate its effects. However, traditional field measurement techniques for gully erosion can be time-consuming and costly, posing challenges in accurately quantifying soil loss from this process. To overcome these limitations, this study aimed to estimate soil loss resulting from gully erosion through the application of machine learning models. The research was conducted in the Choopanlu watershed, situated in West Azarbaijan province, Iran. By leveraging the capabilities of machine learning, the study sought to establish a more efficient and reliable method for assessing soil loss caused by gully erosion. This, in turn, would aid in formulating effective strategies to manage and control land degradation in the region.

Material and methods

The study commenced with field monitoring to identify gullies in the area to achieve this objective. Following this, digital layers of factors influencing gully expansion were prepared to facilitate gully clustering and selection. These factors included topographical characteristics (elevation, slope, aspect, surface curvature, and relative slope position index), vegetation, land use, soil, lithology, and hydroclimate indicators (distance from stream, drainage density, topographic wetness index, annual precipitation, and frequency of heavy rainfall events). Subsequently, the volume of soil lost due to gully erosion during the three-year period (2021-2023) was measured as the dependent variable for the selected gullies through field observations. In this study, three machine learning models were employed for modeling: Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Network (ANN). The cross-validation approach was utilized for modeling process.

Results and discussion

Cochran's formula results indicated that among the 67 identified gullies in the field, a minimum sample size of 58 gullies was required. Following clustering, this number of selected gullies was chosen from the three identified clusters. The annual soil erosion caused by the selected gullies was estimated to be 172 tons in 2021, 196 tons in 2022, and 208 tons in 2023. The results of the multicollinearity analysis indicated that the elevation exhibited collinearity and thus was excluded from the modeling process. Based on the RMSE values and the coefficient of determination, it can be inferred that the RF model demonstrated the best performance, followed by the SVM model with moderate performance, and the ANN model exhibiting the poorest performance in modeling soil loss due to gully erosion. Consequently, the Random Forest model emerged as the most suitable model among the tested methods.

Conclusion

In conclusion, this study successfully applied machine learning models to estimate soil loss from gully erosion in the Choopanlu watershed. The analysis of 67 identified gullies, with a minimum sample size of 58, revealed an increasing trend in annual soil erosion over the three-year period. Multicollinearity analysis led to the exclusion of elevation, and subsequent model comparison indicated that the RF model outperformed the SVM and ANN models in predicting soil loss due to gully erosion. These findings demonstrate the potential of utilizing machine learning, particularly the RF model, for efficient and accurate assessment of soil loss from gully erosion in similar contexts. Further research could explore the application of these models in other regions and investigate additional factors influencing gully erosion processes.

Keywords: Choopanlu watershed, Gully erosion, Machine learning, Modeling, Soil loss

Authors' contribution:

Choubin B: Conceptualization, data curation, methodology, software, writing - original draft preparation.

Rahmati O: Conceptualization, methodology, writing—review and editing.

Soleimanpour SM: Conceptualization, supervision, validation.

Shadfar S: Conceptualization, project administration.

Najafi Eigdir A: Data curation, writing - original draft preparation.

Data Availability Statement:

The datasets are available upon a reasonable request to the corresponding author.

Ethical considerations:

The study was approved by the Ethics Committee of the Agricultural Research, Education and Extension Organization (AREEO). The authors avoided from data fabrication and falsification.

Funding:

The study was funded by the Agricultural Research, Education and Extension Organization (AREEO), Iran, and Grant No. 0-50-29-024-990514.

Conflict of interest:

The authors declare no conflict of interest.

فیلد استنادی
فیلد استنادی