

مقدمه

سیل یکی از فاجعه‌آمیزترین مخاطرات زیست‌محیطی است که صدمات شدیدی به جان‌ها و اموال وارد می‌کند و خساراتی به ارزش میلیاردها دلار در سطح جهان به دنبال دارد [۳۷، ۱۸]. عوامل طبیعی، به‌ویژه شدت بارندگی، که اخیراً به دلیل تأثیرات تغییرات آب و هوایی افزایش یافته است، علت اصلی خطرات سیل هستند. همچنین عوامل انسانی مانند رشد جمعیت، شهرنشینی، سیستم زهکشی ضعیف، کاربری نامناسب زمین، مدیریت ناکافی دشت سیلابی و مدیریت ضعیف مخازن آب تأثیرگذار بوده و خطر سیل را افزایش می‌دهد [۳۵، ۲۲]. سیل و اثرات اجتماعی-اقتصادی آن در سرتاسر جهان با سرعت نگران‌کننده‌ای در حال افزایش است. از آنجایی که سیل یکی از شایع‌ترین، چالش‌برانگیزترین و مخرب‌ترین خطر طبیعی است، شناسایی مناطق مستعد سیل برای استراتژی‌های جامع مدیریت سیل یک موضوع حیاتی است [۴۳، ۳۳]. این امر مستلزم آن است که مدیران و سیاست‌گذاران توجه زیادی به تحقیقات سیل داشته باشند تا مناطق مستعد سیل را برای توسعه آینده ترسیم کنند [۴۴، ۴۰]. بنابراین، تهیه نقشه مناطق مستعد سیل، یک اقدام غیرسازهای و گام بزرگی در مدیریت کلی خطر سیل برای کاهش اثرات سیل است. در چند سال گذشته، علاقه فزاینده‌ای به تحقیق در مورد تشخیص مناطق مستعد سیل و رابطه بین سیل و عوامل محیطی می‌توان مشاهده کرد. وقوع بلایای طبیعی در چند سال اخیر به دلیل عوامل مختلفی از جمله تخریب محیط زیست، تغییرات اقلیمی، رشد سریع جمعیت و تشدید و استفاده نامناسب از اراضی افزایش چشم‌گیری داشته است [۷۱]. قابل ذکر است که در بسیاری از مناطق کشور، وقوع سیل و خسارات ناشی از آن افزایش یافته است و شناسایی مناطق سیل‌خیز و تعیین پتانسیل سیل‌خیزی آن‌ها می‌تواند در بهبود مدیریت آبخیزها بسیار مفید باشد. به این منظور، روش‌های مختلفی برای تعیین میزان رواناب و پهنه‌بندی مناطق سیل‌خیز وجود دارد. بیش‌تر این روش‌ها بر اساس روش‌های نموداری و استفاده از فرمول‌های تجربی، تحلیل آماری داده‌های سیلاب، داده‌های ماهواره‌ای و اطلاعات جغرافیایی انجام می‌شود. این روش‌ها در واقع، به عنوان ابزارهای مهمی می‌توانند در دسترس باشند تا با دقت بیشتری مناطقی را که در معرض خطر سیل قرار دارند شناسایی و مدیریت مناسبی برای آن‌ها انجام شود که مطالعات زیادی با روش‌های مختلف انجام شده است که به چند مورد اشاره شده است. رویکردهای آماری چندمتغیره، هوش مصنوعی و یادگیری ماشین برای تهیه نقشه خطرهای طبیعی

تعیین مناطق مستعد سیل خیزی با استفاده از مدل‌های یادگیری ماشین در حوزه آبخیز شهرستانک شهرستان خوسف

جواد چزگی^۱ و سهیلا پویان^۲

تاریخ دریافت: ۱۴۰۲/۰۵/۱۷ تاریخ پذیرش: ۱۴۰۲/۰۷/۳۰

DOI: 10.22034/17.63.38

چکیده

سیل یکی از مخرب‌ترین پدیده‌های طبیعی است که در سراسر جهان رخ می‌دهد و منجر به آسیب به اموال و زیرساخت‌ها و یا حتی تلفات جانی می‌شود. این پژوهش باهدف تهیه نقشه پهنه‌بندی پتانسیل سیل‌خیزی در حوزه آبخیز شهرستانک شهرستان خوسف با استفاده از الگوریتم‌های یادگیری ماشین، انجام گرفته است. در این تحقیق برای تعیین مناطق دارای پتانسیل خطر سیل در منطقه مورد مطالعه از ۸ پارامتر تأثیرگذار و روش‌های جنگل تصادفی، مدل درخت رگرسیونی پیشرفته و ماشین بردار پشتیبان استفاده شده است. در این راستا ۸۱ مورد مخاطره منطقه سیل‌خیز ثبت شد. به‌طور تصادفی ۷۰ درصد برای فرآیند مدل‌سازی و ۳۰ درصد برای ارزیابی مدل‌های تهیه شده تقسیم‌بندی شد. نتایج اهمیت پارامترها نشان داد پارامتر تراکم زهکشی بیشترین تأثیر را در سیل‌خیزی داشته است. علاوه بر این در تعیین مناطق حساس به وقوع سیل، مدل جنگل تصادفی، درخت رگرسیون پیشرفته و ماشین بردار پشتیبان به ترتیب با منحنی تشخیص عملکرد نسبی (ROC) ۰/۹۹، ۰/۹۸ و ۰/۹۵ دارای دقت بالاتری هستند. منطقه مورد مطالعه از نظر حساسیت سیل‌خیزی به چهار طبقه (خیلی زیاد، زیاد، متوسط و کم) تقسیم شد. به‌صورت کلی، حوزه از نظر سیل‌خیزی متوسط به بالا است.

کلیدواژه‌ها: سیل‌خیزی، تراکم زهکشی، جنگل تصادفی، ماشین بردار پشتیبان، شهرستان خوسف

۱- استادیار گروه مرتع و آبخیزداری، دانشکده منابع طبیعی و محیط زیست، دانشگاه بیرجند، بیرجند، ایران. chezgi@brjand.ac.ir

۲- دستیار پژوهشی بخش منابع طبیعی، دانشکده کشاورزی، دانشگاه شیراز، شیراز، ایران

از جمله سیل مورد آزمایش قرار گرفته‌اند. از آن جمله می‌توان به مطالعات، بویی و همکاران [۶]. ارنر و همکاران [۲۳]. استفاده از روش‌های شبکه‌های عصبی مصنوعی و رگرسیون لجستیک کرنل، واخشوری و زارع [۵۸]. استفاده از منطق فازی، وانگ و همکاران [۶۰]. استفاده از درخت تصمیم، بویی و همکاران [۵]. استفاده از ماشین بردار پشتیبان، یوسف و همکاران [۶۳]. استفاده از جنگل تصادفی، چن و همکاران [۱۲]. استفاده از گروه‌های ANFIS اشاره نمود.

کاظمی و همکاران [۳۰]. مدل‌سازی و پیش‌بینی خطر سیل در شهر بندری نوشهر، را با استفاده از مدل‌های یادگیری ماشین انجام دادند. برای پهنه‌بندی خطر سیل از مدل‌های یادگیری ماشین، درخت تصمیم‌گیری، k نزدیک‌ترین همسایه، یادگیری تجمعی و WLC استفاده شد. برای انجام صحت‌سنجی مدل‌ها و انتخاب بهترین مدل نیز از روش ROC بر اساس دو معیار ارزیابی صحت و سطح زیر نمودار در خصوص بحران سیل استفاده گردید. نتایج نشان داد که مدل یادگیری تجمعی با صحت 0/96 و سطح زیر نمودار 0/90 بهترین و مدل WLC با صحت 0/61 و سطح زیر نمودار 0/62 ضعیف‌ترین مدل پیش‌بینی خطر سیل در شهر نوشهر است.

وانگ و همکاران [۶۱]. به تهیه نقشه حساسیت سیل با استفاده از الگوریتم شبکه عصبی حلقوی پرداختند. به این منظور، ۱۳ عامل محرک سیلاب مربوط به حوادث سیل تاریخی در منطقه شهر شانگیو تهیه شد. عملکرد روش‌های مبتنی بر CNN با استفاده از چندین معیار هدف با روش ماشین بردار پشتیبان (SVM) ارزیابی شد. نتایج نشان داد که کلیه روش‌های مبتنی بر CNN می‌توانند نقشه حساسیت سیل دقیق‌تری را تولید نمایند. بنابراین، الگوریتم پیشنهادی CNN می‌تواند به کاهش و مدیریت سیل کمک کند.

Chen چن و همکاران [۲۰]. بیان کردند که ادغام مدل‌های یادگیری ماشین برای ارزیابی خطر سیل و تعیین مکانیسم بالقوه بین ریسک و عوامل محرک برای بهبود مدیریت سیل بسیار مهم است. در این پژوهش، از شش مدل یادگیری ماشین برای ارزیابی خطر سیل استفاده شد. پس از تعیین مدل بهینه، عوامل محرک سطوح مختلف خطر سیل را بررسی نمودند. نتایج نشان داد که ارتفاع، بارش حداکثر یک‌روزه و تراکم جاده سه عامل مهم محرکه هستند و ۵۲ درصد در خطر سیل منطقه نقش مؤثری داشتند. این مطالعه نه تنها کاربرد یادگیری ماشین و روش‌های یادگیری عمیق برای ارزیابی خطر سیل استفاد شد، بلکه درک افراد از مکانیسم بالقوه خطر سیل عمیق‌تر و بینش‌هایی جدید در راستای مدیریت بهتر خطر سیل ارائه شد.

در سراسر جهان سیل به عنوان پدیده‌ای مخرب از لحاظ اقتصادی و حتی اجتماعی روند افزایشی پیدا کرده است. در واقع خطر سیل بیشترین، چالش برانگیزترین و خسارت‌آورترین خطر طبیعی هستند، شناسایی مناطق مستعد سیل به استراتژی‌های جامع مدیریت سیلاب یک مسئله مهم است [۳، ۲، ۳۳، ۳۴، ۴۳]. بنابراین، تهیه نقشه مناطق مستعد

سیل، گامی اساسی در مدیریت کلی خطر سیل برای کاهش تأثیرات سیل است [۵۳]. پژوهش حاضر در نظر دارد مناطق مستعد سیل را با استفاده از مدل‌های جنگل تصادفی و ماشین بردار پشتیبان در محیط سامانه اطلاعات جغرافیایی در منطقه شهرستانک خوسف پهنه‌بندی کند. در منطقه مذکور پژوهش‌های متفاوتی در مورد خطر سیل انجام گرفته است، اما تاکنون از روش‌های یادگیری ماشین برای مشخص کردن مناطق مستعد سیل استفاده نشده است که در این پژوهش به آن پرداخته شده است.

منطقه مورد مطالعه

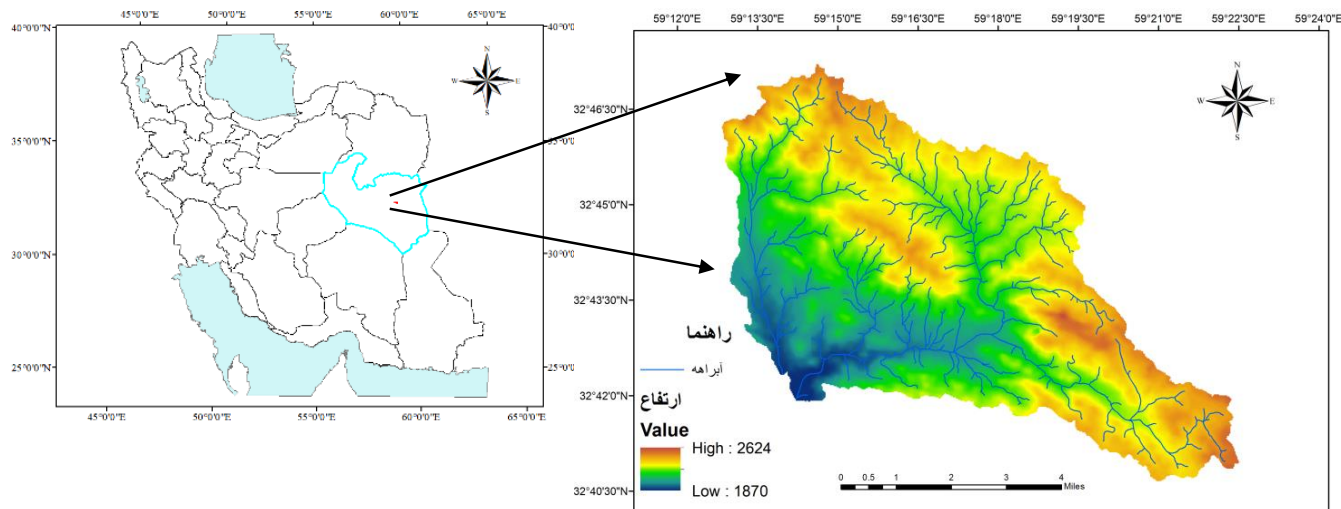
منطقه مورد مطالعه در قسمتی از حوضه آبریز دشت کویر لوت در جنوب شهرستان بیرجند واقع شده است که در طول جغرافیایی ۵۹ درجه و ۱۲ دقیقه و ۴۷ ثانیه تا ۵۹ درجه و ۲۲ دقیقه و ۲۰ ثانیه و عرض جغرافیایی ۳۲ درجه و ۴۰ دقیقه و ۹ ثانیه تا ۳۲ درجه و ۴۷ دقیقه و ۹ ثانیه قرار دارد، وسعت کل دو حوضه ۹۲۳۳ هکتار است که از این مقدار ۶۹۵۶ هکتار سهم حوضه شهرستانک و ۲۲۷۷ هکتار سهم حوضه عیلمی است. نوع اقلیم حوضه به روش آمبرژه از نوع خشک سرد تعیین می‌شود [۵۴]. در این حوضه رودخانه دائمی وجود ندارد و رودخانه اصلی آن هم فصلی است که درواقع سیلاب‌های اتفاق افتاده در حوضه را تخلیه می‌کند. منطقه مورد مطالعه در ۲۳ کیلومتری جنوب شرقی شهر بیرجند و در فاصله ۴۴ کیلومتری جنوب شرقی شهرستان خوسف و در فاصله ۹ کیلومتری مرکز دهستان به نام روستای گل واقع است. موقعیت جغرافیایی محدوده مورد مطالعه، در شکل (۱) نشان داده شده است.

روش تحقیق

در این تحقیق ابتدا عوامل مؤثر بر سیل‌خیزی منطقه براساس نظرات و مطالعات موردی در منطقه مورد استفاده شد. در ادامه داده‌های مورد نظر از سازمان‌های مربوطه، برداشت‌های صحرایی و غیره تهیه شد. در مرحله بعد نقشه‌های همه پارامترها تهیه شده و با استفاده از روش‌های جنگل تصادفی، مدل درخت رگرسیونی پیشرفته و ماشین بردار پشتیبان به نقشه سیل‌خیزی منطقه در نرم افزارهای R و WEKA بدست آمد.

تهیه نقشه پراکنش خطر سیل رخ داده در منطقه مورد مطالعه

نقشه رخداد سیل با پیمایش میدانی گسترده در سطح منطقه مورد مطالعه و با استفاده از GPS و متعاقباً تصاویر Google Earth و ماهواره سنتینل ۲ برداشت شد که در این راستا ۸۱ مورد منطقه سیل‌خیز ثبت شد. به منظور مدل‌سازی هر یک از خطرات سه‌گانه شناسایی شده، به‌طور تصادفی ۷۰ درصد برای فرآیند مدل‌سازی و ۳۰ درصد برای ارزیابی مدل‌های تهیه شده تقسیم‌بندی شد [۴۵، ۶۳].



شکل ۱- موقعیت منطقه مورد مطالعه در ایران و استان خراسان جنوبی
Fig 1. Location of the study area in Iran and South Khorasan province

روش درخت رگرسیون پیشرفته (BRT^2)، جنگل تصادفی (RF^3) است.

روش ماشین بردار پشتیبان (SVM)

فن آوری نوآورانه ذخیره سازی و محبوبیت روزافزون اینترنت، حجم فزاینده ای از داده ها را تولید کرده است. در این حجم وسیع از داده ها، دانش بسیار ارزشمندی در دسترس است، اما پنهان است. SVM یک تکنیک طبقه بندی پیشرفته است که به طور کلی مدل های دقیقی را ارائه می کند، زیرا می تواند غیرخطی ها را در داده ها ثبت کند. با این حال، این نقطه قوت نیز نقطه ضعف اصلی آن است، زیرا مدل های غیرخطی تولید شده معمولاً به عنوان مدل های جعبه سیاه غیرقابل درک در نظر گرفته می شوند. SVM یک روش قدرتمند برای ایجاد طبقه بندی است. هدف آن ایجاد یک مرز تصمیم بین دو کلاس است که امکان پیش بینی برچسب ها از یک یا چند بردار ویژگی را فراهم می کند. هر یک از کلاس های این نزدیک ترین نقاط را بردارهای پشتیبان می نامند. الگوریتم SVM در ابتدا برای ساخت یک طبقه بندی کننده خطی در سال ۱۹۶۳ توسط واپنیک پیشنهاد شد [۵۹،۲۷]. به منظور تهیه نقشه خطر سیل با استفاده از الگوریتم SVM، اساساً چهار نوع تابع Kernel از قبیل خطی، چندجمله ای، سینوسی و تابع پایه شعاعی وجود دارد که در تحقیق حاضر از تابع پایه شعاعی به دلیل عملکرد بهتر آن در مقایسه با دیگر توابع استفاده شده است [۵۶،۴۷].

مدل درخت رگرسیونی پیشرفته^۴ (BRT)

مدل BRT ترکیبی از تکنیک های آماری و یادگیری ماشینی است. مدل BRT با تکنیک های مختلف مطابقت دارد و آن ها را برای بهبود

عوامل مؤثر بر وقوع هر یک از مخاطره های ثبت شده در منطقه

مورد مطالعه

پس از تعیین تهیه نقشه پراکنش خطر، اقدام بعدی شناسایی عوامل مؤثر بر وقوع هر خطر در منطقه مورد مطالعه است. بر اساس بازدیدهای صحرائی و مرور منابع انجام شده [۱۷،۴۵،۳۸] عوامل درجه شیب، جهت شیب، طبقات ارتفاعی (متر)، تراکم زهکشی (کیلومتر بر کیلومتر مربع)، واحدهای سنگ شناسی، کاربری اراضی، شاخص پوشش گیاهی و میانگین بارش سالانه (میلی متر) انتخاب شدند (شکل ۲). نقشه های عوامل شناسایی شده با استفاده از نقشه های پایه توپوگرافی، زمین شناسی، تصاویر ماهواره ای (لندست ۸) موجود تهیه شد.

الگوریتم ماشین بردار پشتیبان (SVM) در سال ۱۹۹۵ توسط واپنیک به عنوان روشی جهت دسته بندی ارائه شد [۴].

جنگل تصادفی (RF)، یک روش یادگیری ترکیبی برای دسته بندی، رگرسیون است، یکی از قوی ترین تکنیک های یادگیری ماشینی است که شامل مجموعه ای از درختان است که هر درخت توسط نمونه های بوت استرپ تولید می شود [۵۱].

با توجه به خصوصیات خطر سیل باید از یک روش کارآمد استفاده شود. مروری بر مطالعات گذشته نشان می دهد، الگوریتم های یادگیری ماشین در تهیه نقشه حساسیت کارایی مناسبی داشته است [۴۸] مدل های مختلف، نه تنها می توانند دقت در برخورد با مسائل پیچیده و عدم قطعیت ها را افزایش دهند، بلکه می توانند باعث ایجاد و توسعه تئوری ها و روش های نوین در مسائل مختلف شوند.

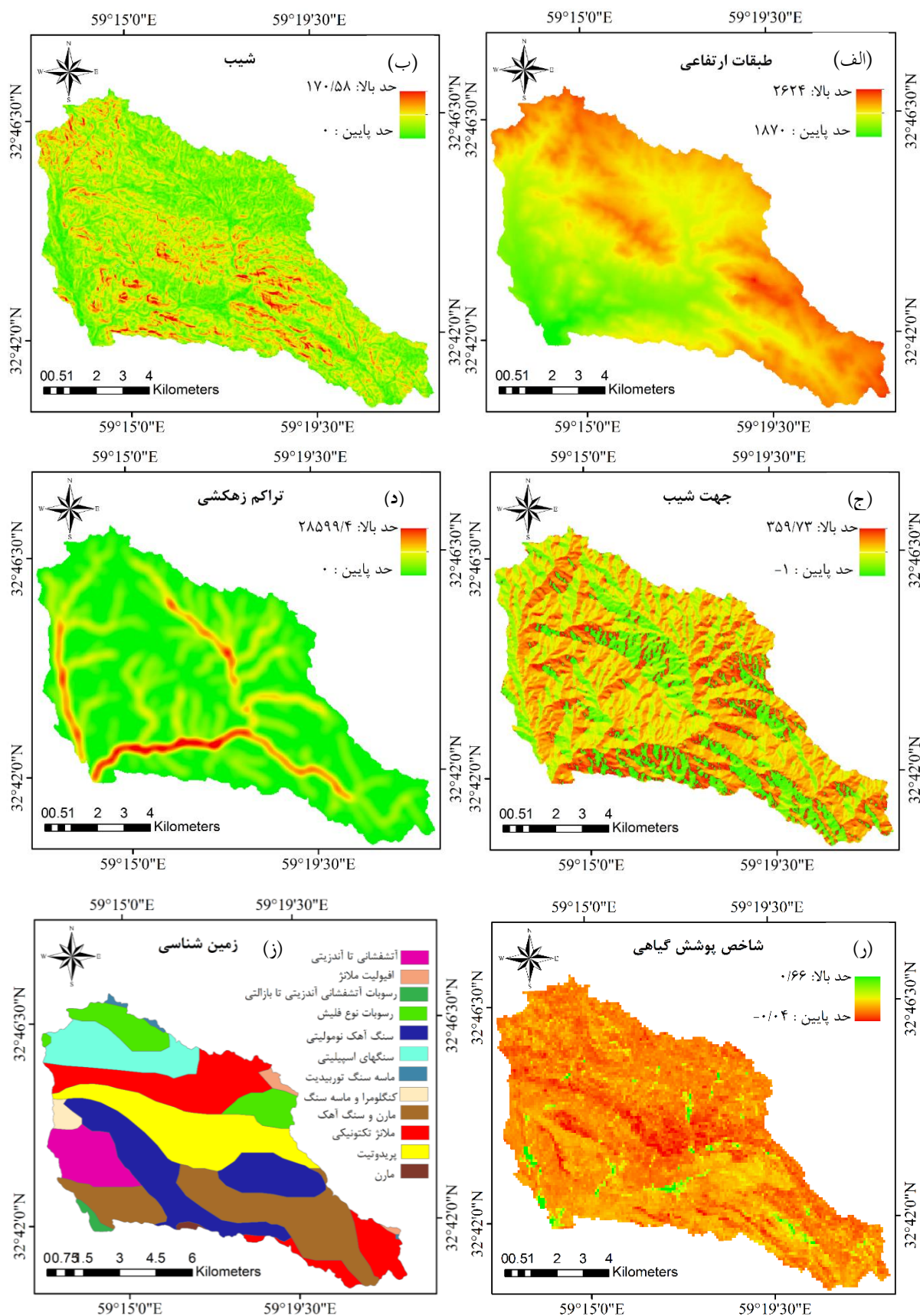
بنابراین برای مدل سازی مکانی خطر سیل، براساس وقوع سیل های مهم گذشته در منطقه و تجزیه و تحلیل عوامل مؤثر بر رخداد و پیدایش آن از تکنیک های یادگیری ماشینی استفاده می شود. تکنیک های یادگیری ماشینی شامل: ماشین بردار پشتیبان (SVM)^۱ و

2. Boosted Regression Tree (BRT)

3. Random Forest (RF)

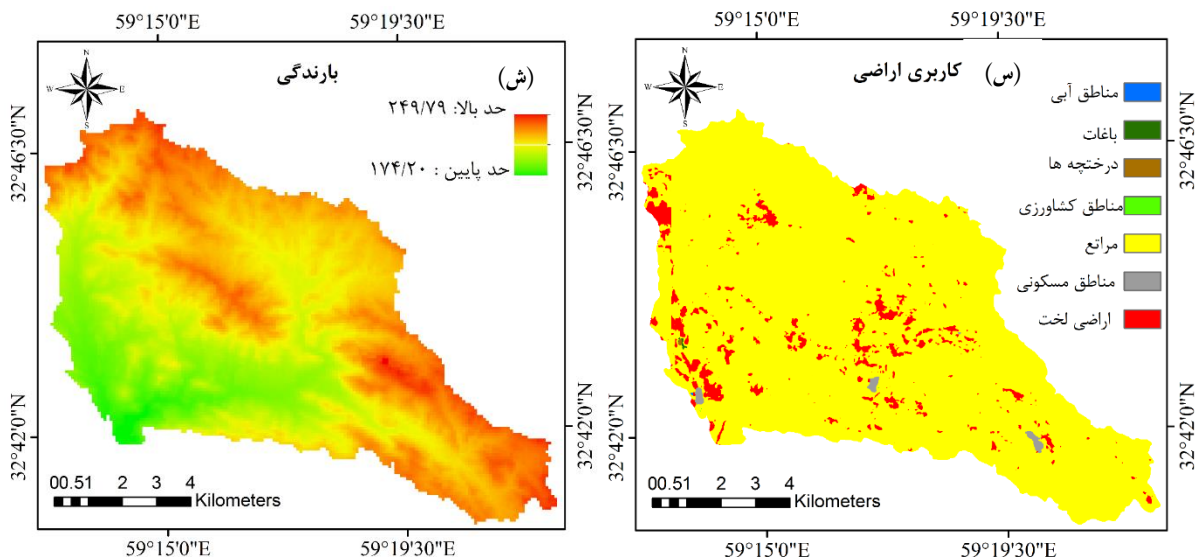
4. Boosted regression tree

1. Support Vector Machine (SVM)



شکل ۲- نقشه‌های پارامترهای: (الف) مدل رقومی ارتفاعی؛ (ب) درصد شیب؛ (ج) جهت شیب؛ (د) تراکم زهکشی؛ (ر) شاخص پوشش گیاهی؛ (ز) زمین‌شناسی

Fig 2. Maps of the parameters: (a) DEM, (b) Slope, (c) Aspect, (d) Drainage density, (e) NDVI, (f) Geology



ادامه شکل ۲- نقشه‌های پارامترهای (س) کاربری اراضی؛ (ش) بارندگی
Continuo Fig 2. Maps of the parameters: (g) Land use, (h) Rain

جنگل تصادفی^۱ (RF)

مدل RF یک الگوریتم یادگیری ماشینی است که از چندین درخت تصمیم تشکیل شده است. توسط بریمن بر اساس نظریه یادگیری گروه بگینگ پیشنهاد شد [۱، ۲۵، ۵۳]. یک مدل RF شامل چندین درخت تصمیم است که توسط الگوریتم یادگیری گروه بگینگ آموزش داده شده است. برای طبقه‌بندی یک نمونه، نتیجه نهایی از طریق مکانیسم رأی‌گیری بسیاری از درختان تصمیم تعیین می‌شود. این روش مشکل عملکرد درختان تصمیم را حل کرده است و دارای موازی و گسترش خوبی برای طبقه‌بندی داده‌های با ابعاد بالا و تحمل عالی برای نویز و ناهنجاری‌ها است [۱۹، ۲۶، ۵۵]. جنگل‌های تصادفی (RF) در دهه گذشته به دلیل توانایی آن‌ها در رسیدگی به مشکلات طبقه‌بندی چند بعدی و رگرسیون با دقت عالی و احتمال کم بیش از حد برازش، شتاب بیش‌تری به دست آورده‌اند. RF یک روش یادگیری مجموعه‌ای است که شامل تجمع بوت استرپ چندین درخت رگرسیون برای پیش‌بینی یک نتیجه، خوب در مدیریت مجموعه داده‌های عظیم، ناهمگونی داده‌ها، فضا‌های ویژگی با ابعاد بالا و ساختارهای داده پیچیده است. در حال حاضر، جنگل تصادفی یکی از بهترین الگوریتم‌های یادگیری است که در علوم زیست محیطی کاربرد زیادی دارد [۳، ۲۸، ۴۸، ۵۱].

ارزیابی مدل

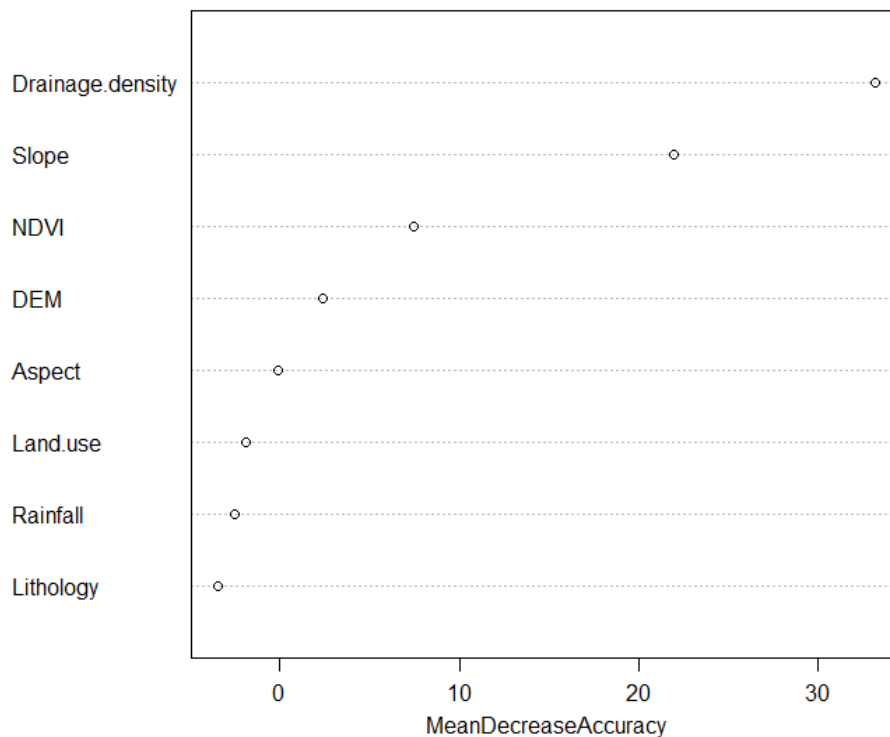
اعتبارسنجی نقشه حساسیت به وقوع سیل با استفاده از منحنی (ROC^۲)

یکی از مهمترین گام‌های اصلی در مدل‌سازی، صحت‌سنجی نقشه‌های خروجی مدل‌سازی است که اهمیت علمی مدل‌ها با آن

عملکرد یک مدل ترکیب می‌کند. دو الگوریتم مختلف، یعنی تقویت و رگرسیون، در مدل استفاده می‌شوند و نقاط قوت این الگوریتم‌ها برای بهبود دقت مدل و کاهش واریانس مدل ترکیب می‌شوند. تقویت یکی از قوی‌ترین روش‌های یادگیری برای بهبود دقت مدل است، با تطبیق مکرر درخت‌های جدید با خطاهای باقی‌مانده (RE) مجموعه درختی موجود. علاوه بر تقویت، مدل BRT از درخت‌های رگرسیون در فرآیند مدل‌سازی استفاده می‌کند. درختان رگرسیون از طبقه‌بندی و رویکردهای درخت رگرسیون از گروه مدل‌های درخت تصمیم طبقه‌بندی می‌شوند [۲۴، ۴۱].

در مدل، از بین پارامترهای مختلف، تعداد درخت‌ها به‌طور خودکار از طریق اعتبارسنجی متقاطع داخلی تنظیم می‌شود. علاوه بر این، نرخ یادگیری، تعداد گره‌ها در یک درخت و کسر کیسه از طریق روش آزمون و خطا تعیین شد. درختان رگرسیون پیشرفته دارای مزایای مهم روش‌های مبتنی بر درخت، مدیریت انواع مختلف متغیرهای پیش‌بینی‌کننده و تطبیق داده‌های از دست رفته هستند. آن‌ها هیچ نیازی به تبدیل داده‌های قبلی یا حذف موارد پرت ندارند، می‌توانند روابط غیرخطی پیچیده را منطبق کنند و به‌طور خودکار اثرات متقابل بین پیش‌بینی‌کننده‌ها را کنترل کنند. نصب چندین درخت در BRT بزرگ‌ترین اشکال مدل‌های تک درختی یعنی عملکرد پیش‌بینی نسبتاً ضعیف آن‌ها را برطرف می‌کند. اگرچه مدل‌های BRT پیچیده هستند، اما می‌توان آن‌ها را به روش‌هایی خلاصه کرد که بینش زیست‌محیطی قدرتمندی به دست می‌دهد و عملکرد پیش‌بینی آن‌ها نسبت به اکثر روش‌های مدل‌سازی سنتی برتر است. ویژگی‌های منحصربه‌فرد BRT تعدادی از مسائل عملی را در برازش مدل ایجاد می‌کند [۲۱].

1. Random Forest (RF)
2. Receiver Operating Characteristics



شکل ۳- اولویت‌بندی تأثیر عوامل مؤثر خطر سیل
Fig 3. Prioritizing the effect of flood risk factors

میانگین کاهشی جینی، اهمیت هر یک از عوامل مؤثر در وقوع خطر سیل مشخص شد. میانگین کاهشی دقت معیار دقیق‌تری برای بررسی اهمیت معیارها در مدل‌سازی طبقه‌بندی است [۳۱، ۴۹، ۶۲]. بنابراین نتایج میانگین کاهشی دقت خطر سیل در شکل (۳) آمده است. همان‌طور که در شکل آمده است در این پژوهش، تراکم زهکشی، شیب و NDVI سه عامل مؤثر در خطر سیل نشان داده شد.

نتایج تهیه نقشه حساسیت‌پذیری

نقشه‌های حساسیت‌پذیری منطقه مورد مطالعه با استفاده از الگوریتم یادگیری ماشین جنگل تصادفی، درخت رگرسیون پیشرفته و ماشین بردار پشتیبان در شکل‌های ۴ تا ۶ ارائه شده است.

هم‌چنین درصد طبقات نقشه‌های حساسیت‌پذیری سیل‌گیری با استفاده از مدل‌های مختلف یادگیری ماشین در شکل (۷) ارائه شده است.

ارزیابی نقشه‌های حساسیت‌پذیری منطقه مورد مطالعه

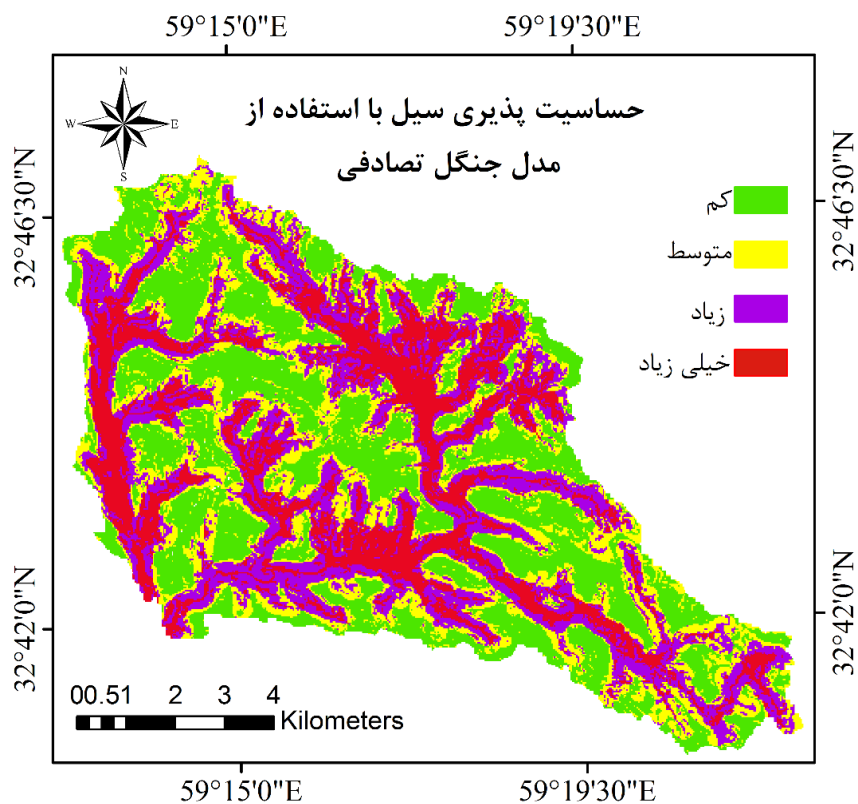
همان‌طور که بیان شد به منظور ارزیابی نقشه‌های حساسیت‌پذیری تهیه شده با استفاده از الگوریتم‌های یادگیری ماشین، از ۳۰ درصد نقاط استفاده نشده در فرآیند مدل‌سازی و منحنی تشخیص عملکرد نسبی (ROC) استفاده شده است. نتایج منحنی تشخیص عملکرد نسبی (ROC) و سطح زیر منحنی نقشه‌های پهنه‌بندی در جدول (۱) و شکل (۸) ارائه شده است.

سنجیده می‌شود [۴۷]. منحنی مشخصه عملکرد سیستم (ROC) مهم‌ترین روشی است که برای تأیید مدل‌های حساسیت استفاده می‌شود [۲۰]. منحنی ROC که نسبت پیکسل‌های پیش‌بینی‌شده درست توسط مدل (نرخ مثبت واقعی)، که به عنوان حساسیت نیز شناخته می‌شود) را در مقابل نسبت پیکسل‌های پیش‌بینی نادرست (نرخ مثبت کاذب)، ترسیم می‌کند. سطح زیر منحنی یک اندازه‌گیری معنادار برای ارزیابی عملکرد مدل است، بنابراین، AUC برای اعتبارسنجی موفقیت مدل استفاده می‌شود. از ۵/۰ (توانایی پیش‌بینی ضعیف) تا ۱ (بالاترین دقت و قابلیت اطمینان) متغیر است. به طور معمول، مقادیر بیش از ۸/۰ نشان‌دهنده عملکرد بسیار خوب مدل است [۱۳].

نتایج

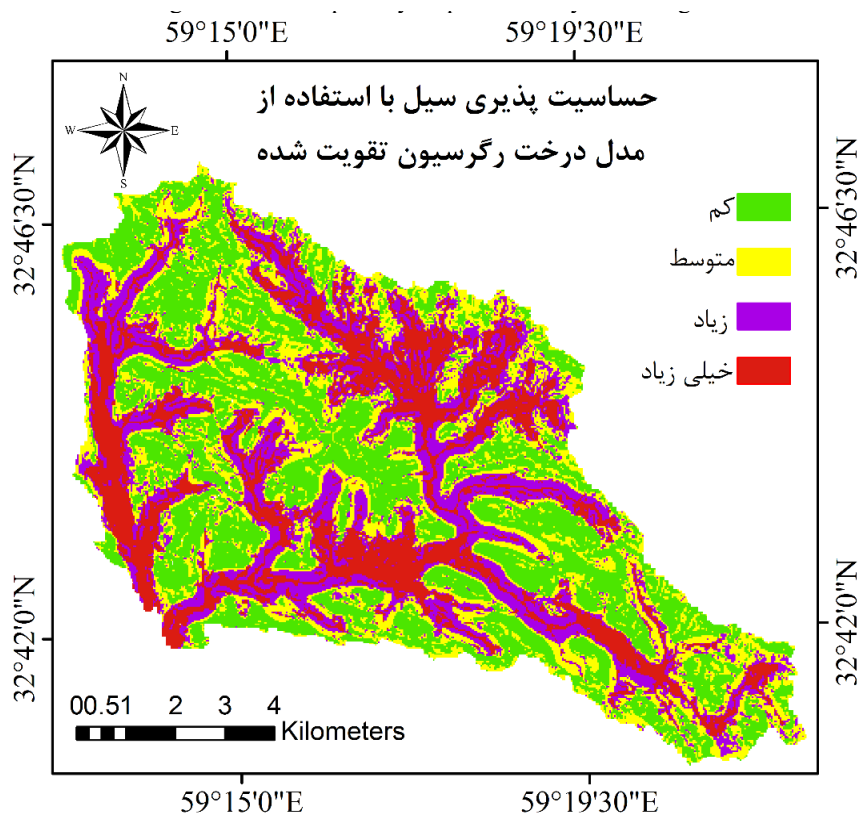
تحقیق حاضر با هدف تعیین مناطق مستعد سیل‌خیزی در منطقه حوزه آبخیز شهرستانک شهرستان خوسف بر اساس تکنیک‌های مختلف یادگیری انجام شده است. بدین منظور پس از جمع‌آوری داده‌های موجود از سیل‌های رخ داده، نقشه پراکنش سیل توسط آن‌ها ترسیم و شناسایی عوامل مؤثر بر وقوع سیل با استفاده از الگوریتم جنگل تصادفی، نقشه‌های حساسیت‌پذیری سیل منطقه مورد مطالعه با استفاده از تکنیک‌های یادگیری ماشین تهیه شد.

نتایج بررسی اهمیت متغیرهای مورد استفاده در خطر سیل در مدل جنگل تصادفی با استفاده از میانگین کاهشی دقت و



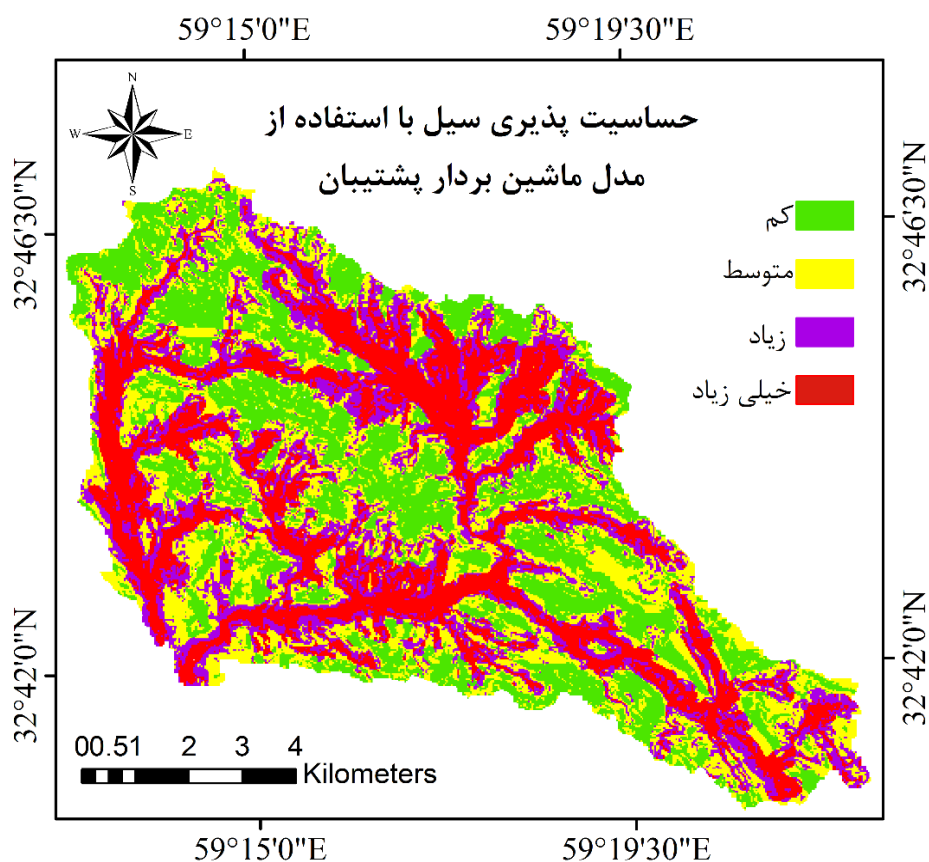
شکل ۴- نقشه حساسیت پذیری سیل گیری منطقه مورد مطالعه با استفاده از الگوریتم جنگل تصادفی

Fig 4. Flood susceptibility map of the study area using RF



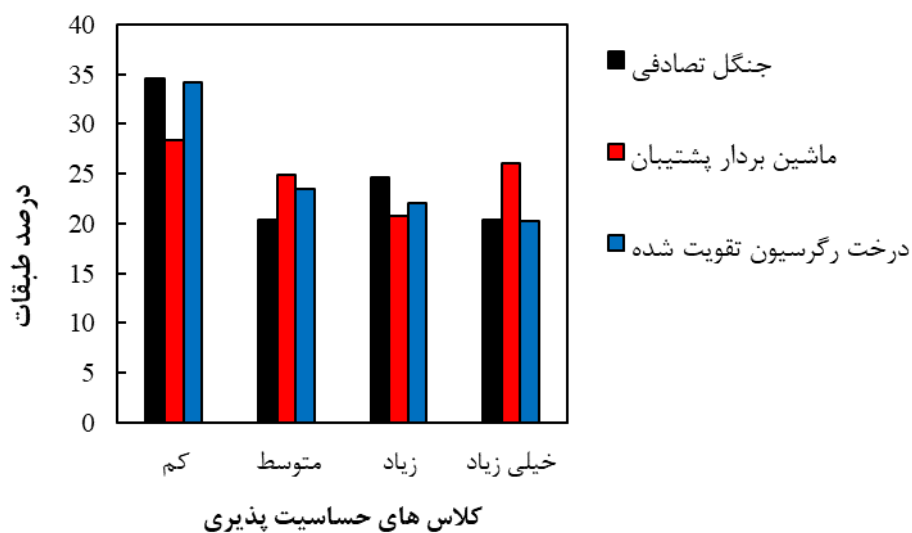
شکل ۵- نقشه حساسیت پذیری سیل گیری منطقه مورد مطالعه با استفاده از الگوریتم درخت رگرسیون تقویت شده

Fig 5. Flood susceptibility map of the study area using the BRT



شکل ۶- نقشه حساسیت پذیری سیل گیری منطقه مورد مطالعه با استفاده از الگوریتم ماشین بردار پشتیبان

Fig 6. Flooding susceptibility map of the study area using SVM



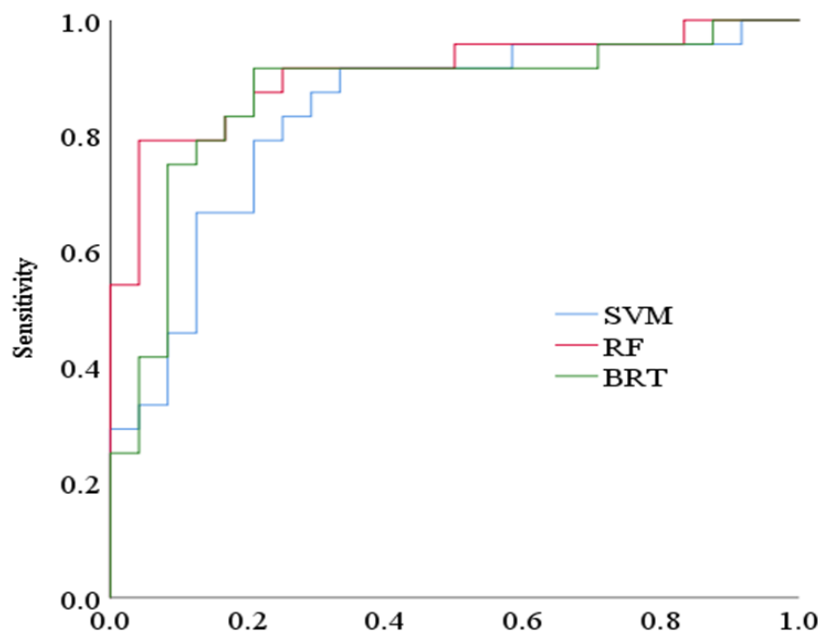
شکل ۷- درصد طبقه های حساسیت پذیری سیل در مدل های مختلف یادگیری ماشین

Fig 7. The percentage of flood sensitivity classes in different machine learning models

جدول ۱- نتایج سطح زیرمنحنی نقشه‌های حساسیت‌پذیری سیل‌گیری با الگوریتم‌های یادگیری ماشین

Table 1. The results of the sub-curve level of flood susceptibility maps with machine learning algorithms

مدل (Model)	AUC	خطای استاندارد (SE)	معنی‌داری مجانبی (AS)	سطح معنی‌داری با اطمینان ۹۵ درصد Significant level 95%	
				حد بالا (Upper line)	حد پایین (lower line)
جنگل تصادفی (RF)	0.908	0.045	0.000	0.820	0.996
درخت رگرسیونی پیشرفته (BRT)	0.870	0.056	0.000	0.760	0.980
ماشین بردار پشتیبان (SVM)	0.837	0.060	0.000	0.720	0.954



شکل ۸- اعتبارسنجی روش‌ها با روش ROC

Fig 8. Validation of methods with ROC method

مسائل است [۱۴۸] و کاربرد گسترده‌ای در تحلیل مخاطرات طبیعی دارد [۱۶، ۳۱]، با این حال، نقطه ضعف این روش می‌تواند حساسیت نقشه‌های تهیه شده به قضاوت کارشناسان و حساسیت به تغییر در وزن‌های متغیرهای مورد استفاده در تحقیق باشد. به عبارت دیگر، این روش ممکن است به دلیل وابستگی به نظریات و انتخاب‌های کارشناسان و همچنین تغییر در وزن‌دهی متغیرها، به دقت و صحت مناسبی در پیش‌بینی وقوع بحران‌های طبیعی نرسد [۹]. قابل ذکر است که یادگیری ماشین از روش‌های مدل‌سازی است که برای درک روابط پیچیده و تهیه نقشه ریسک بین طیف وسیعی از عوامل هواشناسی (باد، فشار، افزایش طوفان و سیلاب) استفاده می‌شود [۴۲]. بنابراین این مدل‌سازی‌ها می‌تواند به پیش‌بینی تأثیر چنین رویدادهای چندمخاطره‌ای کمک کند، جایی که این تأثیرات بسته به موقعیت مکانی و عوامل اقلیمی به‌طور قابل توجهی متفاوت است [۴۷]. در پژوهش حاضر، به‌منظور تهیه نقشه حساسیت‌پذیری سیل‌گیری منطقه مورد مطالعه از سه مدل یادگیری ماشین (الگوریتم

بحث و نتیجه‌گیری

هر سال، به دلیل طغیان آب رودخانه‌ها و سیلاب‌ها، گستره وسیعی از کشور تحت تأثیر قرار می‌گیرد و این موضوع منجر به ویران شدن تأسیسات عمرانی، امکانات ارتباطی، زمین‌های کشاورزی، شهرها و روستاها می‌شود [۴۲]. برای کاهش شدت سیلاب و خسارات ناشی از آن، انسان باید با بهره‌گیری از روش‌های مختلف اقدام به تلاش کند. این اقدامات می‌تواند شامل تمرکز بر عملیات آب‌خیزداری در مناطق خطرزا و بحرانی، استفاده از روش‌های نوین مهندسی آب و همچنین افزایش ظرفیت زیرساخت‌های مقاوم در برابر سیلاب باشد.

مدل‌های مورد استفاده در تهیه نقشه حساسیت‌پذیری سیل‌گیری

طی سال‌های اخیر روش‌های به‌کاربرده شده به‌منظور بررسی و تعیین نقشه‌های پتانسیل سیلاب، شامل روش نسبت فراوانی، آنالیزهای آماری چند متغیره، فرآیند تحلیل سلسله مراتبی، روش ماشین بردار پشتیبان و درخت تصمیم‌گیری است. به‌عنوان مثال، روش تحلیل سلسله مراتبی یکی از پرکاربردترین روش‌ها برای حل

درخت رگرسبون تقویت شده، جنگل تصادفی و ماشین بردار پشتیبان) استفاده شد.

نتایج نشان داد که همه مدل‌ها توانایی بالایی در تعیین مناطق حساس به وقوع سیل در منطقه مورد مطالعه دارند که نشان از پتانسیل بالای این مدل‌هاست که با تحقیقات [۳۲، ۳۶، ۶۵] که بیان داشتند الگوریتم‌های یادگیری ماشین برای تهیه نقشه‌های خطر سیل و مخاطرات طبیعی دیگر دقت و قابلیت بالایی دارند.

در تعیین مناطق حساس به وقوع سیل، مدل جنگل تصادفی، درخت رگرسبون تقویت شده و ماشین بردار پشتیبان به ترتیب دارای دقت بالاتری هستند. که این یافته‌ها با نتایج تحقیقات [۱۱، ۱۲، ۶۶] مطابقت و با نتایج مطالعات پیشین در پیش‌بینی سیلاب با روش ماشین بردار پشتیبان نتایج قابل قبولی را ارائه دادند [۶، ۱۵، ۶۴] همخوانی ندارد.

اهمیت متغیرهای مورد استفاده در خطر سیل گیری

در این پژوهش الگوریتم جنگل تصادفی برای اولویت‌بندی عامل‌های مؤثر به کار برده شد. یک مزیت قابل توجه در الگوریتم جنگل تصادفی این است که امکان مقایسه بالقوه مطالعات در مناطق مختلف جهان را با توجه به ارائه نتایج به صورت کمی، فراهم می‌کند [۱۱، ۳۹، ۴۸، ۵۰]. نتایج پژوهش حاضر با استفاده از الگوریتم جنگل تصادفی نشان داد که عامل‌های تراکم زهکشی، شیب و شاخص پوشش گیاهی بیش‌ترین نقش را در وقوع سیل منطقه داشته است. عمدتاً پوشش سطوح غیرقابل نفوذ، باعث افزایش رواناب و با آبگرفتگی همراه می‌شود [۵۱، ۵۲]. نتایج پژوهش چن و همکاران [۱۰] نشان داد که تراکم زهکشی، تأثیر قابل توجهی در وقوع سیل دارد که با پژوهش حاضر مطابقت دارد. نتایج پژوهش پورقاسمی و همکاران [۴۶]، برای اولویت‌بندی عوامل مؤثر در سیل با استفاده از الگوریتم جنگل تصادفی نشان داد که تراکم زهکشی و شاخص رطوبت توپوگرافی از مؤثرترین عوامل در پیش‌بینی خطر سیل هستند. برخی مطالعات پیشین نشان دادند که ارتفاع مهم‌ترین عامل خطر سیل گیری در منطقه مورد مطالعه است [۲۹، ۵۰]. عرب عامری و همکاران [۳] گزارش دادند که NDVI، ارتفاع و کاربری اراضی، مهم‌ترین عامل در حساسیت به سیل در شمال ایران هستند. وفاخواه و همکاران [۵۷] در پژوهش دیگر نشان دادند که فاصله از رودخانه‌ها، ارتفاع از سطح دریا و تراکم زهکشی مهم‌ترین عامل برای وقوع سیل و کم اهمیت‌ترین عامل زمین‌شناسی است و خاطر نشان کردند که اهمیت متغیرها در نقشه سیل، تحت تأثیر روش تحقیق و ویژگی‌های منطقه مورد مطالعه قرار می‌گیرد. به عبارت دیگر، شرایط مختلف زمین‌شناسی، توپوگرافی، و اقلیم در یک منطقه، می‌تواند اولویت عوامل مؤثر در وقوع سیل را تغییر دهد. بنابراین، متغیرهای مؤثر بر وقوع خطرات طبیعی در یک منطقه، به دلیل متفاوت بودن شرایط توپوگرافی، اقلیمی، هیدرولوژیکی، ژئومورفولوژیکی، خاک‌شناسی و عوامل انسانی، ممکن است با منطقه دیگر متفاوت باشد.

نتیجه‌گیری کلی

سیل را می‌توان به‌عنوان یکی از پر مخاطره‌ترین بلایای طبیعی دانست که منطقه حوزه آبخیز شهرستانک شهرستان خوسف، همواره در معرض آن قرار می‌گیرد. بنابراین درک بهتر از عوامل کنترل سیل، جهت توسعه پایدار مناطق مستعد ابتلا به این خطر مانند استان خراسان جنوبی بسیار مهم است و تهیه نقشه‌های پتانسیل سیل ابزار مهمی برای کاهش آن محسوب می‌شود. بدین منظور در تحقیق حاضر نقشه حساسیت‌پذیری سیل منطقه حوزه آبخیز شهرستانک شهرستان خوسف برای تعیین روش دقیق در شناسایی مناطق مستعد سیل، سه مدل یادگیری ماشین بررسی شد که مدل جنگل تصادفی با دقت ۰/۹۰۸ در مقایسه با مدل درخت رگرسبون تقویت شده با دقت ۰/۸۷۰ و مدل ماشین بردار پشتیبان با دقت ۰/۸۳۷ با استفاده از منحنی تشخیص عملکرد نسبی، به‌عنوان بهترین مدل ارزیابی شد. شناسایی عوامل مؤثر بر وقوع سیل با استفاده از الگوریتم جنگل تصادفی، نیز نشان داد که تأثیرگذارترین عوامل به ترتیب، تراکم زهکشی، شیب و شاخص پوشش گیاهی بوده است. هم‌چنین نتایج بررسی خطر سیل گیری منطقه حوزه آبخیز شهرستانک شهرستان خوسف درصد حساسیت‌پذیری آن‌ها را در چهار طبقه (خیلی زیاد، زیاد، متوسط و کم) نشان داد. به‌طور کلی، برای مدیریت خطر سیل، تهیه نقشه حساسیت‌پذیری می‌تواند نقشی اساسی در برنامه‌ریزی منطقه‌ای کاربری اراضی و مدیریت منابع داشته باشد و به عنوان یک ابزار مفید برای برنامه‌ریزی و مدیریت مخاطرات محیطی، مورد استفاده قرار می‌گیرد و به کاربران امکان می‌دهد تا با شناسایی مناطق پرخطر و تهیه نقشه‌های ریسک، برنامه‌های پیشگیری و مدیریت مناسبی را برای کاهش خطرات طبیعی در نظر بگیرند.

تقدیر و تشکر

این پژوهش در قالب طرح پژوهشی به شماره ابلاغیه ۹۴۲۹/۱۴۰۱ مورخه ۱۴۰۱/۱۲/۲۱ و با استفاده از اعتبارات پژوهشی دانشگاه بیرجند انجام شده است که بدین وسیله تشکر و قدردانی می‌شود.

منابع

1. Alber, M. Buganza Tepole, A. Cannon, W.R. De, S. Dura-Bernal, S. Garikipati, K. and Kuhl, E. 2019. Integrating machine learning and multiscale modeling—perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences. NPJ digital medicine, 2(1): 115.
2. Alvarado-Aguilar, D. Jiménez, J.A. and Nicholls, R. J. 2012. Flood hazard and damage assessment in the Ebro Delta (NW Mediterranean) to relative sea level rise. Natural Hazards, 62: 1301-1321.
3. Arabameri, A. Pourghasemi, H.R. and Yamani, M. 2017.

14. Cheng, Q.S. 1997. Analytic hierarchy process (AHP) and attribute hierarchical model (AHM). *Theory and Practice*. 527.
15. Chu, C.T. Kim, S. Lin, Y.A. Yu, Y. Bradski, G. Olukotun, K. and Ng, A. 2006. Map-reduce for machine learning on multicore. *Advances in neural information processing systems*, 19: 86-102.
16. Dai, F.C. Lee, C.F., and Zhang, X.H. 2001. GIS-based geo-environmental evaluation for urban land-use planning: a case study. *Engineering geology*, 61(4): 257-271.
17. Darabi, H. Choubin, B. Rahmati, O. Haghighi, A. T. Pradhan, B. and Kløve, B. 2019. Urban flood risk mapping using the GARP and QUEST models: A comparative study of machine learning techniques. *Journal of hydrology*, 569: 142-154.
18. Koning, K. Filatova, T. Need, A. and Bin, O. 2019. Avoiding or mitigating fooding: Bottom-up drivers of urban resilience to climate change in the USA. *Glob Environ Change*, 59: 101981.
19. Deng, W. Zhang, X. Zhou, Y. Liu, Y. Zhou, X. Chen, H. and Zhao, H. 2022. An enhanced fast non-dominated solution sorting genetic algorithm for multi-objective problems. *Information Sciences*, 585: 441-453.
20. El-Haddad, B.A. Youssef, A.M. Pourghasemi, H.R. Pradhan, B. El-Shater, A.H. and El-Khashab, M.H. 2021. Flood susceptibility prediction using four machine learning techniques and comparison of their performance at Wadi Qena Basin, Egypt. *Natural Hazards*, 105: 83-114.
21. Elith, J. Leathwick, J.R. and Hastie, T. 2008. A working guide to boosted regression trees. *Journal of animal ecology*, 77(4): 802-813.
22. Erena, S.H. and Worku, H. 2018. Flood risk analysis: causes and landscape based mitigation strategies in Dire Dawa city, Ethiopia. *Geoenviron Disas* 5:16.
23. Erener, A. Mutlu, A. and Düzgün, H.S. 2016. A comparative study for landslide susceptibility mapping using GIS-based multi-criteria decision analysis (MCDA), logistic regression (LR) and association rule mining (ARM). *Engineering geology*, 203: 45-55.
24. Fernández-Delgado, M. Cernadas, E. Barro, S. and Amorim, D. 2014. Do we need hundreds of classifiers to solve real world classification problems?. *The journal of machine learning research*, 15(1): 3133-3181.
25. Golkarian, A. Naghibi, S.A. Kalantar, B. and Pradhan, B. 2018. Groundwater potential mapping using C5. 0, random forest, and multivariate adaptive regression spline models in GIS. *Environmental monitoring and assessment*, 190: 1-16.
26. Gounaridis, D. Chorianopoulos, I. Symeonakis, E. and Applying different scenarios for landslide spatial modeling using computational intelligence methods. *Environmental earth sciences*, 76: 1-20.
4. Ballabio, C. and Sterlacchini, S. 2012. Support vector machines for landslide susceptibility mapping: Staffora River Basin case study, Italy. *Mathematical geosciences*, 44: 47-70.
5. Bui, D.T. Hoang, N.D. Martínez-Álvarez, F. Ngo, P. T.T. Hoa, P.V. Pham, T.D. and Costache, R. 2020. A novel deep learning neural network approach for predicting flash flood susceptibility: A case study at a high frequency tropical storm area. *Science of The Total Environment*, 701: 134413.
6. Bui, D.T. Pradhan, B. Nampak, H. Bui, Q.T. Tran, Q. A. and Nguyen, Q.P. 2016. Hybrid artificial intelligence approach based on neural fuzzy inference model and metaheuristic optimization for flood susceptibility modeling in a high-frequency tropical cyclone area using GIS. *Journal of Hydrology*, 540: 317-330.
7. Caruso, G.D. 2017. The legacy of natural disasters: The intergenerational impact of 100 years of disasters in Latin America. *J. Dev. Econ*. 127: 209-233.
8. Chan, F.S. 2003. Interactive selection model for supplier selection process: an analytical hierarchy process approach. *International Journal of Production Research*, 41(15): 3549-3579.
9. Chang, H. Franczyk, J. and Kim, C. 2009. What is responsible for increasing flood risks? The case of Gangwon Province, Korea. *Natural Hazards*, 48: 339-354.
10. Chen, J. Huang, G. and Chen, W. 2021. Towards better flood risk management: Assessing flood risk and investigating the potential mechanism based on machine learning models. *Journal of environmental management*, 293: 112810.
11. Chen, W. Li, Y. Xue, W. Shahabi, H. Li, S. Hong, H. and Ahmad, B.B. 2020. Modeling flood susceptibility using data-driven approaches of naïve bayes tree, alternating decision tree, and random forest methods. *Science of The Total Environment*, 701: 134979.
12. Chen, W. Panahi, M. and Pourghasemi, H.R. 2017. Performance evaluation of GIS-based new ensemble data mining techniques of adaptive neuro-fuzzy inference system (ANFIS) with genetic algorithm (GA), differential evolution (DE), and particle swarm optimization (PSO) for landslide spatial modelling. *Catena*, 157: 310-324.
13. Chen, W. Shahabi, H. Shirzadi, A. Li, T. Guo, C. Hong, H. and Bin Ahmad, B. 2018. A novel ensemble approach of bivariate statistical-based logistic model tree classifier for landslide susceptibility assessment. *Geocarto International*, 33(12): 1398-

and Ghazali, A.H.B. 2017. Ensemble machine-learning-based geospatial approach for flood risk assessment using multi-sensor remote-sensing data and GIS. *Geomatics, Natural Hazards and Risk*, 8(2): 1080-1102.

39. Norallahi, M. and Kaboli, H. 2021. Urban flood hazard mapping using machine learning models: GARP, RF, MaxEnt and NB. *Natural Hazards*, 106: 119-137.

40. Papaioannou, G. Vasiliades, L. Loukas, A. 2015. multi-criteria analysis framework for potential flood prone areas mapping. *Water Resour Manag*, 29(2): 399-418

41. Park, W. Kim, D. Lu, Y. and Cho, M. 2019. Relational knowledge distillation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 52: 3967-3976.

42. Pilkington, S. and Mahmoud, H. 2017. Spatial and temporal variations in resilience to tropical cyclones along the United States coastline as determined by the multi-hazard hurricane impact level model. *Palgrave Communications*, 3(1): 152-163.

43. Pourghasemi, H.R. and Kerle, N. 2016. Random forests and evidential belief function-based landslide susceptibility assessment in Western Mazandaran Province, Iran. *Environmental earth sciences*, 75: 1-17.

44. Pourghasemi, H.R. and Rossi, M. 2017. Landslide susceptibility modeling in a landslide prone area in Mazandarn Province, north of Iran: a comparison between GLM, GAM, MARS, and M-AHP methods. *Theoretical and Applied Climatology*, 130(1-2): 609-633.

45. Pourghasemi, H.R. Gayen, A. Panahi, M. Rezaie, F. and Blaschke, T. 2019. Multi-hazard probability assessment and mapping in Iran. *Science of the total environment*, 692: 556-571.

46. Pourghasemi, H.R. Kariminejad, N. Amir, M. Edalat, M. Zarafshar, M. Blaschke, T. and Cerda, A. 2020. Assessing and mapping multi-hazard risk susceptibility using a machine learning technique. *Scientific reports*, 10(1): 3203.

47. Pourghasemi, H.R. Moradi, H.R., and Fatemi Aghda, S.M. 2013. Landslide susceptibility mapping by binary logistic regression, analytical hierarchy process, and statistical index models and assessment of their performances. *Natural hazards*, 69: 749-779.

48. Pourghasemi, H.R. Razavi-Termeh, S.V. Kariminejad, N. Hong, H. and Chen, W. 2020. An assessment of metaheuristic approaches for flood assessment. *Journal of Hydrology*, 582: 124536.

49. Pourghasemi, H.R. Yousefi, S. Kornejady, A. and Cerdà, A. 2017. Performance assessment of individual and ensemble data-

Koukoulas, S. 2019. A Random Forest-Cellular Automata modelling approach to explore future land use/cover change in Attica (Greece), under different socio-economic realities and scales. *Science of the Total Environment*, 646: 320-335.

27. Han, D. Chan, L. and Zhu, N. 2007. Flood forecasting using support vector machines. *Journal of hydroinformatics*, 9(4): 267-276.

28. Hong, K.S. and Naseer, N. 2016. Reduction of delay in detecting initial dips from functional near-infrared spectroscopy signals using vector-based phase analysis. *International journal of neural systems*, 26(03): 1650012.

29. Kanani-Sadat, Y. Arabsheibani, R. Karimipour, F. and Nasser, M. 2019. A new approach to flood susceptibility assessment in data-scarce and ungauged regions based on GIS-based hybrid multi criteria decision-making method. *Journal of hydrology*, 572: 17-31.

30. Kazemi, H. Mansouri, N. and Jozi, S.A. 2022. Flood risk zoning in Nowshahr city using machine learning models. *JHRE*, 40(176): 71-86.

31. Laskar, J.I. Sen, M.K. Dutta, S. and Laskar, S.A. 2021. A Flood Resilience Analytics Framework for Housing Infrastructure Systems Based on Dempster-Shafer (Evidence) Theory. *Journal of Performance of Constructed Facilities*, 35(6): 04021073.

32. Lee, S. Kim, J.C. Jung, H.S. Lee, M.J. and Lee, S. 2017. Spatial prediction of flood susceptibility using random-forest and boosted-tree models in Seoul metropolitan city, Korea. *Geomatics, Natural Hazards and Risk*, 8(2): 1185-1203.

33. Marfai, M.A. Sekaranom, A.B. and Ward, P. 2015. Community responses and adaptation strategies toward flood hazard in Jakarta, Indonesia. *Nat Hazards*, 75: 1127-1144

34. Markantonis, V. Meyer, V. and Lienhoop, N. 2013. Evaluation of the environmental impacts of extreme floods in the Evros river basin using contingent valuation method. *Nat Hazards*, 69: 1535-1549

35. Mbow, C. Diop, A. Diaw, A.T. and Niang, C.I. 2008. Urban sprawl development and flooding at Yeumbeul suburb (Dakar-Senegal). *Afr J Environ Sci Technol*, 2: 075-088.

36. Milanović Pešić, A. 2020. Hydrological aspects of the floods in the Kolubara River basin (Serbia)—analyses and flood mitigation measures. *Smart Geography: 100 Years of the Bulgarian Geographical Society*, 117-127.

37. Mishra, K. and Sinha, R. 2020. Flood risk assessment in the Kosi megafan using multi-criteria decision analysis: a hydro-geomorphic approach. *Geomorphology*, 350: 106861.

38. Mojaddadi, H. Pradhan, B. Nampak, H. Ahmad, N.

58. Vakhshoori, V. and Zare, M. 2016. Landslide susceptibility mapping by comparing weight of evidence, fuzzy logic, and frequency ratio methods. *Geomatics, Natural Hazards and Risk*, 7(5): 1731-1752.
59. Vapnik, V. 1999. *The Nature of Statistical Learning Theory*: SpringerVerlag. New York. 481.
60. Wang, S.Y.S., Zhao, L. and Gillies, R.R. 2016. Synoptic and quantitative attributions of the extreme precipitation leading to the August 2016 Louisiana flood. *Geophysical Research Letters*, 43(22): 11-805.
61. Wang, Y. Fang, Z. Hong, H. and Peng, L. 2020. Flood susceptibility mapping using convolutional neural network frameworks. *Journal of Hydrology*, 582: 124482.
62. Westreich, D. Lessler, J. and Funk, M.J. 2010. Propensity score estimation: neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression. *Journal of clinical epidemiology*, 63(8): 826-833.
63. Youssef, A.M. Pradhan, B. and Sefry, S.A. 2016. Flash flood susceptibility assessment in Jeddah city (Kingdom of Saudi Arabia) using bivariate and multivariate statistical models. *Environmental Earth Sciences*, 75(1): 12.
64. Yu, B. Chen, F. and Xu, C. 2020. Landslide detection based on contour-based deep learning framework in case of national scale of Nepal in 2015. *Computers & Geosciences*, 135: 104388.
65. Zhang, Q. Huber, H. Beljaars, S.J. Birnbaum, D. de Best, S. de Kroon, H. and Visser, E.J. 2017. Benefits of flooding-induced aquatic adventitious roots depend on the duration of submergence: linking plant performance to root functioning. *Annals of Botany*, 120(1): 171-180.
66. Zhao, G. Pang, B. Xu, Z. Yue, J. and Tu, T. 2018. Mapping flood susceptibility in mountainous areas on a national scale in China. *Science of the Total Environment*, 615: 1133-1142.
- mining techniques for gully erosion modeling. *Science of the Total Environment*, 609: 764-775.
50. Pouyan, S. Pourghasemi, H.R. Bordbar, M. Rahmanian, S. and Clague, J.J. 2021. A multi-hazard map-based flooding, gully erosion, forest fires, and earthquakes in Iran. *Scientific Reports*, 11(1): 14889.
51. Rahmati, O. and Pourghasemi, H.R. 2017. Identification of critical flood prone areas in data-scarce and ungauged regions: a comparison of three data mining models. *Water resources management*, 31: 1473-1487.
52. Rahmati, O. Zeinivand, H. and Besharat, M. 2016. Flood hazard zoning in Yasooj region, Iran, using GIS and multi-criteria decision analysis. *Geomatics, Natural Hazards and Risk*, 7(3): 1000-1017.
53. Sadler, J.M. Appling, A.P. Read, J.S. Oliver, S.K. Jia, X. Zwart, J.A. and Kumar, V. 2022. Multi-task deep learning of daily streamflow and water temperature. *Water Resources Research*, 58(4): e2021WR030138.
54. Siyadati far, M.H. 2022. Landslide sensitivity and zoning evaluation using support vector machine algorithm (Case study: Shahrstanak-Eilki watershed of Khosf city). Master's thesis in watershed science and engineering. Faculty of Natural Resources and Environment. University of Birjand (in Persian).
55. Speiser, J.L. Karvellas, C.J. Wolf, B.J. Chung, D. Koch, D.G. and Durkalski, V.L. 2019. Predicting daily outcomes in acetaminophen-induced acute liver failure patients with machine learning techniques. *Computer methods and programs in biomedicine*, 175: 111-120.
56. Tehrany, M.S. Pradhan, B. Mansor, S. and Ahmad, N. 2015. Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. *Catena*, 125: 91-101.
57. Vafakhah, M. Hasani Loor, S. Pourghasemi, H.R. and Katebikord, A. 2020. Comparing performance of random forest and adaptive neuro-fuzzy inference system data mining models for flood susceptibility mapping. *Arabian Journal of Geosciences*, 13: 1-16.

Determining Flood-Prone Areas Using Machine Learning Models in the Shahrestank Watershed Area of Khosuf City

J. Chezgi¹ and S. Poyan²

Received: 07-07-2023 Accepted: 20-9-2023

Abstract

Flood is one of the most destructive natural phenomenon that occur worldwide, resulting in damage to properties, infrastructure, and even loss of life. This research aims to create a zoning map of flood-prone areas in the watershed of Khosuf County, using machine learning algorithms. In this study, 8 influential parameters and random forest (RF), boosted regression tree (BRT), and support vector machine (SVM) were used to identify areas at susceptibility of flooding in the study area. A total of 81 flood-prone zones were identified. Randomly, 70% of the data were allocated for the modeling process, and the remaining 30% were used for evaluating the generated models. The results of parameter importance analysis indicated that the drainage density parameter had the most significant impact on flood susceptibility. Furthermore, in determining flood-prone areas, the random forest model, boosted regression tree model, and support vector machine showed higher accuracy with receiver operating characteristics (ROC) curves of 0.99, 0.98, and 0.95, respectively. The study area was divided into four categories of flood susceptibility: very high, high, moderate, and low. Overall, the watershed is considered to have a moderate to high flood susceptibility level.

Keywords: *Flooding, Drainage density, Random forest, Support vector machine, Khosuf city.*

1. Assistant Professor of Rangeland and Watershed Department, Faculty of Natural Resources and Environment, University of Birjand, Birjand. Iran. chezgi@brjand.ac.ir

2. Research Assistant, Department of Natural Resources, Faculty of Agriculture, Shiraz University, Shiraz, Iran