

ارزیابی مکانی حساسیت سیل‌گیری با استفاده از روش‌های تقویت تطبیقی و بسته‌بندی در یادگیری ماشین

مریم جهانبانی^۱، حسین آقامحمدی^{۲*}، محمدحسن وحیدنیا^۳ و زهرا عزیزی^۴

(۱) دانشجوی دکتری رشته سنجش از دور و GIS، دانشکده منابع طبیعی و محیط زیست، دانشگاه آزاد اسلامی، واحد علوم و تحقیقات، تهران، ایران.

(۲) استادیار گروه سنجش از دور و GIS، دانشکده منابع طبیعی و محیط زیست، دانشگاه آزاد اسلامی، واحد علوم و تحقیقات، تهران، ایران.

*رایانامه نویسنده مسئول مکاتبات: aghamohammadi@srbiau.ac.ir

(۳) استادیار مرکز مطالعات سنجش از دور و GIS، دانشکده علوم زمین، دانشگاه شهید بهشتی، تهران، ایران.

(۴) استادیار گروه سنجش از دور و GIS، دانشکده منابع طبیعی و محیط زیست، دانشگاه آزاد اسلامی، واحد علوم و تحقیقات، تهران، ایران.

<https://doi.org/10.71916/jrnr.2024.10310>

تاریخ پذیرش: ۱۴۰۲/۱۲/۰۹

تاریخ دریافت: ۱۴۰۲/۰۸/۲۹

چکیده

سیل هر ساله میلیاردها دلار خسارت به کشورها وارد می‌سازد و از این طریق معیشت افراد را تهدید می‌کند. این امر سبب ایجاد تهدیدات اجتماعی-اقتصادی قابل توجهی برای جمعیت سراسر جهان شده که باید کنترل و مهار گردد. در این راستا الگوریتم‌های یادگیری ماشین به همراه سیستم‌های اطلاعات جغرافیایی، ابزارهای اولیه‌ای هستند که در مدل‌سازی و تحلیل کنترل سیلاب موثر می‌باشند. هدف از پژوهش حاضر شناسایی بخشی از مناطق در معرض خطر سیل در حوضه آبریز رودخانه هراز در استان مازندران با استفاده از روش‌های مجموعه‌ای در الگوریتم‌های یادگیری ماشینی است. روند پژوهش بدین صورت بود که ابتدا داده‌های نقاط سیلابی با استفاده از تحقیقات میدانی و همچنین به‌دست آوردن آمار سیل‌های گذشته از سازمان آب منطقه‌ای مازندران تهیه شد. سپس ۷۰ درصد از حدود ۲۴۰ موقعیت نمونه برای مدل‌سازی و ۳۰ درصد باقی‌مانده برای اعتبارسنجی نقشه‌های تولیدشده استفاده شد. معیارهای موثر در این خصوص شامل زاویه شیب، جهت شیب، توپوگرافی، نوع خاک، پوشش زمین، فاصله از رودخانه، بارندگی، شاخص پوشش گیاهی تفاوت نرمال شده، شاخص انتقال رسوب، شاخص رطوبت توپوگرافیک و شاخص تراکم آبراهه برای وزن دادن تأثیر هر عامل با استفاده از الگوریتم‌های یادگیری ماشین جمعی بود. براساس نتایج این مطالعه، منحنی مشخصه عملکرد سیستم (ROC) ترسیم شد و مساحت زیر منحنی (AUC) برای اعتبارسنجی نقشه منطقه مستعد سیل محاسبه گشت. یافته‌ها نشان داد مدل تقویت تطبیقی دقیق‌تر از مدل بگینگ در تهیه نقشه حساسیت سیل‌گیری بود. تهیه نقشه حساسیت سیل‌گیری نقش محوری در توانمندسازی برنامه‌ریزان و مدیران شهری جهت کاهش و حفاظت پیشگیرانه در برابر پیامدهای نامطلوب سیل بازی می‌کند و مقامات مدیریت سیل در وزارت نیرو می‌توانند از مدل مجموعه پیشنهادی جهت کمک به مدیریت بلایا و کاهش خطرات در مطالعات آتی استفاده نمایند.

واژه‌های کلیدی: الگوریتم بگینگ، الگوریتم تقویت تطبیقی، سیل، سیستم اطلاعات مکانی، مدل‌های یادگیری ماشین ترکیبی.

مقدمه

دلایل تکرار فراوان این پدیده برهم خوردن تعادل هیدرولوژیکی و اکولوژی همانند تغییرات کاربری در طول حوضه (Yari et al., 2019) و تغییر اقلیم و همچنین از بین بردن جنگل‌ها (Vajda et al., 2020) می‌باشند. در این پژوهش به تهیه نقشه خطرپذیری سیلاب در سطح حوضه

سیل یکی از رخدادهای هیدرواقليمی و از پدیده‌های مخاطره‌انگیز طبیعی است که جوامع بشری را همواره تهدید می‌کند. سیل یعنی هرگونه افزایش در جریان رودخانه که از ظرفیت رودخانه بیشتر باشد و پهنه‌ای از زمین را که در شرایط عادی زیر آب نیست، فرا گیرد (Yang et al., 2015).

آبریز هراز پرداخته شد. رود هراز، رودی در استان مازندران در شمال ایران است که از دره لار در جنوب کوه دماوند سرچشمه گرفته و به دریای خزر می‌ریزد. آب زراعی کشاورزان آمل، فریدون‌کنار، بخشی از بابل و نور نیز از این رودخانه تامین می‌گردد. در تاریخ ۲۶ اسفند ۱۳۹۷ بر اثر بارندگی‌های شدید در استان‌های شمالی ایران از جمله گلستان و مازندران سیل جاری شد. این سیل موجب خسارت‌های مالی برای ساکنین این استان‌ها و همچنین فوت ۱۳ نفر و مصدوم شدن حداقل ۱۱۶ نفر شد. یکی از راهکارهای اساسی جهت کاهش خسارت ناشی از سیل، تهیه و استفاده از نقشه‌های حساسیت خطرپذیری سیل در سیاست‌گذاری‌ها و برنامه‌ریزی‌های عملیاتی و اجرایی است. برای تهیه این‌گونه نقشه‌ها از سیستم اطلاعات مکانی و تصاویر سنجنش از دور استفاده می‌گردد. سیستم‌های اطلاعات مکانی (جغرافیایی) سیستمی است که در داده‌کامپیوتر مکان‌محور مورد استفاده قرار می‌گیرد. به منظور دقیق‌تر شدن نقشه خطرپذیری سیل از الگوریتم‌های یادگیری ماشین مانند آدابوست^۱ و بگینگ^۲ جهت مدل‌سازی استفاده می‌شود. این الگوریتم‌ها جز طبقه‌بندی‌های ترکیبی هستند و تاکنون مطالعات بسیاری در زمینه سیل و تهیه نقشه حساسیت سیل با روش‌های مختلف صورت پذیرفته است. این مدل‌ها انواع مختلفی منجمله تحلیل تصمیم‌گیری چند معیاره (Abdullah et al., 2021) مانند تاپسیس (Rafiei-Sardooi et al., 2021; Akay et al., 2021)، فرآیند تحلیل شبکه (Gudiyangada et al., 2020) و روش تحلیل سلسله مراتبی (Bera et al., 2022; Bui et al., 2023) دارند. روش‌های تحلیل تصمیم چندمعیاره، زمانی استفاده می‌شود که بیش از یک معیار در نظر گرفته شود. تصمیم‌گیرندگان معیارهای ارزیابی را برای انتخاب از بین گزینه‌ها وزن می‌کنند. روش‌های آماری برای ارزیابی رابطه سیلاب و عوامل موثر بر آن جز روش‌های سریع و آسان می‌باشند. روش‌های آماری مورد استفاده در این امر عبارتند از: نسبت فراوانی و وزن شواهد (Plataridis & Mallios, 2023)، رگرسیون لجستیک (Shafapour Tehrani et al., 2019; Pham et al., 2021; Ali et al., 2019) و تحلیل مولفه اصلی (Sahana & Patel, 2019). مدل‌های

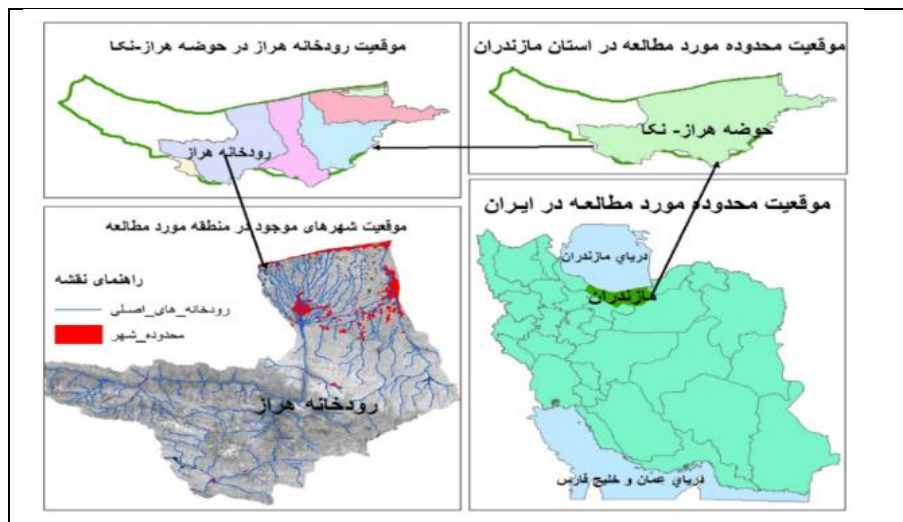
هیدرولوژیکی مانند ابزار ارزیابی آب و خاک نیز مورد استفاده قرار می‌گیرند (Wijayarathne & Coulibaly, 2020). در سال‌های اخیر، روش‌های مختلف متعددی جهت تهیه نقشه حساسیت خطرپذیری سیلاب با افزایش علاقه به تکنیک‌های یادگیری ماشین استفاده شده است (Hasanuzzaman et al., 2022; Seydi et al., 2022; Saikh & Mondal, 2023). تکنیک‌های رایج یادگیری ماشین، مانند ماشین‌های بردار پشتیبان (Khosravi et al., 2019)، شبکه‌های عصبی مصنوعی (Wang et al., 2020)، جنگل تصادفی (Chen et al., 2019; Avand et al., 2019) و درختان تصمیم (Chen et al., 2020) همه در پیش‌بینی مناطق مستعد سیلاب کارایی دارند. مزیت این است که مدل‌های ترکیبی می‌توانند نقاط ضعف مدل‌های فردی را حذف کنند تا نتایج دقیق‌تری به دست آورند (Nguyen et al., 2023).

مواد و روش‌ها

داده‌ها: منطقه مورد مطالعه در محدوده جغرافیایی ۵۲ درجه، ۳ دقیقه، ۳۰ ثانیه طول شرقی و ۳۶ درجه، ۳۴ دقیقه، ۶ ثانیه عرض شمالی قرار گرفته است (شکل ۱). محدوده مورد نظر در شبکه مختصات UTM در زون ۳۹ قرار داشته و فاصله آن تا ساحل دریا به کمتر از ۳۰۰ متر می‌رسد. ارتفاع آن از سطح دریاهای آزاد از ۲۰ تا ۴۳ متر است که شیب عمومی آن بین ۳-۰ درصد می‌باشد. وجود دریای مازندران آب و هوای معتدل و مرطوب را در این منطقه به وجود آورده است. حوضه آبریز رودخانه هراز و رودخانه‌های بین هراز و قره‌سو با نام اختصاری هراز- نکا یکی از حوضه‌های باز ایران است که در تقسیم‌بندی حوضه‌های آبریز ایران، حوضه فرعی به‌شمار می‌رود و زیرمجموعه حوضه آبریز دریای مازندران است. مساحت این حوضه، ۱۸،۶۴۴ کیلومتر مربع است و مهم‌ترین رودهای آن بابلرود، هراز و تجن می‌باشند (افشین، ۱۳۷۴). پس از سرازیر شدن رود هراز از دره لار، رواناب‌های دره قاضی‌کلا و جنگل‌دره از کوه‌های شمیم‌کوه، آب زیارو و لاسم از کوه‌های غربی قزقان‌چای فیروزکوه و آب رودهای جاری از دهکده‌های دلارستاق نیز به این رود وارد می‌شوند. شیب رودخانه هراز در محدوده کوهستانی بسیار متغیر است و طول آن ۱۸۵ کیلومتر می‌باشد (ستوده، ۱۳۴۹).

1 Adaboost

2 Bagging



شکل ۱. موقعیت جغرافیایی رودخانه هراز در ایران و استان مازندران

سطح حوضه‌های آبریز را تحت تاثیر قرار می‌دهد. در حوضه‌های با درجه ناهمواری بالا نظیر حوضه آبریز هراز متغیر شیب پراکنش فضایی سکونتگاه‌های انسانی را به شدت متاثر می‌سازد که به نوبه خود می‌تواند باعث افزایش یا کاهش مخاطرات ناشی از فرآیند سیلاب گردد. به این ترتیب متغیر شیب در پهنه‌بندی خطر سیلاب حوضه آبریز هراز مشارکت داده شد. در این رابطه، لایه شیب حوضه از روی تصویر مدل رقومی ارتفاع منطقه تهیه گردید. بین شیب منطقه و سرعت جریان سطحی همبستگی قوی مثبت می‌توان یافت. مناطقی که با کاهش ناگهانی شیب مواجه هستند، احتمال دارد با حجم عظیمی از آب روبه‌رو شوند که موجب سیل‌گیری شدید این نواحی شود. جهت شیب یکی دیگر از پارامترهای اصلی مورفومتریک که با وقوع سیل رابطه معکوس دارد، جهت شیب است. جهت شیب می‌تواند شرایط هیدرولوژیکی منطقه را تحت تاثیر قرار دهد. نقشه جهت شیب منطقه در نرم‌افزار ArcGIS و با استفاده از لایه مدل رقومی ارتفاع در ۹ طبقه تهیه شد. یکی دیگر از پارامترهای اصلی مورفومتریک که با وقوع سیل رابطه معکوس دارد جهت شیب است. جهت شیب می‌تواند شرایط هیدرولوژیکی منطقه را تحت تاثیر قرار دهد. نقشه جهت شیب منطقه در نرم‌افزار ArcGIS و با استفاده از لایه مدل رقومی ارتفاع در ۹ طبقه تهیه شد که به شرح زیر می‌باشد: شمال (۰-۲۲/۵ درجه)، شمال شرقی (۲۲/۵ - ۶۷/۵)، شرقی (۶۷/۵-۱۱۲/۵)، جنوب شرقی (۱۱۲/۵-۱۵۷/۵)، جنوب (۱۵۷/۵ - ۲۰۲/۵)، جنوب غرب (۲۰۲/۵ - ۲۴۷/۵)، غرب

پارامترهای موثر در وقوع سیل: متغیرهای مختلفی بر وقوع سیل در یک حوضه آبریز تاثیر می‌گذارند و تعیین پهنه‌های سیلابی حوضه آبریز فاکتور مهمی در بررسی دشت های سیلابی است. در این رابطه، مهمترین عوامل و فاکتورهای موثر در وقوع سیل شامل ارتفاع، شیب زمین، جهت شیب، میزان بارندگی، شاخص رطوبت توپوگرافیک^۱، شاخص توان رودخانه^۲، شاخص پوشش گیاهی نرمال شده^۳، کاربری اراضی، گروه‌های خاک، فاصله از رودخانه، فاصله از آبراهه، تراکم آبراهه، شاخص قدرت آبراهه و شاخص حمل رسوب هستند (Saha et al., 2022).

ارتفاع، شیب زمین و جهت شیب زمین: ارتفاع نسبت به سطح آزاد دریاها یکی از مهم‌ترین متغیرهایی به‌شمار می‌رود که رفتار هیدرولوژیکی حوضه‌های آبریز و همچنین مخاطرات مرتبط با آنها را تحت تاثیر قرار می‌دهد. به‌منظور ارزیابی ارتفاع در سطح حوضه آبریز هراز از مدل رقومی ارتفاع منطقه استفاده به‌عمل آمد. به‌طور کلی رابطه معکوسی بین خطر وقوع سیل و ارتفاع وجود دارد. فراوانی وقوع سیل با افزایش ارتفاع کاهش می‌یابد، در نتیجه ارتفاعات پست‌تر از حساسیت بیشتری برای وقوع سیل برخوردارند. با توجه به استقرار اغلب سکونتگاه‌های منطقه در قسمت‌های کم‌ارتفاع حوضه و تمرکز فعالیت‌های کشاورزی و سایر فعالیت‌های انسانی در این بخش از حوضه فرآیندهای مرتبط با سیلاب، رودخانه‌ها می‌توانند تبدیل به مخاطراتی جدی شوند. شیب سطح زمین یکی از مهم‌ترین متغیرهایی به‌شمار می‌رود که خطر وقوع سیلاب در

1 Topographic Wetness Index (TWI)

2 Stream Power Index (SPI)

3 Normalized Difference Vegetation Index (NDVI)

(۲۹۲/۵-۲۴۷/۵)، شمال غرب (۳۳۷/۵-۲۹۲/۵) و شمال (۳۳۷/۵-۳۶۰).

بارندگی و بافت خاک: از دیگر فاکتورهای مهم در وقوع

سیل، عامل بارندگی است که در ارزیابی پتانسیل سیل حوضه‌های آبریز دارای اهمیت فراوانی است. بارندگی رابطه معنی داری با دبی رودخانه دارد و به طور مستقیم وقوع سیل را کنترل می‌کند. داده‌های مربوط به متوسط بارندگی سالیانه از سازمان کل هواشناسی استان مازندران اخذ گردید و سپس به منظور کنترل کیفی و آزمون همگنی آنها و رفع نواقص آماری از آزمون Run Test در محیط نرم افزار SPSS استفاده شد. سپس لایه اطلاعاتی آنها در محیط نرم افزار ARCGIS پیاده سازی شد. پس از بررسی دقت روش های مختلف درون یابی در نرم افزار ARC GIS، لایه توزیع میانگین بارندگی سالانه حوزه آبخیز هراز بر اساس روش کریجینگ ساده به دلیل داشتن کمترین مقدار خطای میانگین ریشه مربعات^۱ تهیه شد. خاک‌ها از نظر رنگ، بافت، مواد تشکیل دهنده و دانه بندی انواع مختلفی دارند. عمق خاک‌ها نیز متفاوت است. نوع خاک در یک منطقه نوع گیاهانی که می‌تواند در آن رشد کنند را تعیین می‌کند. پوشش گیاهی نقش مهمی در جلوگیری از فرسایش خاک دارد. جذب آب در خاک با توجه به تخلخل، تراکم و بافت خاک متفاوت است. آب بارش زمانی به سیلاب تبدیل می‌گردد که شدت بارش از شدت نفوذ آب به داخل خاک بیشتر باشد، بنابراین پارامتر نفوذپذیری خاک در کنار پارامترهایی همچون نوع پوشش گیاهی از مهمترین فاکتورهای ایجاد سیلاب در هر حوضه‌ای محسوب می‌گردد. نقشه بافت هیدرولوژیکی خاک به منظور استفاده در مدل‌های یادگیری ماشین تهیه شد و به منظور تهیه لایه‌های اطلاعاتی از محیط نرم افزار ARC GIS استفاده گشت.

شاخص پوشش گیاهی نرمال شده: شاخصی است که

خصوصیات پوشش گیاهی منطقه و تراکم آنها را توصیف می‌کند. این امر خصوصیتی که بر رواناب سطحی و قابلیت نفوذپذیری یک منطقه تاثیر می‌گذارد را مشخص می‌سازد. مناطق دارای تراکم پوشش گیاهی کمتر بیشتر مستعد سیل هستند. در پژوهش حاضر، به منظور تهیه نقشه تراکم پوشش گیاهی منطقه از شاخص تفاضل پوشش گیاهی نرمال شده

استفاده شد. از تصویر سنجنده OLI مربوط به ماهواره LANDSAT 8 در قالب نرم افزار ENVI استفاده شد. این شاخص بین -۱ و +۱ می‌باشد. با افزایش پوشش گیاهی، میزان ارزش حاصل از این نسبت گیری نیز افزایش می‌یابد. مقادیر منفی شاخص پوشش گیاهی نرمال شده (اعداد نزدیک به -۱) نشان دهنده پهنه‌های آبی است. مقادیر نزدیک به صفر (بین -۰/۱ تا +۰/۱) معمولاً نشان دهنده سطوح برهنه سنگی، ماسه‌ای یا برفی هستند. مقادیر پایین و مثبت شاخص (حدود +۰/۲ تا +۰/۴) نشان دهنده پوشش درختچه‌ای و علفزار و مقادیر بالای شاخص (اعداد نزدیک به +۱) نشان دهنده جنگل‌ها می‌باشند. مقادیر شاخص پوشش گیاهی نرمال شده از طیف نور قرمز و مادون قرمز نزدیک به دست می‌آید (Chen et al., 2019). در سنجنده لندست هشت باند ۴ نشان دهنده نور قرمز در طول موج ۰,۶۳۰ تا ۰,۶۸۰ و باندهای ۵ نشان دهنده نور مادون قرمز نزدیک در بازه ۰,۸۴۵ تا ۰,۸۸۵ میکرومتر استفاده می‌شود. لایه پوشش گیاهی در هشت کلاس دریاچه‌ها و خلیج؛ جنگل راش، بلوط و پهن برگ (خزری)؛ مراتع متوسط در کوهستان‌ها (نسبتاً سرد)؛ اراضی کشاورزی آبی؛ اراضی کشاورزی دیم به همراه دیم‌زارهای خزری؛ پوشش گیاهی ویژه دشت‌های شور، تپه‌های شنی و دشت‌های سیلابی؛ جنگل بلوط، ارس و گاه پسته، بادام، گز و کهور؛ و مراتع مرغوب و استپی گاه با درخت‌های پراکنده طبقه بندی شدند.

کاربری اراضی: کاربری اراضی هر منطقه اهمیت زیادی

برای پاسخ‌های هیدرولوژیکی در دوره‌های زمانی مختلف دارد. این امر در تحقیق نشان داد تغییرات در کاربری اراضی می‌تواند احتمال وقوع سیل در منطقه را افزایش دهد (Saha et al., 2022). با استفاده از روش طبقه بندی نظارت شده و الگوریتم حداکثر احتمال در محیط نرم افزار ENVI به هیجده کلاس طبقه بندی شد.

فاصله از آبراه‌ها: در پژوهش حاضر متغیر فاصله از

رودخانه به عنوان یکی از مهم ترین متغیرهای موثر بر خطر سیلاب مورد استفاده قرار گرفت. در این رابطه می‌توان بیان داشت سیلاب‌ها زمانی تبدیل به مخاطره می‌شوند که جامعه انسانی را در معرض تهدید قرار دهند. در واقع، سیلاب‌ها جز دینامیک ذاتی هر رودخانه‌ای به شمار می‌روند و جلوگیری از

کنترل می‌کند. این شاخص کنترل توپوگرافی محلی در فرآیندهای هیدرولوژیکی را کمی می‌سازد. از شاخص رطوبت توپوگرافیک به‌طور گسترده برای توصیف توزیع فضایی و گستره پهنه‌های اشباع و سطوح منبع متغیر جهت تولید رواناب استفاده می‌شود.

شاخص‌های قدرت جریان و حمل رسوب: شاخص

قدرت جریان، قدرت فرآیند فرسایشی ناشی از رواناب سطحی را نشان می‌دهد. پارامتر شاخص حمل رسوب نیز نشان‌دهنده قدرت جریان آب از نظر فرسایش است که بر شرایط هیدرولوژیکی اثر می‌گذارد. شاخص انتقال رسوب عامل مهمی در مدل‌سازی حساسیت به سیل است. این شاخص شکل کلی رواناب را نشان می‌دهد و مناطق با رواناب بیشتر، انتقال رسوب بالاتر و حساسیت کمتری برای سیل دارد.

در مطالعه حاضر، از انواع داده‌های مختلف برای مدل‌سازی حساسیت به سیل استفاده گردید. برای تولید نقشه‌های پوشش زمین و کاربری زمین از سنجنده لندست ۸ (OLI) از ردیف ۱۳۸ و ستون ۴۲ با اندازه پیکسل زمینی ۳۰ متر در تاریخ ۱۹ مارس ۲۰۱۹ از وب سایت رسمی سازمان زمین‌شناسی ایالات متحده (USGS) استفاده شد. لایه‌های هیدرولوژیکی و توپوگرافی با استفاده از ASTER GDEM (نسخه ۲) با اندازه پیکسل زمینی ۳۰ متر استخراج شد. نقشه خاک از داده‌های موجود در سازمان نقشه‌برداری به‌دست آمد، درحالی‌که داده‌های بارندگی از سازمان آب و هواشناسی تهیه شد. در این پژوهش پایگاه داده از ۲۴۰ نقطه سیلابی استفاده شد که از کل مکان‌های سیلابی، ۷۰ درصد را برای آموزش الگوریتم‌های انتخاب شده و ۳۰ درصد را نیز برای اهداف اعتبارسنجی مدل استفاده شد.

روش کار: پژوهش حاضر به لحاظ هدف از نوع تحقیقات

کاربردی است که به شیوه کمی انجام شده است. در این تحقیق با استفاده از الگوریتم‌های یادگیری ماشین ترکیبی مناطق مستعد خطر سیلاب رودخانه هراز مشخص گردید. ابتدا داده‌های هیدرولوژی و توپوگرافی و داده‌های مربوط به نقاط سیلابی و دیتاست‌های مورد نیاز برای ارزیابی معیارهای موثر در سیلاب رودخانه هراز و تصاویر ماهواره‌ای مورد نیاز منطقه جمع‌آوری شد. پس از برازش و تکمیل داده‌های ناقص

رخداده آن امری غیرممکن به‌نظر می‌رسد. بدین ترتیب هنگامی که انسان خود را در معرض این پدیده عادی رودخانه‌ها قرار می‌دهد آنگاه فرآیند سیلاب تبدیل به مخاطره می‌گردد. پرواضح است که نواحی مجاور رودخانه‌ها بیشتر تحت تاثیر سیلاب‌های رودخانه‌ها قرار می‌گیرند و با دور شدن از بستر رودخانه‌ها خطر سیلاب کاهش پیدا می‌کند. این امر ایجاب می‌کند متغیر فاصله از رودخانه در تهیه نقشه مناطق در خطر سیلاب مورد استفاده قرار گیرد. با افزایش فاصله، شیب و ارتفاع بیشتر می‌شود. در نتیجه مناطق دور از کانال رودخانه آسیب‌پذیری کمتری از وقوع سیل دارند، بدین منظور در پژوهش حاضر پس از استخراج شبکه زهکشی حوضه مطالعاتی نسبت به تهیه لایه فاصله از رودخانه اقدام گردید. لازم به توضیح است در تهیه لایه فاصله از رودخانه تنها آبراهه‌های با رده بالاتر از ۳ مورد توجه قرار گرفتند. برای تهیه نقشه فاصله از رودخانه از نقشه رقومی رودخانه در نرم‌افزار ArcGIS استفاده شد و به ۱۰ کلاس از ۲۵ متر تا ۲۵۰ متر تقسیم‌بندی گردید.

تراکم جریان: یکی دیگر از عواملی است که تاثیر مستقیم

بر سیل دارد. می‌توان آن را به صورت کل طول نهرها در یک حوضه زهکشی تقسیم بر مساحت کل حوضه زهکشی بیان کرد. به‌طور کلی، مناطقی که تراکم جریان بالاتری دارند بیشتر در معرض سیل هستند (Nguyen, 2023). نقشه توپوگرافی برای محاسبه و استخراج تراکم رودخانه استفاده شد. به‌منظور تهیه لایه‌های اطلاعاتی از محیط نرم‌افزار ARC GIS استفاده شد.

شاخص رطوبت توپوگرافیک: این پارامتر در سال ۱۹۷۹

توسط بیون و کرکی^۱ معرفی شد که نشان‌دهنده تغییرات فضایی رطوبت در حوضه آبریز است. شاخص رطوبت توپوگرافیک مقدار انباشت آب در هر پیکسل از حوضه آبریز را نشان می‌دهد. به‌طور کلی مقادیر زیاد شاخص رطوبت توپوگرافیک و وقوع سیل همبستگی قوی با یکدیگر دارند و جز شاخص‌هایی به‌شمار می‌روند که از روی تصویر مدل رقومی ارتفاع استخراج می‌شوند. در طی سال‌های اخیر برای ارزیابی خطر وقوع سیلاب در سطح حوضه‌های آبریز مورد استفاده قرار گرفته‌اند. شاخص رطوبت توپوگرافیک بر این فرض استوار است که توپوگرافی حرکت آب را در چشم‌انداز

طبقه‌بند یک قسمت از مجموعه داده^۶ را مشاهده کرده و باید مدل خود را بر اساس همان قسمت از داده‌ها که در اختیارش قرار گرفته است، بسازد. یعنی کل مجموعه داده به هر کدام از طبقه‌بندها داده نمی‌شود. فرض می‌شود مجموعه داده دارای ۷ نمونه است. برای هر کدام از طبقه‌بندها، یک زیرمجموعه از داده‌های اصلی انتخاب می‌شود. انتخاب این زیرمجموعه با جایگزینی خواهد بود. یعنی یک نمونه می‌تواند چند بار هم انتخاب شود. برای مثال به طبقه‌بند شماره ۱ نمونه‌های ۳، ۱، ۱ و ۲ داده شده است. نمونه ۱ دو مرتبه به طبقه‌بند شماره ۱ داده شده است. طبق روال، هر طبقه‌بند با استفاده از داده‌هایی که در اختیار دارد یک مدل ساخته و بقیه‌ی کار انجام می‌شود. تحقیقات نشان داده است روش بگینگ برای الگوریتم‌هایی مانند شبکه‌های عصبی یا درخت‌های تصمیم که با تغییر کم نمونه‌ها ممکن است طبقه‌های مختلفی ایجاد کنند، می‌تواند مفید باشد.

الگوریتم آدابوست، یک روش یادگیری جمعی است و معروف‌ترین الگوریتم از خانواده الگوریتم‌های بوسستینگ^۷ است که توسط فروند، شاپیره ارایه شده است (Freund & Schapire, 1997). آدابوست اختصار کلمه تقویت سازگار^۸ است و همان‌طور که از اسم این تکنیک مشخص است هدف آن تقویت سازگار توان مدل یادگیری ماشین در حل مسئله است. در روش آدابوست در هر تکرار، تمرکز کلاسه‌بند مربوطه بیشتر بر تشخیص نمونه‌هایی است که به اشتباه توسط کلاسه‌بندهای مراحل قبل برچسب‌گذاری شده‌اند. در این روش، اولین کلاسه‌بند با یک بوت‌استرتب از مجموعه آموزشی، آموزش داده می‌شود. سپس کلاسه‌بند مرحله اول با تمام نمونه‌های آموزشی تست می‌شود و تعیین می‌گردد که کدام نمونه‌ها توسط این کلاسه‌بند به درستی قابل تشخیص می‌باشند و کدام نمونه‌ها به اشتباه کلاسه‌بندی شده‌اند. سپس احتمال انتخاب نمونه‌هایی که به اشتباه کلاسه‌بندی شده‌اند، برای نمونه‌گیری مرحله بعد افزایش می‌یابد و احتمال انتخاب نمونه‌هایی که به درستی کلاسه‌بندی شده‌اند، کاهش می‌یابد. بنابراین، کلاسه‌بندهای بعدی به احتمال زیاد با نمونه‌هایی آموزش داده می‌شوند که توسط کلاسه‌بندهای مراحل قبلی به درستی قابل تشخیص نبوده و احتمال آنکه مجدداً با نمونه‌هایی که توسط کلاسه‌بندهای قبلی به درستی قابل تشخیص بوده‌اند

اقدام به تهیه نقشه‌های معیار موثر بر سیلاب منطقه گردید. در تهیه نقشه حساسیت سیلاب در منطقه مورد مطالعه، زاویه شیب، جهت شیب، ارتفاع از سطح دریا، نوع خاک، کاربری اراضی، فاصله از رودخانه، میزان بارندگی سالانه، شاخص پوشش گیاهی نرمال‌شده، شاخص حمل رسوب، شاخص رطوبت توپوگرافیک، شاخص قدرت جریان و تراکم جریان حایز اهمیت بودند (شکل ۴). فرآیند انجام پژوهش به این صورت بود که ابتدا فهرست وقوع سیل‌های گذشته منطقه مورد مطالعه تهیه شد و در ادامه پارامترهای موثر در وقوع سیل شناسایی شد و با استفاده از الگوریتم یادگیری ماشین ترکیبی، نقشه مناطق در خطر وقوع سیل در حوزه آبخیز هراز در شمال ایران تهیه گردید. در یادگیری ماشین، روش کلاسه‌بندی جمعی^۱ از چندین الگوریتم یادگیری اصطلاحاً ضعیف^۲ که مدل‌های پایه هستند و با ترکیب چندین مدل ضعیف، مدل پیچیده‌ای ایجاد می‌شود، بهره گرفته شد. یادگیری ترکیبی در واقع روش یادگیری ماشین با ناظر است که ابتدا یاد می‌گیرد و سپس با آنچه که یاد گرفته است، پیش‌بینی‌هایی را روی داده‌های ناشناخته انجام می‌دهد. به عبارتی یادگیر جمعی ترکیب چند یادگیر ضعیف برای تولید یک یادگیر قوی می‌باشد. یادگیر ضعیف طبقه‌بندی است که فقط کمی از طبقه‌بند تصادفی بهتر است. طبقه‌بندهای ترکیبی (دسته‌بندهای ترکیبی)^۳ از ترکیب چندین طبقه‌بند (کلسیفایر) استفاده می‌کنند. هر طبقه‌بند یک مدل را بر روی داده‌های آموزشی می‌سازد تا به وسیله آن بتواند تفاوت‌ها را در طبقه‌های مختلف درک کند. طبقه‌بند ترکیبی، اما به جای اینکه خود یک مدل بسازد از مدل‌های ساخته شده توسط بقیه طبقه‌بندها استفاده کرده و با یک رای‌گیری، مشخص می‌کند که کدام طبقه را برای نمونه جاری باید برگزیند. دو روش معروف برای ایجاد الگوریتم‌های کلاسه‌بندی جمعی یا الگوریتم‌های طبقه‌بند ترکیبی وجود دارند که عبارت از آدابوست^۴ و بگینگ^۵ هستند.

الگوریتم بگینگ یک تکنیک یادگیری جمعی هست که برای حداقل کردن واریانس مدل استفاده می‌شود. تکنیک بگینگ در سال ۱۹۹۶ توسط برمن مطرح شد (Breiman et al., 1996). در روش بگینگ یک زیرمجموعه از مجموعه داده‌ی اصلی به هر کدام از طبقه‌بندها داده می‌شود. یعنی هر

5 Bagging
6 Data Set
7 Boosting
8 Adaptive Boosting

1 Ensemble
2 weak Learner
3 Ensemble Methods
4 AdaBoost

تشخیص کلاس منفی حایز اهمیت باشد. از متداول‌ترین معیارها که معمولاً در کنار حساسیت بررسی می‌شود، پارامتر ویژگی^۴ است که به آن نرخ پاسخ‌های منفی درست نیز می‌گویند. ویژگی به معنی نسبتی از موارد منفی است که آزمایش آنها را به درستی به‌عنوان نمونه منفی تشخیص داده است (رابطه ۳).

$$\text{Specificity (TNR)} = \text{TN} / (\text{TN} + \text{FP}) \quad (۳)$$

دو پارامتر حساسیت و ویژگی نیز مشابه معیار دقت، معمولاً به صورت درصد بیان می‌شوند. واضح است که پیش‌بینی عالی، پیش‌بینی است که مقادیر حساسیت و ویژگی مربوط به آن، هر دو صد درصد باشند، اما احتمال وقوع این اتفاق در واقعیت بسیار کم است و همیشه یک حداقل خطایی وجود دارد. پارامترهای حساسیت و ویژگی، بنابر ماهیتی که دارند همواره در رقابت با یکدیگر هستند، یعنی افزایش یکی با کاهش دیگری همراه است و برعکس. همین وضعیت منجر به تولید ابزاری دیگر برای ارزیابی کیفیت دسته‌بندی شده است.

معیار دیگر منحنی مشخصه عملکرد سیستم^۵ می‌باشد که عبارت است از منحنی که ارتباط بین دو پارامتر حساسیت و ویژگی را بیان می‌کند. محور عمودی منحنی مشخصه عملکرد سیستم نشان‌دهنده نرخ مثبت صحیح (حساسیت)، و محور افقی نشان‌دهنده مقدار نرخ مثبت غلط (ویژگی) است. نتایج مختلف دسته‌بندی نشانگر نقاط مختلف بر این نمودار هستند و در نهایت یک منحنی را تشکیل می‌دهند. مساحت زیر منحنی^۶، به‌عنوان معیار برای ارزیابی عملکرد دسته‌بند مورد استفاده قرار می‌گیرد. منحنی مشخصه عملکرد سیستم یک معیار کارآیی، بر اساس مقادیر آستانه متغیر برای مسایل طبقه‌بندی است. به بیان ساده‌تر، معیار AUC-ROC، درباره قابلیت مدل در زمینه تشخیص کلاس‌ها به ما اطلاعات می‌دهد. هر چه سطح زیر منحنی بالاتر باشد، مدل بهتر است. در حالت ایده‌آل، مساحت زیر منحنی برابر با بیشترین مقدار خود، یعنی یک است. بنابراین، هر چه مساحت زیر نمودار به عدد یک نزدیک‌تر باشد، به معنای بهتر بودن عملکرد دسته‌بند است. فلوجارت پژوهش در شکل (۲) نشان داده شده است.

و آموزش داده می‌شوند، کاهش می‌یابد (Aydin & Iban, 2023).

ارزیابی الگوریتم یادگیری ماشین با استفاده از ماتریس درهم ریختگی (Confusion Matrix): الگوریتم یادگیری ماشین را می‌توان با تکنیک‌های تخمین دقت که داده‌ها را در یک مجموعه آموزشی و آزمایشی تقسیم می‌کند (معمولاً ۳/۲ مجموعه آموزشی و ۳/۱ تعیین مجموعه آزمایشی) اعتبارسنجی کرد و عملکرد مدل آموزشی را ارزیابی نمود. در بحث دسته‌بندی^۱ یک مجموعه داده با استفاده از روش‌های دسته‌بندی، هدف دستیابی به بالاترین دقت ممکن در دسته‌بندی و تشخیص دسته‌ها است. ماتریس درهم ریختگی، نتایج حاصل از طبقه‌بندی را بر اساس اطلاعات واقعی موجود، نمایش می‌دهد. حال بر اساس این مقادیر می‌توان معیارهای مختلف ارزیابی دسته‌بند و اندازه‌گیری دقت را تعریف کرد. پارامتر دقت^۲، متداول‌ترین، اساسی‌ترین و ساده‌ترین معیار اندازه‌گیری کیفیت یک دسته‌بند است و میزان تشخیص صحیح دسته‌بند در مجموع دو دسته را نشان می‌دهد. این پارامتر در واقع نشان‌گر میزان الگوهایی است که درست تشخیص داده شده‌اند (رابطه ۱).

رابطه (۱) $\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FN} + \text{FP} + \text{TN})$ معیار دیگر، معیار حساسیت^۳ است که آن را «نرخ پاسخ‌های مثبت درست» نیز می‌گویند. حساسیت به معنی نسبتی از موارد مثبت است که آزمایش آنها را به درستی به‌عنوان نمونه مثبت تشخیص داده است. معیاری است که مشخص می‌کند دسته‌بند به چه اندازه در تشخیص تمام مناطق پرخطر موفق بوده است. به کمک این معیار، مناطق کم‌خطری که توسط دسته‌بند به اشتباه به‌عنوان مناطق پرخطر تشخیص داده شده‌اند، هیچ تأثیری در محاسبه این پارامتر ندارد و در واقع زمانی که پژوهشگر از این پارامتر به‌عنوان معیار ارزیابی برای دسته‌بند خود استفاده می‌کند، هدفش دستیابی به نهایت صحت در تشخیص نمونه‌های کلاس مثبت است (رابطه ۲).

$$\text{Sensitivity (TPR)} = \text{TP} / (\text{TP} + \text{FN}) \quad (۲)$$

در نقطه مقابل این معیار، ممکن است در مواقعی صحت

4 Specificity

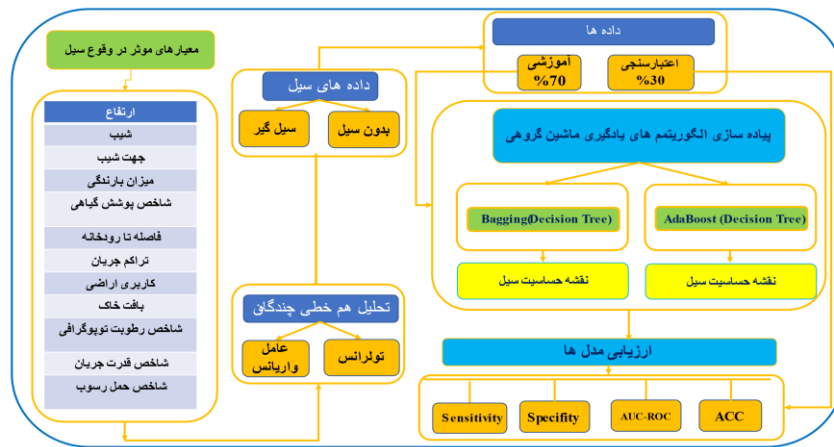
5 Receiver Operating Characteristic Curve (ROC)

6 Area under the curve (AUC)

1 Classification

2 Accuracy

3 Sensitivity

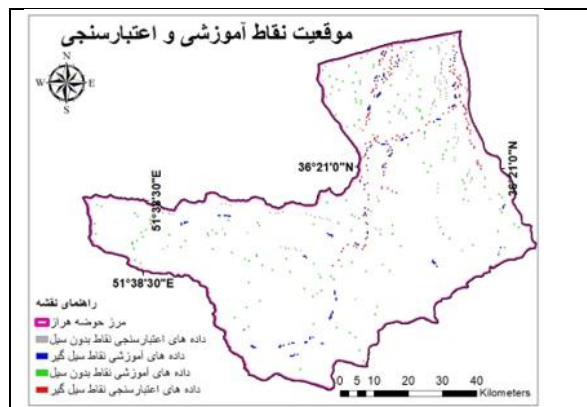


شکل ۲. فلوجارت پژوهش

نتایج

به منظور ارزیابی پتانسیل سیل، تجزیه و تحلیل سیل های ثبت شده ضروری است. صحت وقایع سیلابی گذشته تاثیر بسیار بالایی بر صحت نقشه حساسیت سیلاب دارد (Mia et al., 2023). حوزه آبخیز هراز در دهه های اخیر به شدت تحت تاثیر سیل های خطرناک بوده است. در این مرحله برای ایجاد مجموعه داده های آموزشی و اعتبارسنجی از روش نمونه برداری استفاده شد. با استفاده از دستگاه GPS و بازدیدهای میدانی و بهره گیری از امکانات نرم افزاری

Google Earth و همچنین اطلاعات دبی و بارش شرکت مدیریت منابع آب ایران و سازمان آب منطقه ای مازندران و سازمان هواشناسی کشور یک نقشه از موقعیت ۲۴۰ نقطه سیلابی و غیرسیلابی در حوزه آبخیز هراز تهیه گردید. این ۲۴۰ نقطه به دو گروه تقسیم بندی گردید. ۷۰ درصد از موقعیت نقاط برای آموزش و مدل سازی و ۳۰ درصد موقعیت نقاط که در مدل سازی به کار رفته نشدند برای اعتبارسنجی مورد استفاده قرار گرفتند (شکل ۳).



شکل ۳. موقعیت نقاط آموزشی و اعتبارسنجی

تحلیل هم خطی چندگانه

تست هم خطی بین پارامترها به منظور برآورد همبستگی بین متغیرهای سیل به وسیله دو شاخص ضریب تولرانس^۱ و عامل تورم واریانس^۲ صورت گرفت. اگر مقدار ضرایب تحمل کمتر یا مساوی ۰/۱ باشد و مقدار تورم واریانس بیشتر یا مساوی پنج باشد نشانگر هم خطی پارامترها است (Mia et al., 2023). مقدار ضریب تولرانس بین ۰ و ۱ متغیر است.

هرچه مقدار آن بزرگتر باشد (نزدیک به ۱)، همپوشی با متغیرهای دیگر و در نتیجه هم خطی کمتر است. عامل تورم واریانس با تولرانس نسبت معکوس دارد، یعنی با افزایش مقدار تولرانس، عامل تورم واریانس کاهش می یابد. متغیرهایی که ممکن است همبستگی داخلی زیادی داشته باشند باید به منظور بهبود عملکرد مدل های پیش بینی کننده از فرآیند تحلیل حذف شوند. بنابراین در این بخش از پژوهش تجزیه و

1 Tolerance (TOL)

2 Variance Inflation Factor (VIF)

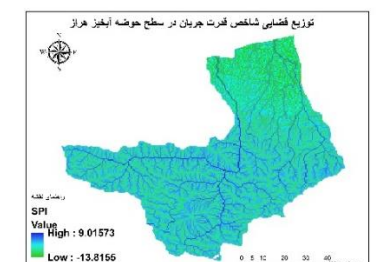
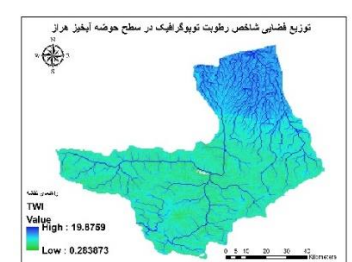
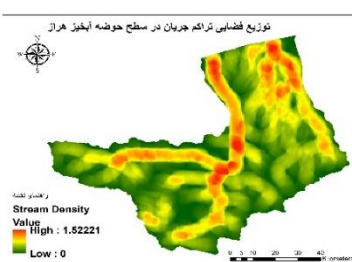
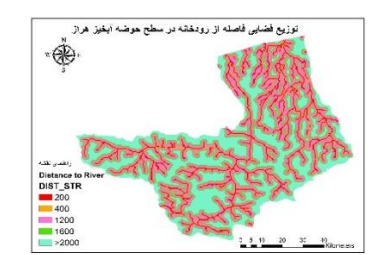
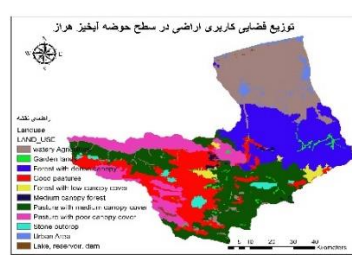
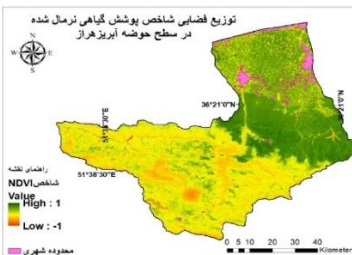
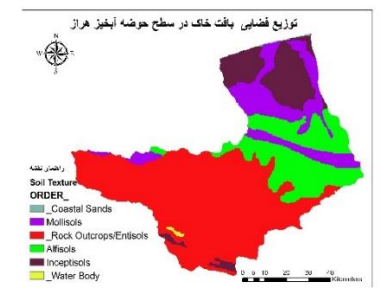
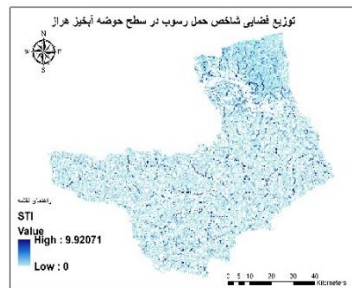
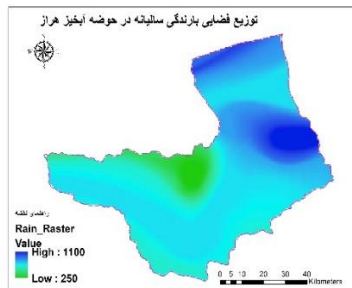
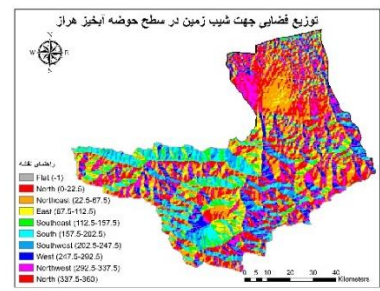
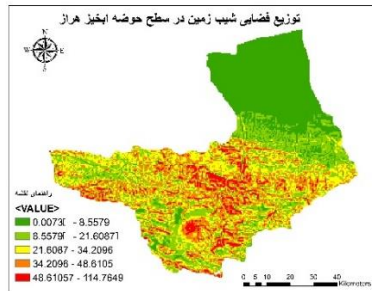
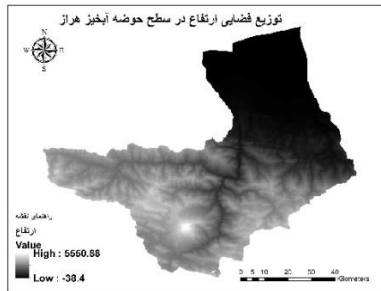
ارزیابی مکانی حساسیت سیل‌گیری با استفاده از روش‌های تقویت تطبیقی و بسته‌بندی در یادگیری ماشین/۱۰۹

تحلیل هم‌خطی چندگانه انجام گرفت (جدول ۱). برای هر

یک از پارامترها مقدار عامل تورم واریانس نیز محاسبه شد که

جدول ۱. نتایج تحلیل هم‌خطی چندگانه

پارامترهای موثر در خطر وقوع سیل	همبستگی با دیگر پارامترها	ضریب تولرانس	عامل تورم واریانس
ارتفاع	۰/۶۵	۰/۳۵	۲/۸۵
شیب	۰/۷۵	۰/۲۵	۴
جهت شیب	۰/۴	۱/۶۷	۰/۶
متوسط بارندگی	۰/۵۵	۰/۴۵	۲/۲
شاخص پوشش گیاهی	۰/۶	۰/۴	۲/۵
کاربری اراضی	۰/۵	۰/۵	۲
فاصله از آبراهه	۰/۷	۰/۳	۳/۳۳
گروه‌های خاک	۰/۳	۰/۷	۱/۴۳
تراکم جریان	۰/۲	۰/۵	۱/۳۲
شاخص رطوبت توپوگرافیک (TWI)	۰/۸	۰/۲۵	۴
شاخص قدرت جریان (SPI)	۰/۴۵	۰/۵۵	۱/۸۲
شاخص حمل رسوب (STI)	۰/۵۵	۰/۴۵	۲/۲



شکل ۴. توزیع فضایی معیارهای موثر در تهیه نقشه حساسیت سیل در سطح حوضه آبریز هراز

پیاده‌سازی الگوریتم‌های یادگیری ماشین ترکیبی

مرحله اولیه مستلزم آماده‌سازی دقیق داده‌ها برای تجزیه و تحلیل بعدی می‌باشد. این امر شامل مجموعه‌ای از متغیرهای موثر در وقوع سیل است. قبل از شروع داده‌کاوی در الگوریتم‌های یادگیری ماشین، کاربر باید مطمئن شود که همه مجموعه داده‌ها دارای سیستم مختصات یکسانی برای لایه‌های رستری و برداری در محیط نرم‌افزاری ArcGIS هستند. یک فرآیند ارزیابی کیفیت داده به‌ویژه در قالب بسته GeoPandas، برای تسهیل تجزیه و تحلیل اعمال شد. متعاقباً داده‌های سیستم اطلاعات جغرافیایی آماده شده به برنامه پایتون وارد شد و در نتیجه امکان استفاده از آن در الگوریتم‌های یادگیری ماشین فراهم شد.

انتخاب ویژگی‌های گام بعدی برای مدل‌سازی حساسیت به سیل می‌باشد. برای دستیابی به این هدف از تحلیل همبستگی استفاده گردید. از قابلیت‌های کتابخانه scikit-learn برای پیاده‌سازی و ارزیابی یکپارچه الگوریتم‌های یادگیری ماشین و سایر راهبردهای مجموعه‌ای که برای مدل‌سازی حساسیت به سیل در پایتون طراحی شده‌اند، استفاده گردید.

در الگوریتم بگینگ در ابتدا نوع و تعداد مدل‌های پایه مشخص شد، سپس با کمک رویکرد بوت‌استرپ^۱ برای هر مدل پایه یک داده آموزشی از روی داده آموزشی انتخاب و آموزش داده شد. مفهوم بوت‌استرپ آن است که داده‌ها به صورت تصادفی از روی داده آموزشی انتخاب شوند، منتهی در این داده انتخاب شده ممکن است یک نمونه آموزشی چندین بار انتخاب شود. در الگوریتم بگینگ، داده جدید به تک‌تک مدل‌های پایه آموزش دیده ارائه می‌شود، تک‌تک مدل‌ها بر اساس دانشی که دارند، یک خروجی‌ای برای داده جدید تخمین می‌زنند. در نهایت نظر همه مدل‌ها با هم ترکیب می‌شوند تا خروجی داده جدید را تخمین زنند. جهت طبقه‌بندی، یک رای‌گیری ساده بین مدل‌ها اتفاق می‌افتد و داده جدید به کلاسی تعلق دارد که بیشترین رای را نسبت به آن کلاس در مقایسه با سایر کلاس‌ها داشته باشد. به عبارتی الگوریتم بگینگ مستلزم آموزش مدل‌های متعدد بر

زیرمجموعه‌های تصادفی داده‌های آموزشی است. در این چارچوب، از درخت تصمیم به‌عنوان طبقه‌بندی‌کننده پایه استفاده شد. نتایج به‌دست‌آمده از روش‌شناسی متعاقباً در محیط ArcGIS برای تولید نقشه‌های حساسیت به سیل گنجانده شد.

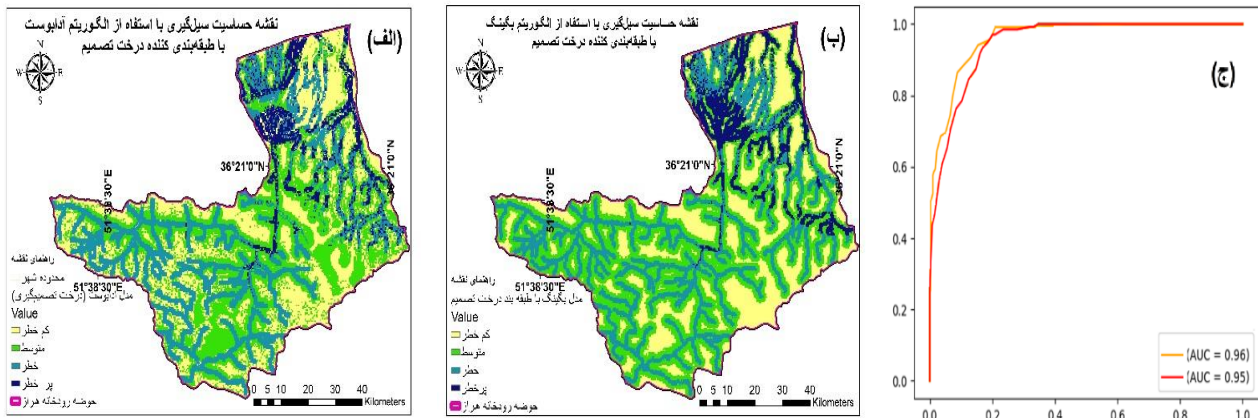
در الگوریتم آدابوسست به‌طور متوالی یک طبقه‌بندی ضعیف برای نسخه‌های اصلاح‌شده داده‌ها اعمال گردید. با افزایش وزن مشاهدات طبقه‌بندی نشده، هر یادگیرنده ضعیف بر خطای قبلی تمرکز می‌یابد. پیش‌بینی‌ها از طریق رای اکثریت وزنی جمع‌آوری شد. در این قسمت الگوریتم‌های یادگیری ماشین ترکیبی بر اساس مجموعه داده‌های آموزشی و ارزیابی عملکرد این مدل براساس مجموعه داده‌های اعتبارسنجی در بسته نرم‌افزاری پایتون^۲ انجام گرفت. برای اجرای الگوریتم آدابوسست از درخت تصمیم‌گیری CART به‌عنوان طبقه‌بندی‌کننده پایه^۳ انتخاب شد. به‌منظور انتقال لایه‌های اطلاعاتی از محیط نرم‌افزار ARCGIS به محیط نرم‌افزار پایتون و اجرای مراحل الگوریتم در این نرم‌افزار، ابتدا این لایه‌ها به فرمت ASCII تبدیل شد و سپس با استفاده از نرم‌افزار SPSS به فرمت جدول عددی (CSV) درآمد. این جدول عددی داده وارد نرم‌افزار پایتون گشت و به‌عنوان داده اولیه در محیط نرم‌افزاری پایتون با استفاده از الگوریتم‌های یادگیری ماشین پردازش شد.

نقشه‌های حساسیت به سیل‌گیری با استفاده از روش‌های یادگیری ماشینی مورد استفاده در این مطالعه، به چهار کلاس مجزا، با حساسیت کم، متوسط، زیاد و بسیار زیاد طبقه‌بندی شدند (شکل ۵). عمدتاً مناطق در دسته‌های حساسیت کم و متوسط قرار گرفتند، در حالی‌که مناطق با حساسیت بالا و بسیار بالا عمدتاً در بخش شمالی حوضه هراز متمرکز شده‌اند. علاوه بر این، بخشی از مناطق مستعد سیل در بخش شرقی منطقه مورد مطالعه مشاهده شد. جدول (۲) خلاصه‌ای جامع از نتایج را ارائه می‌دهد که مناطق مربوط به هر کلاس حساسیت را بر حسب کیلومتر مربع به صورت درصدی از کل مساحت مشخص می‌کند.

1 Bootstrap (the random sampling with replacement)

2 python

3 Base Classifier



شکل ۵. نقشه حساسیت سیلاب با استفاده از الگوریتم یادگیری ماشین ترکیبی (آدابوست) با استفاده از طبقه‌بندی‌کننده درخت تصمیم در سطح حوضه آبریز هراز، (ب) نقشه حساسیت سیلاب با استفاده از الگوریتم یادگیری ماشین ترکیبی (بگینگ) با استفاده از طبقه‌بندی‌کننده درخت تصمیم در سطح حوضه آبریز هراز و (ج) منحنی مشخصه عملکرد بر اساس داده‌های آموزشی و اعتبارسنجی جهت اعتبارسنجی مدل آدابوست با رنگ نارنجی و بگینگ با رنگ قرمز با استفاده از طبقه‌بندی درخت تصمیم

جدول ۲. مساحت هر یک از طبقات در ارتباط با حساسیت سیلاب

الگوریتم	مساحت	طبقه‌بندی کلاس‌ها		
		کم خطر	متوسط	خطر
آداوست با طبقه‌بند درخت تصمیم	کیلومترمربع	۸۹۶۸/۱۵۳۸	۵۷۱۸/۱۴۲۳	۸۸۷۷/۱۲۵۸
	درصد	۳۳/۳۶	۳۰/۸۶	۲۷/۲۹
بگینگ با طبقه‌بند درخت تصمیم	کیلومترمربع	۲۲۲/۱۳۵۶	۷۸۵۳/۱۶۰۵	۵۱۰۱/۱۰۹۶
	درصد	۲۹/۴۰	۳۴/۸۱	۲۳/۷۷

شد. در مطالعه دیگر مدل‌های بگینگ و آنتروپی شانون توسط شهابی (۱۴۰۰) استفاده شد و نشان داد مدل بگینگ با دقت بالاتری (۰/۹۶) نسبت به مدل آنتروپی شانون (۰/۸۸) است. جدول (۳)، نتایج حاصل از تجزیه و تحلیل مقایسه‌ای بین مطالعه حاضر و ۴ مطالعه مرتبط اخیر را ارائه می‌کند. این مطالعه با در نظر گرفتن ۱۲ عامل مهم در مدل‌سازی با درصد بالایی با مقالات مورد مقایسه مطابقت دارد. قابل ذکر است، مطالعه حاضر به AUC چشمگیر ۰/۹۶ دست یافت که در بین ۴ مطالعه برتر در این زمینه رتبه‌بندی شد. علاوه بر این، اجرای روش ترکیبی AdaBoost-Decision Tree یک دستاورد پیشگام در مطالعات اخیر است که باید مد نظر قرار گیرد.

بحث و نتیجه‌گیری

سیل‌ها تهدیدی اساسی برای مناطق شهری و اکوسیستم‌ها به‌شمار می‌روند و در ردیف خطرناک‌ترین بلایای طبیعی قرار می‌گیرند. حوضه آبریز هراز به دلیل موقعیت توپوگرافی خاص خود و دریافت بارش مناسب در طول سال و به‌خصوص در فصل بهار، از حوضه‌های با حساسیت سیل‌خیزی زیاد محسوب می‌شود. مطالعاتی بر روی حوزه آبخیز هراز با روش‌های مختلف انجام شده است. نوهانی و همکاران (۱۳۹۵) از روش آنتروپی شانون برای تخمین حساسیت به سیلاب در حوزه هراز استفاده کردند. نتایج ایشان نشان داد روش آنتروپی شانون با دستیابی به دقت ۹۱/۴۲ درصد انجام

جدول ۳. تجزیه و تحلیل مقایسه‌ای مطالعات در تهیه نقشه‌های حساسیت سیل‌گیری

شماره	مطالعات	تعداد معیارهای موثر تعداد نقاط		تأثیرگذارترین متغیرها	تعداد مدل‌ها	مدل بهتر	بهرترین AUC به‌دست آمده
		در سیل	سیلابی				
۰	مطالعه حاضر	۱۲	۲۴۰	ارتفاع، فاصله از رودخانه، شیب زمین	۲	AdaBoost- Decision Tree Ensemble Learning	۰/۹۶
۱	Anand & Pradhan, 2023	۱۵	۱۸۹	ارتفاع، شیب زمین، فاصله از رودخانه	۶	Evidential Belief Function (EBF)	۰/۹۱
۲	Costache et al., 2023	۱۰	۱۵۸	ارتفاع، شیب زمین، فاصله از رودخانه، بارندگی	۶	Iterative Classifier Optimizer	۰/۹۵
۳	Mia et al., 2023	۱۶	۳۵۰	بارندگی، ارتفاع، فاصله از رودخانه	۴	ADT-ICO, NB-ICO, ANN-ICO, and DLNN-ICO	۰/۹۳
۴	Saikh et al., 2023	۱۴	۲۰۰	ارتفاع، شیب زمین، بارندگی، پوشش گیاهی، فاصله از رودخانه	۷	Reduced Error Pruning Tree (REPTree)	۰/۹۲

شهری است تا به طور موثر تاثیر سیل را در مناطق آسیب پذیر کاهش دهند.

سپاسگزاری و قدردانی

نویسندگان مایل هستند نهایت قدردانی خود را از آقای مهندس علی اکبری، دانشجوی دکتری رشته الکترونیک دانشگاه تهران برای حمایت و کمک وی در تکمیل فرآیند پیاده سازی الگوریتم های یادگیری ماشین ترکیبی ابراز نمایند.

منابع

افشین، ی.ا. (۱۳۷۴) رودخانه های ایران، جلد اول، تهران: انتشارات وزارت نیرو، شرکت مهندسين مشاور جاماب، صفحات: ۵۳۴-۵۴۰.

ستوده، م. (۱۳۴۹) از آستارا تا استرآباد، جلد چهارم، چاپ اول، تهران: انتشارات انجمن آثار و مفاخر، ۱۱۰۴ صفحه.

شهابی، ه. (۱۴۰۰) پهنه بندی حساسیت وقوع سیل در مناطق شمالی ایران با استفاده از الگوریتم های پیشرفته داده کاوی، منطقه مورد مطالعه حوزه آبخیز هراز. نشریه برنامه ریزی منطقه ای، ۱۱(۴۱): ۱۶۵-۱۸۲.

نوهانی، ا. دارابی، ف. معروفی نیا، ا. و خسروی، خ. (۱۳۹۵) ارزیابی مدل آنتروپی شانون در تهیه نقشه حساسیت و احتمال به وقوع سیل در حوزه آبخیز هراز. مخاطرات محیط طبیعی، ۵(۱۰): ۹۹-۱۱۶.

Abdullah, M.F., Siraj, S. and Hodgett, R.E. (2021) An overview of multi-criteria decision analysis (MCDA) application in managing water-related disaster events: Analyzing 20 years of literature for flood and drought events. *Water*, 13(10): 1358. Retrieved from <https://doi.org/10.3390/w13101358/>

Akay, H. (2021) Flood hazards susceptibility mapping using statistical, fuzzy logic, and MCDM methods. *Soft Computing*, 25(14): 9325-9346. Retrieved from <https://doi.org/10.1007/s00500-021-05903-1/>

Ali, S.A., Khatun, R., Ahmad, A. and Ahmad, S.N. (2019) Application of GIS-based analytic hierarchy process and frequency ratio model to flood vulnerable mapping and risk area estimation at Sundarban region, India. *Modeling Earth Systems and Environment*, 5(3): 1083-1102. Retrieved from <https://doi.org/10.1007/s40808-019-00593-z/>

Anand, A.K. and Pradhan, S.P. (2023) Evaluation of bivariate statistical and hybrid models for the preparation of flood hazard susceptibility maps in the Brahmani River Basin, India.

پژوهش حاضر تلاشی به منظور بررسی حساسیت خطر وقوع سیل در این حوضه بوده است. بنابراین از ۲۴۰ نقطه مشاهداتی و میدانی در ارتباط با مناطق سیل گیر و بدون سیل و همچنین پارامترهای موثر در وقوع سیل استفاده شد. از الگوریتم های یادگیری ماشین ترکیبی مانند آدابوست و بگینگ با استفاده از طبقه بندی کننده درخت تصمیم جهت تهیه نقشه های حساسیت به سیل در منطقه مازندران ایران بهره گرفته شد. تجزیه و تحلیل طیف وسیعی از پارامترها شامل ارتفاع، شیب، جهت شیب، کاربری اراضی، نوع خاک، بارندگی، شاخص تفاوت نرمال شده پوشش گیاهی، فاصله تا رودخانه، شاخص رطوبت توپوگرافی، شاخص توان جریان و شاخص انتقال رسوب را صورت پذیرفت و از تکنیک های ارزیابی دقت آماری دقیق از جمله دقت، حساسیت، ویژگی و منحنی مشخصه عملکرد سیستم استفاده شد. یافته ها نشان داد الگوریتم آدابوست با استفاده از طبقه بند درخت تصمیم بالاترین دقت را در بین مدل ها دارد.

پیشنهادها

با توجه به نتایج و چشم انداز تهیه نقشه حساسیت سیل، تلاش های تحقیقاتی آینده بر چندین راه امیدوارکننده متمرکز خواهد شد. اولاً، پالایش دقت مدل های حساسیت به سیل یک هدف اصلی باقی می ماند. هدف این است که از پتانسیل تکنیک های یادگیری ماشینی در حال ظهور استفاده شود و به طور بالقوه الگوریتم های یادگیری عمیق پیشرفته یکپارچه شود تا قابلیت های پیش بینی بیشتر گردد. علاوه بر این، ترکیب داده ها و استفاده از فن آوری های سنسج از راه دور، توسعه مدل های پویا و پاسخگو را قادر می سازد که با شرایط متغیر محیطی سازگار شوند. علاوه بر این، گسترش دامنه تحقیقات به مناطق جغرافیایی وسیع تر و مناطق آب و هوایی متنوع، درک جامع تری از الگوهای حساسیت به سیل را تسهیل می کند. تلاش های مشترک برای ایجاد پروتکل ها و معیارهای استاندارد شده جهت تهیه نقشه های حساسیت سیل گیری به مقایسه پذیری مطالعات آینده کمک می کند و رویکرد یکپارچه تر را در جامعه تحقیقاتی تقویت می کند. در نهایت، تحقیق حاضر نه تنها به دنبال ارتقای دقت و قابلیت اطمینان تهیه نقشه های حساسیت سیل گیری است، بلکه به دنبال ارایه بینش ها و ابزارهای عملی برای سیاست گذاران و برنامه ریزان

- and System Sciences, 55(1): 119-139. Retrieved from <https://doi.org/10.1006/jcss.1997.1504/>
- Gudiyangada Nachappa, T., Tavakkoli Piralilou, S., Gholamnia, K., Ghorbanzadeh, O., Rahmati, O. and Blaschke, T. (2020) Flood susceptibility mapping with machine learning, multi-criteria decision analysis and ensemble using Dempster Shafer Theory. *Journal of Hydrology*, 590(125275): 125275. Retrieved from <https://doi.org/10.1016/j.jhydrol.2020.125275/>
- Hasanuzzaman, M., Islam, A., Bera, B. and Shit, P.K. (2022) A comparison of performance measures of three machine learning algorithms for flood susceptibility mapping of river Silabati (tropical river, India). *Physics and Chemistry of the Earth* (2002), 127(103198): 103198. Retrieved from <https://doi.org/10.1016/j.pce.2022.103198/>
- Khosravi, K., Shahabi, H., Pham, B.T., Adamowski, J., Shirzadi, A., Pradhan, B., Dou, J., Ly, H.-B., Gróf, G., Ho, H. L., Hong, H., Chapi, K. and Prakash, I. (2019) A comparative assessment of flood susceptibility modeling using Multi-Criteria Decision-Making Analysis and Machine Learning Methods. *Journal of Hydrology*, 573: 311–323. Retrieved from <https://doi.org/10.1016/j.jhydrol.2019.03.073/>
- Mia, M.U., Chowdhury, T.N., Chakraborty, R., Pal, S.C., Al-Sadoon, M.K., Costache, R. and Islam, A.R.M.T. (2023) Flood susceptibility modeling using an advanced deep learning-based iterative classifier optimizer. *Land*, 12(4): 810-810. Retrieved from <https://doi.org/10.3390/land12040810/>
- Nguyen, H.D. (2023) Spatial modeling of flood hazard using machine learning and GIS in Ha Tinh province, Vietnam. *Journal of Water and Climate Change*, 14(1): 200–222. Retrieved from <https://doi.org/10.2166/wcc.2022.257/>
- Pham, B.T., Jaafari, A., Van Phong, T., Yen, H.P.H., Tuyen, T.T., Van Luong, V., Nguyen, H.D., Van Le, H. and Foong, L.K. (2021) Improved flood susceptibility mapping using a best first decision tree integrated with ensemble learning techniques. *Geoscience Frontiers*, 12(3): 101105. Retrieved from <https://doi.org/10.1016/j.gsf.2020.11.003/>
- Plataridis, K. and Mallios, Z. (2023) Flood susceptibility mapping using hybrid models optimized with Artificial Bee Colony. *Journal of Hydrology*, 624(129961): 129961. Retrieved from <https://doi.org/10.1016/j.jhydrol.2023.129961/>
- Rafiei-Sardooi, E., Azareh, A., Choubin, B., Mosavi, A.H. and Clague, J.J. (2021) Evaluating urban flood risk using hybrid method of TOPSIS and machine learning. *Environmental Earth Sciences*, 82(16): 395–408. Retrieved from <https://doi.org/10.1007/s12665-023-11069-w/>
- Avand, M., Janizadeh, S., Naghibi, S.A., Pourghasemi, H.R., Khosrobeigi Bozchaloei, S. and Blaschke, T. (2019) A comparative assessment of random forest and k-nearest neighbor classifiers for gully erosion susceptibility mapping. *Water*, 11(10): 2076. Retrieved from <https://doi.org/10.3390/w11102076/>
- Aydin, H.E. and Iban, M.C. (2023) Predicting and analyzing flood susceptibility using boosting-based ensemble machine learning algorithms with SHapley Additive exPlanations. *Natural Hazards (Dordrecht, Netherlands)*, 116(3): 2957–2991. Retrieved from <https://doi.org/10.1007/s11069-022-05793-y/>
- Bera, S., Das, A. and Mazumder, T. (2022) Evaluation of machine learning, information theory and multi-criteria decision analysis methods for flood susceptibility mapping under varying spatial scale of analyses. *Remote Sensing Applications Society and Environment*, 25(100686): 100686. Retrieved from <https://doi.org/10.1016/j.rsase.2021.100686/>
- Breiman, L. (1996) Bagging predictors. *Machine Learning*, 24(2): 123–140. <https://doi.org/10.1007/bf00058655/>
- Bui, Q.D., Luu, C., Mai, S.H., Ha, H.T., Ta, H.T. and Pham, B.T. (2023) Flood risk mapping and analysis using an integrated framework of machine learning models and analytic hierarchy process. *Risk analysis: An official publication of the Society for Risk Analysis*, 43(7): 1478–1495. Doi: 10.1111/risa.14018/
- Chen, J., Li, Q., Wang, H. and Deng, M. (2019) A machine learning ensemble approach based on random forest and radial basis function neural network for risk evaluation of regional flood disaster: A case study of the Yangtze River Delta, China. *International Journal of Environmental Research and Public Health*, 17(1): 49-49. Retrieved from <https://doi.org/10.3390/ijerph17010049/>
- Costache, R., Abdo, H.G., Pratap Mishra, A., Pal, S.C., Islam, A.R.M.T., Pande, C.B. and Albanai, J.A. (2023) Using fuzzy and machine learning iterative optimized models to generate the flood susceptibility maps: case study of Prahova River basin, Romania. *Geomatics Natural Hazards and Risk*, 14(1); 2281241. Doi: 10.1080/19475705.2023.2281241/
- Freund, Y. and Schapire, R. (1997) A Decision-Theoretic generalization of on-line learning and an application to boosting. *Journal of Computer*

- <https://doi.org/10.1080/19475705.2018.1506509/>
- Vajda, V., McLoughlin, S., Mays, C., Frank, T.D., Fielding, C.R., Tevyaw, A., Lehsten, V., Bocking, M. and Nicoll, R.S. (2020) End-Permian (252 Mya) deforestation, wildfires and flooding—An ancient biotic crisis with lessons for the present. *Earth and Planetary Science Letters*, 529(115875): 115875. Retrieved from <https://doi.org/10.1016/j.epsl.2019.115875/>
- Wang, Y., Fang, Z., Hong, H. and Peng, L. (2020) Flood susceptibility mapping using convolutional neural network frameworks. *Journal of Hydrology*, 582(124482): 124482. Retrieved from <https://doi.org/10.1016/j.jhydrol.2019.124482/>
- Wijayarathne, D.B. and Coulibaly, P. (2020) Identification of hydrological models for operational flood forecasting in St. John's, Newfoundland, Canada. *Journal of Hydrology, Regional Studies*, 27(100646): 100646. Retrieved from <https://doi.org/10.1016/j.ejrh.2019.100646/>
- Yang, Y.C.E., Ray, P.A., Brown, C.M., Khalil, A.F. and Yu, W.H. (2015). Estimation of flood damage functions for river basin planning: a case study in Bangladesh. *Natural Hazards (Dordrecht, Netherlands)*, 75(3): 2773–2791. Retrieved from <https://doi.org/10.1007/s11069-014-1459-y/>
- Yari, M., Soltani-Gerdefaramarzi, S. and Ghasemi, M. (2019) Investigation of the effect of to land use changes on flood hydrograph and fluctuations of groundwater level in part of Ghareh Souh Basin. *Journal of Geography and Environmental Hazards*, 8(3): 41-58. Doi: 10.22067/GEO.V8I3.80976/
- International Journal of Disaster Risk Reduction: IJDRR*, 66(102614): 102614. Retrieved from <https://doi.org/10.1016/j.ijdr.2021.102614/>
- Saha, S., Gayen, A. and Bayen, B. (2022) Deep learning algorithms to develop Flood susceptibility map in Data-Scarce and Ungauged River Basin in India. *Stochastic Environmental Research and Risk Assessment: Research Journal*, 36(10): 3295–3310. Retrieved from <https://doi.org/10.1007/s00477-022-02195-1/>
- Sahana, M. and Patel, P.P. (2019) A comparison of frequency ratio and fuzzy logic models for flood susceptibility assessment of the lower Kosi River Basin in India. *Environmental Earth Sciences*, 78(10): 1-27. Retrieved from <https://doi.org/10.1007/s12665-019-8285-1/>
- Saikh, N.I. and Mondal, P. (2023) GIS-based machine learning algorithm for flood susceptibility analysis in the Pagla river basin, Eastern India. *Natural Hazards Research*, 3(3): 420–436. Retrieved from <https://doi.org/10.1016/j.nhres.2023.05.004/>
- Seydi, S.T., Kanani-Sadat, Y., Hasanlou, M., Sahraei, R., Chanussot, J. and Amani, M. (2022) Comparison of machine learning algorithms for Flood Susceptibility Mapping. *Remote Sensing*, 15(1): 192-192. Retrieved from <https://doi.org/10.3390/rs15010192/>
- Shafapour Tehrany, M., Kumar, L., Neamah Jebur, M. and Shabani, F. (2019) Evaluating the application of the statistical index method in flood susceptibility mapping and its comparison with frequency ratio and logistic regression methods. *Geomatics Natural Hazards and Risk*, 10(1): 79–101. Retrieved from

Spatial flood susceptibility assessment using boosting and bagging in machine learning techniques

Maryam Jahanbani¹, Hossein Aghamohammadi^{2*}, Mohammad H. Vahidnia³ and Zahra Azizi⁴

- 1) PhD Student, Department of Remote Sensing and GIS, Faculty of Natural Resources and Environment, Science and Research Branch, Islamic Azad University, Tehran, Iran.
- 2) Assistant Professor, Department of Remote Sensing and GIS, Faculty of Natural Resources and Environment, Science and Research Branch, Islamic Azad University, Tehran, Iran. *Corresponding Author Email Address: aghamohammadi@srbiau.ac.ir
- 3) Assistant Professor, Center for Remote Sensing and GIS Research, Faculty of Earth Sciences, Shahid Beheshti University, Tehran, Iran
- 4) Assistant Professor, Department of Remote Sensing and GIS, Faculty of Natural Resources and Environment, Science and Research Branch, Islamic Azad University, Tehran, Iran.

Date of Submission: 2023/11/20

Date of Acceptance: 2024/02/28

Abstract

Floods cause billions of dollars in damage to countries every year and threaten people's livelihoods. This has caused significant socio-economic threats to the world's population, which must be controlled and contained. In this regard, machine learning algorithms along with geographic information systems are primary tools that are effective in flood control modeling and analysis. The Purpose of this research is to identify a part of flood-sensitive regions across the Heraz catchment area in Mazandaran province using ensemble methods in machine learning algorithms. The research process is as follows: first, the data of flood points were prepared. Next, 70% of approximately 240 sample positions were used for modeling and map preparation. The remaining 30%, which were randomly selected, were used to validate the produced maps. Then, the effective factors, including slope angle, slope direction, topography, soil type, land cover, distance from the river, annual rainfall, normalized difference vegetation index, index of sediment transmittance, index of topographic wetness, and index of stream density have been used to weight the impact of each factor using machine learning algorithms. Based on the results of this study, the system performance characteristic curve (ROC) was drawn, and the area under the curve (AUC) was calculated to validate the flood-prone area map. Findings demonstrated that the Adaptive Boosting model is more accurate than the Bagging model in preparing a flood sensitivity map. Predictive susceptibility mapping plays a pivotal role in enabling urban planners and managers to mitigate and safeguard proactively against the adverse consequences of flooding. Flood management authorities in the Ministry of Energy can employ the proposed ensemble model to assist disaster management and mitigate hazards in future studies.

Keywords: Adaptive Boosting algorithm, Bagging algorithm, Ensemble machine learning models, Flood, Spatial information system.