



مرکز تحقیقات آموزش کشاورزی و منابع طبیعی فارس

پژوهش‌های آب‌نخرداری

شاپا: ۲۰۳۸-۲۹۸۱



مراکز تحقیقات آموزش و پژوهش کشاورزی

پیش‌بینی مکانی استعداد سیل‌گیری آبخیز زرينه‌رود با استفاده از مدل یادگیری ماشینی بیز ساده

فرنوش محمدی^{۱*}، عارف بهمنی^۲

۱- دانشجوی دکتری تخصصی، گروه احیاء مناطق خشک و کوهستانی، دانشکده منابع طبیعی دانشگاه تهران، کرج، ایران

۲- معاون آبخیزداری اداره کل منابع طبیعی و آبخیزداری استان کردستان، سنندج، ایران

چکیده مبسوط

مقدمه و هدف

سیلاب‌ها هرساله سبب خسارت‌های مالی و تلفات جانی پرشماری بوده و تأثیر نابودگری بر توسعه پایدار کشور دارند. با این وجود، عموماً رخدادهای واقعی سیل‌گیری در تحلیل‌های مکانی و مدل‌سازی استفاده‌نشده و نقشه‌های استعداد سیل‌گیری صرفاً براساس دیدگاه کارشناسی و روش‌های تصمیم‌گیری چندمعیاره تهیه شده‌اند. از سوی دیگر، برنامه‌های مدیریتی و اقدامات دستگاه‌های اجرایی معمولاً بدون توجه به پهنه‌های مختلف استعداد سیل‌گیری آبخیزها تدوین می‌شوند. این پژوهش، با هدف بهره‌گیری از داده‌های مشاهده‌ای رخدادهای سیل‌گیری و تعیین کارایی مدل نابو بیز در زمینه پیش‌بینی مکانی استعداد سیل‌گیری در آبخیز زرينه‌رود انجام شد.

مواد و روش‌ها

پایگاه داده رخدادهای سیل‌گیری، براساس اطلاعات رخدادهای سیل در آبخیز زرينه‌رود که به‌وسیله دفتر مدیریت بحران استانداری و شرکت آب منطقه‌ای ثبت‌شده بود، تهیه شد. با توجه به اینکه عامل‌های محیطی مختلفی در شکل‌گیری سیل و آب‌گرفتگی زمین مجاور رودها نقش دارند، بدون درنظر گرفتن آن‌ها، مدل‌سازی سیل امکان‌پذیر نیست. از این‌رو، بعد از بررسی منابع مختلف، سیزده عامل محیطی مؤثر بر سیل‌گیری شامل بلندی زمین، جهت دامنه، تراکم زهکشی، کاربری زمین، سنگ‌شناسی، انحنای سطح، انحنای مقطع، میانگین سالانه بارندگی، درصد شیب، بافت خاک، شاخص توان جریان، فاصله از آبراهه و شاخص رطوبت پستی‌بلندی انتخاب شدند. وضعیت هم‌خطی چندگانه عامل‌های محیطی با استفاده از آماره عامل تحمل بررسی شد. لایه‌های رستری عامل‌های محیطی به‌عنوان متغیرهای مستقل به مدل یادگیری ماشینی بیز ساده معرفی شدند.

نوع مقاله: پژوهشی

*مسئول مکاتبات: farnoushmohammadi@ut.ac.ir

استناد: محمدی، ف.، بهمنی، ع. ۱۴۰۴. پیش‌بینی مکانی استعداد سیل‌گیری آبخیز زرينه‌رود با استفاده از مدل یادگیری ماشینی بیز ساده. پژوهش‌های آبخیزداری. ۳۸(۴): ۱۲۷-۱۰۸.

شناسه دیجیتال: 10.22092/wmrj.2025.369572.1623

تاریخ دریافت: ۱۴۰۴/۰۳/۰۴، تاریخ بازنگری: ۱۴۰۴/۰۳/۱۱، تاریخ پذیرش: ۱۴۰۴/۰۳/۳۱، تاریخ انتشار: ۱۴۰۴/۱۰/۱۱
پژوهش‌های آبخیزداری، سال ۱۴۰۴ دوره ۳۸، شماره ۴، شماره پیاپی ۱۴۹، زمستان ۱۴۰۴، صفحه‌های ۱۰۸ تا ۱۲۷.

© نویسنده‌گان

ناشر: مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان فارس



نقاط موقعیت‌های رخداد سیل‌گیری براساس روش تصادفی مکانی، با نسبت سی درصد و هفتاد درصد به دو گروه آموزش و اعتبارسنجی تقسیم شد. پس از اجرای مدل، نقشه استعداد سیل آبخیز زرينه‌رود تولید شد؛ به گونه‌ای که هر سلول نشان‌دهنده احتمال سیل‌گیری در آن محدوده بود. دقت نقشه استعداد سیل‌گیری با استفاده از آماره‌های مستقل و وابسته به آستانه و داده‌های گروه اعتبارسنجی ارزیابی شد.

نتایج و بحث

براساس نتایج این پژوهش، متغیرهای مستقل مد نظر بدون هم‌خطی چندگانه بودند و به‌عنوان عامل‌های پیش‌بینی‌کننده در فرایند مدل‌سازی قابل استفاده بودند. نتایج اعتبارسنجی براساس آماره مساحت زیر منحنی مشخصه عملکرد گیرنده نشان داد که دقت نقشه استعداد سیل‌گیری ۹۳/۶٪ بود. بر پایه آماره‌های وابسته به آستانه کارایی مدل یادگیری ماشینی بیز ساده براساس آماره Accuracy ۸۵/۷٪، براساس آماره Precision ۸۲/۶٪ و براساس آماره Recall ۹۰/۴٪ به‌دست آمد.

نتیجه‌گیری و پیشنهادها

کارایی مدل یادگیری ماشینی بیز ساده برای پیش‌بینی مکانی استعداد سیل‌گیری در مقیاس آبخیز مناسب بود و می‌توان از متغیرهای مختلفی برای تجزیه و تحلیل مکانی بهره برد. نقشه استعداد سیل‌گیری باید به‌عنوان مبنای برنامه‌ریزی عملیات سامان‌دهی رود (مانند ساخت دیوار ساحلی و لایه‌روبی بستر)، مدیریت سیل (مانند رعایت حریم رودها) و آبخیزداری (مانند ساخت سازه‌های آبخیزداری در بالادست مناطق مستعد سیل‌گیری) در آبخیز زرينه‌رود، مد نظر باشد. از این‌رو، پیشنهاد می‌شود از این مدل برای تهیه نقشه استعداد سیل‌گیری براساس داده‌های سیلاب‌های تاریخی در دیگر آبخیزهای کشور استفاده شود.

واژگان کلیدی

استان کردستان، تجزیه و تحلیل مکانی، سیل، مدل‌سازی، مدیریت بحران

مقدمه

به‌وجود آورد که خسارت‌های آن ۳۶۵۰ میلیارد ریال برآورد شد (خبرگزاری جمهوری اسلامی، کد خبر ۸۳۲۹۵۱۱۵). از این‌رو، پژوهش علمی پدیده سیل‌گیری^۱، شناسایی مناطق پرخطر و ارائه راهکارهای مدیریتی و پیشگیرانه، ضرورتی اجتناب‌ناپذیر در مسیر توسعه پایدار و کاهش خسارت‌های مالی و تلفات جانی ناشی از این مخاطره طبیعی است. سیل‌گیری زمانی رخ می‌دهد که ظرفیت مجرای رود از آب پر شده و جریان آب به دشت سیلابی وارد شود.

شناسایی پهنه‌های مستعد بلایای طبیعی در یک آبخیز یکی از گام‌های اساسی در مهار مخاطره‌های طبیعی و تدوین برنامه‌های مدیریت جامع آبخیزهای آبخیز است (تاپا و همکاران ۲۰۲۲). تعیین پهنه‌های

سیل یکی از ویران‌گرترین و شایع‌ترین مخاطره‌های طبیعی در سراسر جهان به‌شمار می‌آید که همه‌ساله خسارت‌های جبران‌ناپذیری به زیرساخت‌ها، منابع طبیعی، زمین کشاورزی و زندگی انسان‌ها وارد می‌کند. کشور ایران نیز به‌دلیل شرایط خاص اقلیمی، پستی‌بلندی ناهمگون، پراکنش نامتوازن زمانی-مکانی بارندگی و همچنین ویرانی پوشش گیاهی و دخالت‌های انسانی مانند تغییر کاربری زمین، رخدادهای سیل مہیبی را تجربه کرده است (رجبی‌زاده و همکاران ۲۰۱۹). افزایش فراوانی و شدت رخداد سیلاب‌ها در دهه پیشین نه فقط تهدیدی جدی برای پایداری محیط‌زیست به‌شمار می‌آید، بلکه هزینه‌های هنگفتی نیز بر اقتصاد کشور تحمیل می‌کند (براتی و همکاران ۲۰۱۲). به‌عنوان مثال، بارش‌های فروردین ۱۳۹۸ در استان کردستان، سیلاب‌های پرشماری

1- Flood inundation

شد. نتایج این پژوهش نشان داد دقت مدل جنگل تصادفی بیشتر بود. عرب‌عامری و همکاران (۲۰۲۲) در آبخیز کیسر استان مازندران برای شناسایی مناطق مستعد سیل‌گیری از مدل ماشین بردار پشتیبان^۸ استفاده کردند و دریافتند استفاده از روش‌های فراکاوشی^۹ برای واسنجی مدل می‌تواند مفید باشد. سلیم و همکاران (۲۰۲۲) برای پیش‌بینی استعداد سیل‌گیری شهر برلین آلمان، عملکرد مدل‌های شبکه عصبی مصنوعی، جنگل تصادفی و ماشین بردار پشتیبان را مقایسه کردند. نتایج ایشان نشان داد که قابلیت مدل‌های بررسی‌شده در تجزیه و تحلیل مناطق مستعد سیل‌گیری زیاد بود. بنابراین، می‌توان گفت با استفاده از مدل‌های یادگیری ماشینی و با بهره‌گیری از الگوریتم‌های پیچیده می‌توان، توزیع احتمالاتی برای استخراج الگوی پنهان داده‌ها و روابط میان متغیرهای مستقل و وابسته را کشف و برآورد کرد (سیاه‌کمری و همکاران ۲۰۱۸).

مدل‌های یادگیری ماشینی در پژوهش‌های داخلی و خارجی در زمینه ارزیابی و پیش‌بینی استعداد سیل‌گیری به‌کار گرفته شده‌اند. اگرچه عموماً این مدل‌ها در کشف روابط غیرخطی میان متغیرها و شناسایی مناطق مستعد سیل‌گیری، کارایی زیادی دارند، اما آنچه که باعث تنوع عملکرد آنها می‌شود ساختار و معماری مختلف مدل‌ها است. از سوی دیگر، عامل‌هایی مانند تعداد کم داده‌های رخداد سیل‌گیری می‌تواند بر عملکرد مدل‌ها تأثیرگذار باشد و اندازه کارایی مدل در این شرایط باید بررسی شود. مدل یادگیری ماشینی بیز ساده^{۱۰} (NB) یکی از مدل‌های یادگیری ماشینی است که در پژوهش‌های پیش‌بینی سیل در خارج از کشور استفاده شده است. چن و همکاران (۲۰۲۰) در منطقه کوانان کشور چین، فم و همکاران (۲۰۲۰) در کشور ویتنام و وانگ و همکاران (۲۰۲۱) در شهر زنگشو کشور چین برای پیش‌بینی استعداد سیل‌گیری از مدل یادگیری ماشینی بیز ساده استفاده کردند و نتایج قابل قبولی به‌دست آمد. در

سیل‌گیر با استفاده از روش‌های نوین تحلیل مکانی و رویکردهای مختلف مدل‌سازی، این امکان را فراهم می‌آورد تا تصمیم‌گیران و برنامه‌ریزان دستگاه‌های اجرایی مرتبط با سیل بتوانند برای پیشگیری، مهار و آمادگی در برابر رخداد‌های احتمالی سیل تصمیم‌های مناسبی بگیرند (کی و همکاران ۲۰۲۱). بدون آگاهی از پهنه‌های خطر سیل‌گیری در یک آبخیز، برنامه‌ریزی در زمینه تعیین حریم رودها و تخصیص منابع به‌طور غیربهرینه بوده و حتی توسعه شهری و زیرساخت‌ها گاهی اوقات در مناطق پرخطر انجام خواهد شد (خواج‌نژاد و همکاران ۲۰۲۴). از این‌رو، ارزیابی دقیق خطرپذیری سیل و تهیه نقشه‌های استعداد سیل‌گیری در مقیاس‌های مختلف محلی، منطقه‌ای و ملی، نقش مهمی در مدیریت بحران، کاهش خسارت‌های جانی و مالی، افزایش تاب‌آوری جامعه‌های و دستیابی به توسعه پایدار دارد.

پژوهشگران برای تهیه نقشه استعداد سیل‌گیری از رویکردهای مختلف مدل‌سازی مانند مدل‌های آماری، داده‌کاوی و یادگیری ماشینی مختلفی استفاده کرده‌اند (سامان‌تا و پالسمان‌تا ۲۰۱۸؛ معظم و همکاران ۲۰۲۰). در این راستا، تهرانی و کومار (۲۰۱۸) برای پیش‌بینی استعداد سیل منطقه‌ای در بریزبن کشور استرالیا از داده‌های مشاهده‌ای رخداد سیل‌گیری و مدل‌های تابع شواهد قطعی^۲، نسبت فراوانی^۳ و وایازی لجستیک^۴ استفاده کردند. دمیسی و همکاران (۲۰۲۴) در منطقه‌ای در ایالت کانزاس آمریکا، کارایی مدل‌های یادگیری ماشینی مختلفی در زمینه پیش‌بینی استعداد سیل‌گیری را ارزیابی کردند. نتایج این پژوهش نشان داد دقت مدل‌های جنگل تصادفی^۵ و تقویت شیب شدید^۶، بیشتر بود. در پژوهشی در کشور رومانی، عابدی و همکاران (۲۰۲۲) با استفاده از مدل‌های جنگل تصادفی، درخت وایازی تقویت‌شده^۷ و تقویت شیب شدید، نقشه استعداد سیل‌گیری تولید

2- Evidential belief function

3- Frequency ratio

4- Logistic regression

5- Random Forest

6- Extreme gradient boosting

7- Boosted regression trees

8- Support vector machine

9- Meta-heuristic

10- Naive Bayes

نقشه می‌تواند مبنای برنامه‌ریزی برای عملیات آب‌بخیزداری و مدیریت سیل باشد.

مواد و روش‌ها

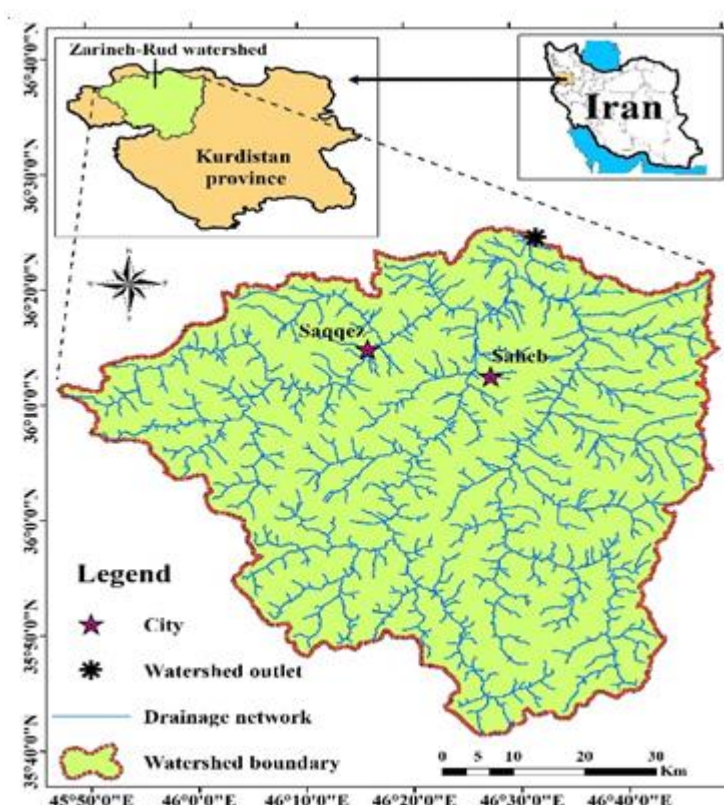
معرفی منطقه پژوهش‌شده

محدوده جغرافیایی آب‌بخیز زرینه‌رود در استان کردستان در شمال غرب ایران در شکل ۱ نشان داده شده است. آب‌بخیز زرینه‌رود میان طول‌های جغرافیایی $46^{\circ} 46'$ تا $46^{\circ} 48'$ شرقی و عرض‌های جغرافیایی $35^{\circ} 42'$ تا $36^{\circ} 23'$ شمالی است. این آب‌بخیز که در بخش شمالی استان است، شامل مناطق کوهستانی و ناهموار بوده و مساحت آن ۴۴۸۵ کیلومتر مربع است. آب‌بخیز زرینه‌رود شبکه‌ای گسترده از آبراهه‌ها و رودها دارد و آب از بخش‌های جنوبی به سمت بخش‌های شمالی جریان دارد.

ایران، حبیبی و همکاران (۲۰۲۳) برای پیش‌بینی استعداد سیل‌گیری آب‌بخیز سرداب‌رود استان مازندران از مدل بیز ساده استفاده کردند. با این وجود، به دلیل اطلاعات کم در زمینه کاربرد مدل بیز ساده در پژوهش سیل در بخش‌های مختلف کشور، ارزیابی کارایی این مدل در زمینه پیش‌بینی استعداد سیل‌گیری در کشور ایران نیازمند پژوهش‌های بیشتری است.

آب‌بخیز زرینه‌رود در استان کردستان یکی از مناطق مستعد سیل در کشور است که در دهه پیشین سیلاب‌ها، خسارت‌های مالی و تلفات جانی زیادی در این آب‌بخیز به همراه داشته‌اند (اعلمی و همکاران ۲۰۲۴؛ چوبین و همکاران ۲۰۲۵).

از این رو، این پژوهش به منظور ارزیابی عملکرد مدل یادگیری ماشینی بیز ساده در زمینه پیش‌بینی مکانی سیل‌گیری و تولید نقشه استعداد سیل‌گیری در آب‌بخیز زرینه‌رود، اجرا شد. پس از تأیید علمی و فنی نقشه استعداد سیل‌گیری به دست آمده از مدل‌سازی، این



شکل ۱- موقعیت جغرافیایی آب‌بخیز زرینه‌رود.

Figure 1- Geographical location of the Zarineh-Rud Watershed.

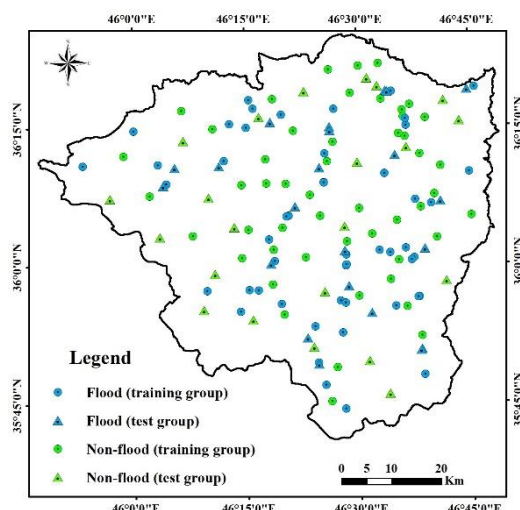
این آبخیز، دو شهر اصلی سقز و صاحب و روستاهای پرشماری دارد. خروجی اصلی این آبخیز در بخش شمال‌شرقی آبخیز است، جایی که جریان‌های سطحی به سمت دریاچه ارومیه در استان آذربایجان غربی هدایت می‌شوند. در دهه پیشین رخدادهای سیلابی بزرگی در آبخیز زرينه رخ داده است که گاه‌آمنجر به خسارت‌های مالی شدیدی شده است. در این راستا، یک رخداد بارش در تیرماه ۱۳۹۴، سبب خسارت‌های زیادی به اماکن مسکونی، تجاری و اداری شهر سقز در استان کردستان شد که اندازه آن ۲۱۰ میلیارد ریال برآورد شد (خبرگزاری جمهوری اسلامی، کد خبر ۸۱۶۹۰۰۴۳). براساس روش طبقه‌بندی اقلیمی دومارتن، این آبخیز در پهنه نیمه‌مرطوب است. از این‌رو، تابستان‌های این آبخیز معتدل و زمستان‌های آن سرد است. در برخی سال‌ها، حتی در اواخر فصل بهار و اوایل فصل تابستان بارش‌های سیل‌آسا به شکل تگرگ رخ داده که منجر به جاری‌شدن سریع رواناب و شکل‌گیری سیلاب شده است. مثلاً در تیرماه ۱۳۹۴ یک سیلاب مهیب در آبخیز رخ داد و سبب خسارت‌هایی در محدوده شهر سقز به اماکن مسکونی و تجاری شد (حسینی و همکاران ۲۰۲۱).

روش پژوهش

جمع‌آوری داده‌های سیل‌گیری

جمع‌آوری اطلاعات و ثبت داده‌ها اهمیت بسیار زیادی دارد. در این پژوهش، شواهد زمینی رخداد سیل‌گیری مهم‌ترین داده لازم برای آموزش مدل بود. داده‌های مکانی موقعیت رخداد سیل‌گیری به‌عنوان متغیر وابسته (متغیر هدف) به مدل معرفی شد. مشخصات سیل از با ارزش‌ترین اطلاعات تاریخی یک آبخیز به‌شمار می‌آید و دستگاه‌های اجرایی سعی بر ثبت آن‌ها داشته‌اند. موقعیت جغرافیایی کلیه رخدادهای

سیل‌گیری ثبت‌شده در بانک‌های اطلاعات اداره کل منابع طبیعی و آبخیزداری، شرکت آب منطقه‌ای، اداره کل راهداری و حمل و نقل جاده‌ای، دفتر مدیریت بحران استانداری و جمعیت هلال احمر استان کردستان ثبت‌شده بودند، که برای انجام این پژوهش جمع‌آوری شدند. سپس، صحت داده‌های موقعیت سیل‌گیری در بازدیدهای میدانی و مصاحبه با افراد بومی بررسی شد. براساس تحلیل‌های انجام‌شده، مجموعاً ۷۲ موقعیت رخداد تاریخی سیل‌گیری در آبخیز زرينه‌رود ثبت‌شده بود، که در این پژوهش استفاده شد. از آنجایی که اجرای مدل‌های یادگیری ماشینی عموماً نیازمند داده‌های رخداد و نبود رخداد سیل هستند، براساس تحلیل‌های کارشناسی، ۷۲ موقعیت نبود رخداد سیل‌گیری شناسایی شد که از دیدگاه علمی احتمال سیل‌گیری آن‌ها بسیار کم بود. به دلیل اینکه مدل‌سازی نیازمند دو گروه داده شامل آموزش و اعتبارسنجی است، داده‌های رخداد و نبود رخداد سیل به شکل تصادفی به دو گروه آموزش و اعتبارسنجی تقسیم شدند. گروه آموزش شامل ۷۰٪ از مجموعه داده‌های رخداد و نبود رخداد سیل بود که برای شناسایی روابط میان متغیرها و همچنین الگوی پیش‌بینی مدل، استفاده شد (چاودوری و همکاران ۲۰۲۵). از سوی دیگر، داده‌های گروه اعتبارسنجی در مرحله آموزش استفاده نشدند و صرفاً پس از پایان فرایند مدل‌سازی و تولید نقشه استعداد سیل، به‌منظور تعیین دقت مدل به‌کارگرفته شد. نتایج مدل‌سازی و خروجی مدل بدون تایید آماره‌های ارزیابی در مرحله اعتبارسنجی، بدون اعتبار می‌باشند (تراوره و همکاران ۲۰۲۴). موقعیت داده‌های گروه آموزش و گروه اعتبارسنجی به تفکیک نقاط رخداد و نبود رخداد سیل در شکل ۲ ارائه شده است.



شکل ۲- موقعیت رخداد های سیل گیری در آبخیز زرینه رود.

Figure 2- Location of the flood inundation events in Zarineh-Rud Watershed.

وجود داشته باشد. مدل سازی استعداد سیل گیری در شرایط وجود هم خطی چندگانه صحیح نبوده و بر عملکرد مدل تأثیرگذار است (چان و همکاران ۲۰۲۲). بنابراین، بررسی این موضوع میان متغیرهای مستقل قبل از اجرای مدل، ضروری است. در این پژوهش از عامل تحمل^{۱۲} برای تعیین هم خطی چندگانه استفاده شد. آماره عامل تحمل با استفاده از رابطه ۱ محاسبه شد (سلیمان و همکاران ۲۰۲۱).

$$TOL = R_{ij}^2 - 1 \quad (1)$$

TOL: آماره عامل تحمل، R_{ij}^2 ضریب تعیین دو متغیر i و j است.

هنگامی که اندازه آماره عامل تحمل کمتر از ۰/۱ باشد، مشکل هم خطی چندگانه در میان متغیرهای مستقل وجود دارد (دوتا و دکا ۲۰۲۴). در این شرایط، از میان متغیرهایی که همبستگی زیادی با یکدیگر داشتند، فقط یکی از آنها برای مدل سازی انتخاب شد.

مدل یادگیری ماشینی بیز ساده

مدل بیز ساده^{۱۳} (NB) یکی از مدل های پرکاربرد یادگیری ماشینی است که در سال های پیشین پژوهشگران برای حل مسائل طبقه بندی در علوم

انتخاب عامل های محیطی مؤثر بر رخداد سیل گیری عامل های محیطی مختلف پستی بلندی، آب شناختی، زمین ریخت شناختی، کاربری زمین، پوشش گیاهی، خاک شناسی و زمین شناسی بر رخداد سیل تأثیرگذارند (رحمتی و همکاران ۲۰۲۰؛ ویدیا و همکاران ۲۰۲۴). در این پژوهش، تعداد ۱۳ عامل محیطی مؤثر بر سیل گیری شامل بلندی، انحنای سطح، انحنای مقطع، شاخص موقعیت پستی بلندی، موقعیت شیب نسبی، اندازه شیب، جهت شیب، شاخص رطوبت پستی بلندی، شاخص توان جریان، میانگین سالانه بارندگی، فاصله از جریان، تراکم زهکشی، سنگ شناسی، بافت خاک، کاربری زمین انتخاب شدند. رابطه میان عامل ها با فرایند شکل گیری سیل معنی دار بود و تغییرات مکانی هر یک از آنها در این آبخیز زیاد بود (چن و همکاران ۲۰۲۱). هنگامی که تغییرات عامل های محیطی کم و از دیدگاه فضایی یکنواخت باشند، اطلاعات زیادی برای مدل فراهم نشده و نمی توان با استفاده از آن مناطق مستعد سیل گیری از دیگر مناطق را تفکیک کرد.

بررسی هم خطی چندگانه عامل های محیطی

هم خطی چندگانه^{۱۱}، هنگامی رخ می دهد که میان عامل های محیطی (متغیرهای مستقل) همبستگی

12- Tolerance
13- Naive Bayes

11- Multicollinearity

آبی در آن‌ها امکان‌پذیر نیست. افزون بر این، برخی از مدل‌های یادگیری ماشینی نیازمند داده‌های زیادی برای آموزش هستند که کاربرد آن‌ها نیز محدودیت زیادی دارد. در این پژوهش، استفاده از مدل یادگیری ماشینی بیز ساده برای پیش‌بینی استعداد سیل‌گیری یک آبخیز امکان‌سنجی شد و کارایی آن از دیدگاه معیارهای آماری ارزیابی و کاوش شد.

اعتبارسنجی نتایج مدل

در این پژوهش، از آماره‌های مختلف مستقل از آستانه^{۱۸} و وابسته به آستانه^{۱۹} برای ارزیابی قابلیت پیش‌بینی و عملکرد مدل استفاده شد. آماره‌های مستقل از آستانه به‌عنوان نمایه‌های اصلی تعیین عملکرد مدل مطرح شدند و آماره‌های وابسته به آستانه برای اطمینان بیشتر از کارایی مدل و یا دستیابی به جزئیات بیشتر در زمینه عملکرد مدل استفاده شدند (رحمتی و همکاران ۲۰۱۹).

مساحت زیرمنحنی مشخصه عملکرد گیرنده (AUC) به‌عنوان معتبرترین آماره ارزیابی دقت مدل‌های طبقه‌بندی به‌شکل گسترده استفاده شده است. این آماره به‌طور مستقل از آستانه محاسبه می‌شود. از این‌رو، دیدگاه کارشناسی بر آن تأثیرگذار نیست (فراآینی و همکاران ۲۰۱۰). در این پژوهش، داده‌های رخداد و نبود رخداد سیل‌گیری گروه اعتبارسنجی روی نقشه خروجی مدل پیاده‌سازی شد و داده‌های متناظر احتمال سیل‌گیری استخراج شد. براساس داده‌های استخراج‌شده، منحنی مشخصه عملکرد گیرنده در نرم‌افزار آماری R رسم شد و مساحت زیر منحنی نامبرده به‌عنوان آماره ارزیابی دقت مدل محاسبه شد. از آنجایی که داده‌های گروه اعتبارسنجی در مراحل آموزش و ساخت مدل یادگیری ماشینی بیز ساده استفاده نشدند، آماره ارزیابی مساحت زیرمنحنی مشخصه عملکرد گیرنده بیانگر کارایی و عملکرد پیش‌بینی مدل بود. طبقه‌بندی عملکرد مدل براساس اندازه آماره مساحت زیرمنحنی مشخصه عملکرد گیرنده اینگونه بود که اندازه آماره کمتر از ۰/۷

محیطی مانند پیش‌بینی مخاطره‌های طبیعی به آن توجه داشته‌اند (وانگ و همکاران ۲۰۲۰). این مدل بر پایه قضیه بیز^{۱۴} و فرض استقلال عامل‌های محیطی (متغیرهای مستقل) ساخته شده است. قضیه بیز نظریه‌ای است که براساس رخداد یک رویداد، احتمال یک رویداد دیگر را توصیف می‌کند. این یک مفهوم اساسی در احتمال و آمار است که در هنگام مشاهده شواهد جدید برای به روزرسانی احتمالات استفاده می‌شود. این قضیه به نام آماردان معروف توماس بیز^{۱۵} نام‌گذاری شده است.

این مدل با محاسبه احتمال پسین هر طبقه و با توجه به داده‌های عامل‌های محیطی ورودی، طبقه‌بندی را انجام می‌دهد و به‌رغم ساده بودن ساختار آن، در عمل به‌ویژه در حل مسائل پیچیده مانند علوم طبیعی و علوم پزشکی کاربرد گسترده‌ای دارد (مسوادی و همکاران ۲۰۲۱). معادله اصلی مدل بیز به‌شکل رابطه ۲ است (برار ۲۰۲۵).

$$P(C|X) = \frac{P(X|C) \times P(C)}{P(X)} \quad (2)$$

$P(C|X)$: احتمال اینکه سلولی^{۱۶} از لایه متغیر مستقل X متعلق به طبقه C (پهنه سیل‌گیر) باشد (احتمال پسین)، $P(X|C)$: احتمال مشاهده متغیر مستقل X در شرایطی که متعلق به طبقه C (پهنه سیل‌گیر) باشد، $P(C)$: احتمال پیشین طبقه C (پهنه سیل‌گیر) و $P(X)$: احتمال کلی مشاهده ویژگی X است.

سرانجام، با این مدل می‌توان احتمال رخداد سیل‌گیری را با توجه به اندازه‌های عامل‌های محیطی برای هر سلول محاسبه کرد. مزیت اصلی این مدل، سادگی، سرعت زیاد و کارایی مناسب حتی با داده‌های کم حجم است (پاجیلا و همکاران ۲۰۲۳). از آنجایی که داده‌های کامل رخداد سیل‌گیری در آبخیزهای کشور موجود نیست، عموماً به‌عنوان مناطق کم‌داده^{۱۷} شناخته شده‌اند و استفاده از مدل‌های آب‌شناختی و

18- Cut-off independent threshold

19- Cut-off dependent threshold

14- Bayes' Theorem

15- Thomas Bayes

16- Pixel

17- Data-scarce

می‌شود. زمانی اهمیت این موضوع بیشتر می‌شود که دستگاه‌های اجرایی با محدودیت منابع مالی و انسانی روبرو باشند. نقشه پهنه‌بندی استعداد سیل‌گیری به‌روشنی مناطق بحرانی که باید در اولویت باشند را مشخص می‌کند.

نتایج و بحث

عامل‌های محیطی مؤثر بر رخداد سیل‌گیری

سیل، یک پدیده پیچیده است که در اثر مشارکت عامل‌های محیطی مختلف رخ می‌دهد. در این پژوهش، ۱۳ عامل محیطی مؤثر بر سیل‌گیری از جمله عامل‌های پستی‌بلندی، خاک‌شناسی، آب‌شناسی، کاربری، اقلیمی، زمین‌شناختی و محیطی برای استفاده در مدل‌سازی انتخاب شدند که نقشه آن‌ها در شکل ۳ ارائه شده است.

در نظر گرفتن عامل‌های محیطی مختلف برای تجزیه و تحلیل مکانی و پیش‌بینی استعداد سیل‌گیری ضروری است. تنوع و تغییرات مکانی در هر یک از متغیرهای بلندی زمین، جهت دامنه، تراکم زهکشی، کاربری زمین، سنگ‌شناسی، انحنا سطح، انحنا مقطع، میانگین سالانه بارندگی، درصد شیب، بافت خاک، شاخص توان جریان، فاصله از آبراهه و شاخص رطوبت پستی‌بلندی در آبخیز زرينه‌رود مشاهده شد. تغییرات اندازه بلندی در این آبخیز قابل مشاهده و از ۱۳۶۳ تا ۳۱۴۸ متر متغیر بود. جهت‌های گوناگون دامنه در بخش‌های متفاوت پراکنش داشت. تراکم زهکشی در این آبخیز از صفر تا ۱/۷ کیلومتر بر کیلومتر مربع متغیر بود و بیشترین اندازه این متغیر در بخش‌های شمالی آبخیز مشاهده شد. کاربری‌های مسکونی، کشاورزی، زمین بایر، باغ، مرتع و پهنه آبی در این آبخیز شناسایی شد. از دیدگاه سنگ‌شناسی، ۲۶ واحد گوناگون در این آبخیز شناسایی شد که نشان‌دهنده تنوع و پیچیدگی سنگ‌شناسی منطقه بود.

با توجه به پستی‌بلندی نسبتاً متنوع این آبخیز، دامنه‌های آن از دیدگاه انحنا سطح و انحنا مقطع به طبقه‌های مقعر، محدب و تخت طبقه‌بندی شدند. اندازه بارندگی سالانه در منطقه پژوهش‌شده از ۳۵۶ تا

نشان‌دهنده عملکرد ضعیف مدل، اندازه آماره میان ۰/۷ تا ۰/۸ نشان‌دهنده عملکرد خوب مدل، اندازه آماره میان ۰/۸ تا ۰/۹ نشان‌دهنده عملکرد خیلی خوب و اندازه آماره میان ۰/۹ تا ۱ بیانگر عملکرد عالی مدل بود (عرب‌عامری و همکاران ۲۰۲۰).

در این پژوهش، آماره‌های مختلف وابسته به آستانه شامل Accuracy، Precision و Recall برای اطمینان از عملکرد مدل محاسبه شدند. در ماتریس خطا چهار مؤلفه مثبت درست (TP)، منفی درست (TN)، مثبت نادرست (FP) و منفی نادرست (FN) بود که آماره‌های وابسته به آستانه براساس آن‌ها تعیین شدند. هر یک از این آماره‌ها با استفاده از رابطه‌های ۳، ۴ و ۵ محاسبه شد (فراتینی و همکاران ۲۰۱۰).

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

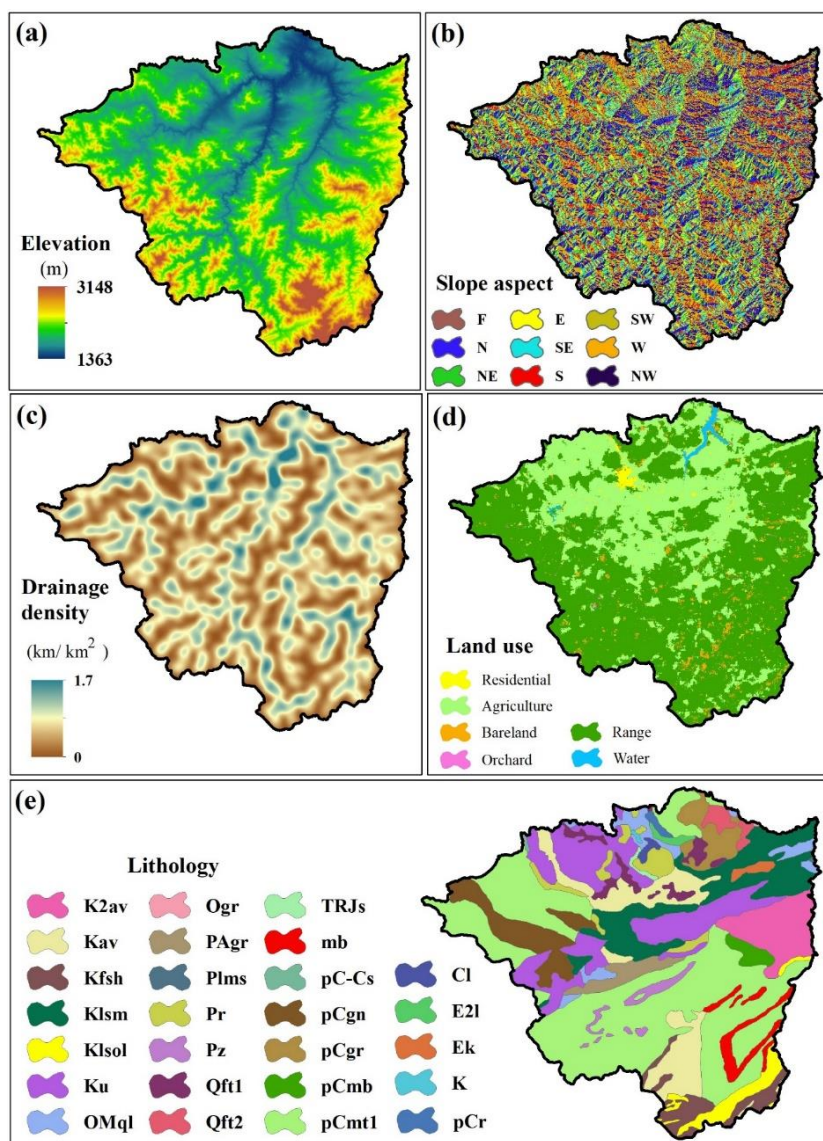
$$Recall = \frac{TP}{TP + FN} \quad (5)$$

تولید نقشه پهنه‌بندی استعداد سیل‌گیری

نقشه احتمال رخداد سیل‌گیری (نتیجه مدل‌سازی) یک نقشه کمی است که ارزش عددی هر سلول میان صفر تا یک متغیر است (چوبین و همکاران ۲۰۲۳). به‌منظور درک بهتر وضعیت سیل‌گیری آبخیزها، لایه‌های رستری نقشه به‌دست آمده از فرایند مدل‌سازی، با استفاده از روش طبقه‌بندی گام یکسان، پهنه‌بندی شد. نقشه پهنه‌بندی استعداد سیل‌گیری شامل پنج طبقه مختلف استعداد خیلی کم، استعداد کم، استعداد متوسط، استعداد زیاد و استعداد خیلی زیاد بود. این نقشه اطلاعات مفیدی برای مدیران و کارشناسان منابع طبیعی و آبخیزداری فراهم می‌آورد تا روی پهنه‌های مستعد سیل‌گیری زیاد و خیلی زیاد تمرکز کنند و منابع مالی و فعالیت‌های اجرایی را برای این پهنه‌ها برنامه‌ریزی کنند. در برخی شرایط، عملیات آبخیزداری و مهار سیلاب بدون توجه به وضعیت بالقوه سیل‌گیری در بخش‌هایی از آبخیز اجرا

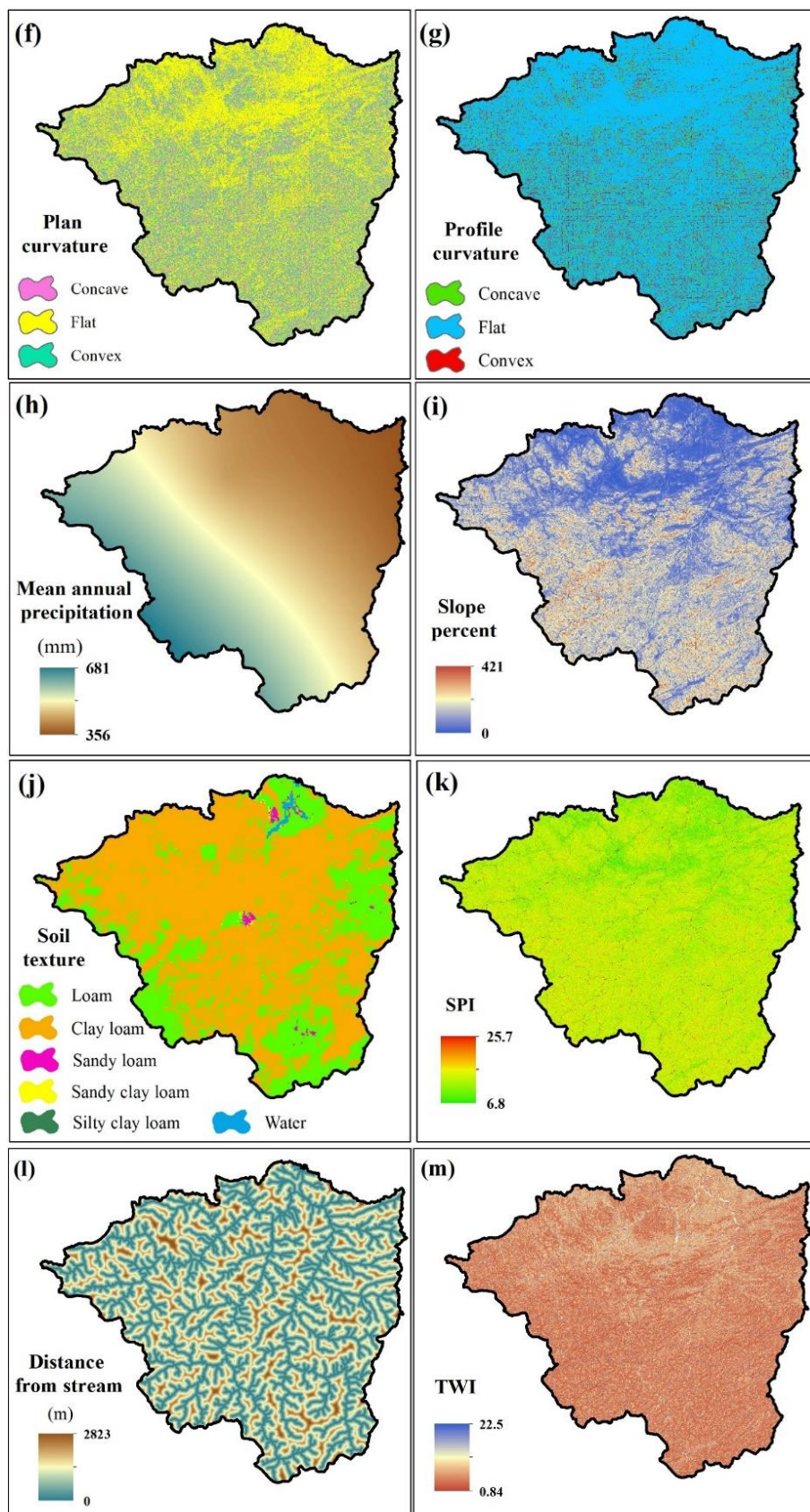
این آبخیز، تغییرات مکانی از دیدگاه شیب زمین قابل مشاهده بود.

۶۸۱ میلی‌متر متغیر بود. اندازه شیب زمین در این آبخیز از صفر تا ۴۲۱٪ متغیر بود. به بیان دیگر، در



شکل ۳- متغیرهای مستقل برای پیش‌بینی استعداد سیل‌گیری در آبخیز زرینه‌رود: (a) بلندی، (b) جهت شیب، (c) تراکم زهکشی، (d) کاربری زمین، (e) سنگ‌شناسی، (f) انحنا سطح، (g) انحنا مقطع، (h) میانگین سالانه بارندگی، (i) درصد شیب، (j) بافت خاک، (k) شاخص توان جریان، (l) فاصله از جریان و (m) شاخص رطوبت پستی‌بلندی.

Figure 3- Independent variables for predicting flood susceptibility in the Zarineh-Rud Watershed: a) elevation, b) slope aspect, c) drainage density, d) land use, e) lithology, f) plan curvature, g) profile curvature, h) mean annual precipitation, i) slope percent, j) soil texture, k) stream power index (SPI), l) distance from stream, and m) topographic wetness index (TWI).



شکل ۳- (ادامه).

Figure 3- (continued).

وضعیت هم‌خطی چندگانه

نتایج آزمون هم‌خطی چندگانه در جدول ۱ ارائه شده است. اندازه آماره عامل تحمل برای هر یک از متغیرهای مستقل آبخیز زرینه‌رود در این پژوهش بزرگتر از ۰/۱ به دست آمد. از این رو، متغیرهای مستقل استفاده شده بدون هم‌خطی چندگانه بودند و تمامی آن‌ها از دیدگاه آماری به عنوان عامل‌های پیش‌بینی‌کننده به مدل معرفی شدند.

در این آبخیز، طبقه‌های بافت خاک شامل لومی، لومی رسی، لومی شنی، لومی رسی شنی، و لومی رسی سیلتی بود. همچنین، بیشترین اندازه شاخص توان جریان مربوط به بخش‌های جنوبی آبخیز بود. با توجه به گستردگی شبکه زهکشی، در بیشتر بخش‌های این آبخیز مسیر زهکشی وجود داشت و اندازه فاصله تا نزدیکترین آبراهه یا رود از صفر تا ۲۸۲۳ متر متغیر بود. اندازه شاخص رطوبت پستی‌بلندی نیز از ۰/۸۴ تا ۲۲/۵ متغیر بود و بیشترین اندازه آن در بخش‌های شمالی که رودهای اصلی جریان داشتند، به دست آمد.

جدول ۱- آماره عامل تحمل متغیرهای مستقل استفاده شده در آبخیز زرینه‌رود.

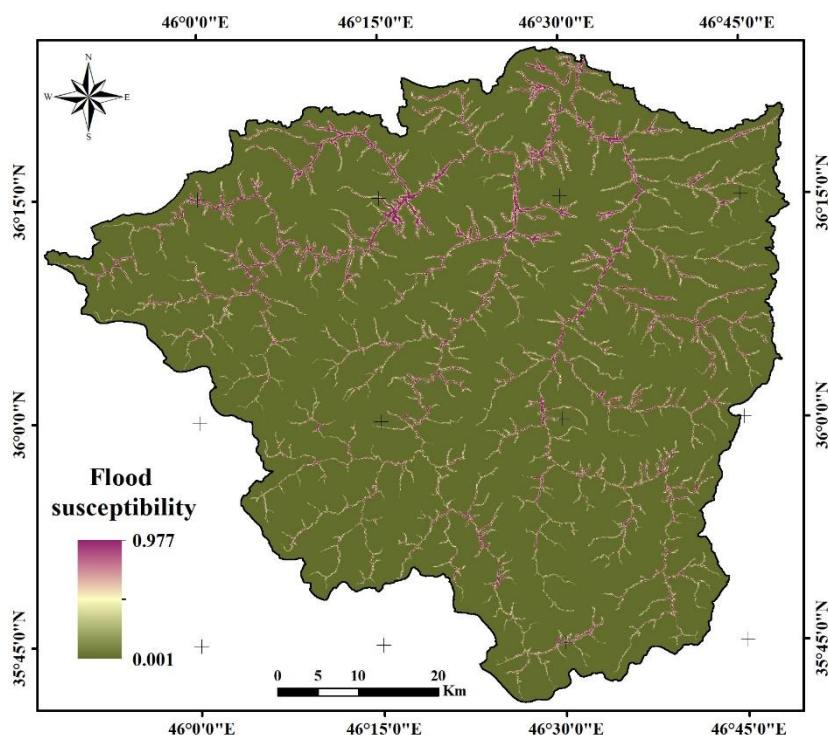
Table 1- Tolerance statistic for the independent variables used in the Zarineh-Rud Watershed.

No.	Variable	Tolerance	No.	Variable	Tolerance
1	Soil texture	0.781	8	TWI	0.121
2	SPI	0.125	9	Lithology	0.905
3	Distance from stream	0.265	10	Plan curvature	0.839
4	Slope aspect	0.903	11	Precipitation	0.762
5	Elevation	0.313	12	Profile curvature	0.784
6	Drainage density	0.258	13	Slope percent	0.185
7	Land use	0.742			

الگوی مکانی استعداد سیل‌گیری

نقشه استعداد سیل‌گیری آبخیز زرینه‌رود که با استفاده از مدل یادگیری ماشینی بیز ساده و براساس رخدادهای تاریخی سیل این آبخیز تولید شد، در شکل ۴ نشان داده شده است. اندازه استعداد سیل در بخش‌های مختلف این آبخیز متفاوت بود و بسته به شرایط محیطی از ۰/۰۰۱ تا ۰/۹۷۷ متغیر بود.

براساس نتایج به دست آمده، سلول‌هایی که بالقوه سیل‌گیری زیادی داشتند عموماً در بخش‌های شمالی، جنوبی و شرقی آبخیز زرینه‌رود مشاهده شدند. تغییرات مکانی درجه استعداد سیل‌گیری در این نقشه مؤید پیچیدگی پدیده سیل و مشارکت عامل‌های محیطی مختلف در مقیاس آبخیز بود.



شکل ۴- نقشه استعداد سیل‌گیری آبخیز زرینه‌رود با استفاده از مدل بیز ساده.

Figure 4- Flood susceptibility map of the Zarineh-Rud Watershed using the Naive Bayes Model.

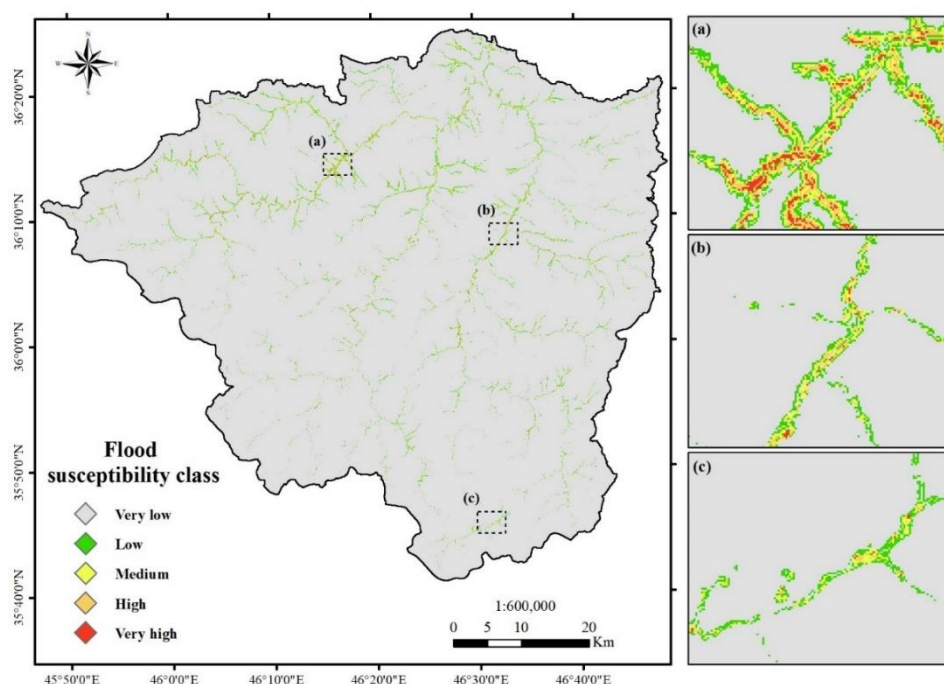
از این رو، امکان متمرکز کردن عملیات اجرایی و برنامه‌های مدیریتی روی پهنه‌های با استعداد سیل‌گیری زیاد و خیلی زیاد می‌تواند کارایی اقدامات را در مقیاس آبخیز به‌طور چشمگیری افزایش دهد. اهمیت این نکته هنگامی بیشتر خواهد بود که منابع مالی و انسانی برای مدیریت سیل آبخیز محدود باشد و تصمیم‌گیران باید سطح محدودی از یک آبخیز را برای اجرای عملیات انتخاب کنند.

نقشه پهنه‌بندی استعداد سیل‌گیری در شکل ۵ نشان داده شده است. پهنه‌بندی نقشه استعداد سیل‌گیری با استفاده از روش طبقه‌بندی گام یکسان^{۲۰} انجام شد (پال و سینقا ۲۰۲۲). در این نقشه، آبخیز زرینه‌رود به پنج طبقه استعداد خیلی کم (صفر تا ۰/۲)، کم (۰/۲ تا ۰/۴)، متوسط (۰/۴ تا ۰/۶)، زیاد (۰/۶ تا ۰/۸) و خیلی زیاد (۰/۸ تا ۱) تقسیم شد.

سه قسمت از این نقشه بزرگ‌نمایی شد. قسمت اول (a) شامل محدوده شهری سقز و روستاهای قبل و بعد از آن بود. قسمت دوم (b) شامل زمین محدوده روستاهای خانه‌میران و چناره بود. قسمت سوم (c) نیز زمین روستای گوگجه بود. بر اساس بخش‌های بزرگ‌نمایی‌شده این نقشه، تغییرات مکانی زیادی از دیدگاه وضعیت استعداد سیل‌گیری در بخش‌های مختلف آبخیز زرینه‌رود مشاهده شد.

نقشه پهنه‌بندی استعداد سیل‌گیری به شکل روشنی معرف مناطق پرخطر برای مدیران و کارشناسان است.

20- Equal interval classification



شکل ۵- نقشه پهنه‌های استعداد سیل‌گیری آبخیز زرینه‌رود با استفاده از مدل بیز ساده.

Figure 5- Flood susceptibility map of the Zarineh-Rud Watershed using the Naive Bayes Model.

نتایج پژوهش رحمتی و همکاران (۲۰۲۴) که از مدل تحلیل تفکیک‌کننده انعطاف‌پذیر برای پیش‌بینی استعداد سیل‌گیری آبخیز زرینه‌رود استفاده کردند، حدود ۶۲ هزار هکتار از این آبخیز در پهنه‌های استعداد سیل‌گیری زیاد و خیلی‌زیاد بودند. با توجه به تفاوت در ساختار پیش‌بینی مدل‌های پیش‌بینی‌کننده، اختلاف در تعیین پهنه‌های استعداد سیل‌گیری قابل انتظار بود. تطابق نقشه‌های پیش‌بینی به‌دست‌آمده از پژوهش رحمتی و همکاران (۲۰۲۴) با یافته‌های این پژوهش بیانگر آن بود که با استفاده از مدل یادگیری ماشینی بیز ساده بخش‌های کمتری از آبخیز زرینه‌رود در پهنه‌های استعداد سیل‌گیری زیاد و خیلی‌زیاد بودند. در این آبخیز، پژوهش محدودی در زمینه شناسایی مناطق مستعد سیل اجرا شده است و انجام پژوهش‌هایی با استفاده از رویکردهای مختلف در این آبخیز ضروری است. هنگامی که یک مدل سطح بیشتری را به‌عنوان منطقه بحرانی (با استعداد سیل‌گیری زیاد و خیلی‌زیاد) پیش‌بینی کند، ممکن است در شرایط محدودیت بودجه، عملیات اجرایی

مساحت پهنه‌های استعداد سیل‌گیری آبخیز زرینه‌رود در جدول ۲ ارائه شده است. در این آبخیز براساس تحلیل‌های انجام‌شده، بیشترین سطح پهنه‌ها به‌ترتیب مربوط به پهنه‌های استعداد خیلی‌کم سیل‌گیری (۴۳۷۵۶۵/۷۸ هکتار) و استعداد کم سیل‌گیری (۶۵۹۸/۹۸ هکتار) بود. مساحت پهنه استعداد متوسط سیل‌گیری نیز حدود ۲۷۶۴ هکتار از زمین آبخیز زرینه‌رود بود. همچنین، نتایج نشان داد حدود ۱۲۹۶ هکتار از این آبخیز در پهنه استعداد سیل‌گیری زیاد بود. افزون بر آن، حدود ۲۹۷ هکتار از این آبخیز در پهنه استعداد خیلی‌زیاد سیل‌گیری بود. از آنجایی که پهنه‌های استعداد زیاد و خیلی‌زیاد سیل‌گیری نیازمند توجه ویژه هستند، حدود ۱۵۹۳ هکتار از این آبخیز از دیدگاه سیل‌گیری در وضعیت بحرانی هستند. براساس نتایج این پژوهش می‌توان گفت، تمام برنامه‌های عملیاتی و مدیریتی در راستای مهار سیل باید به محدوده ۱۵۹۳ هکتار معطوف شود تا از هدررفت منابع و اجرای اقدامات پیشگیرانه سیل در مناطق با استعداد سیل‌گیری کمتر جلوگیری شود. براساس

زمین است و گروه سخت شامل عملیات فیزیکی مانند ساخت دیواره‌های ساحلی و سیل‌بندها است.

جدول ۳- مناطق روستایی مستعد سیل‌گیری در آبخیز زرینه‌رود.

Table 3- Rural areas prone to flooding in the Zarineh-Rud Watershed.

No.	Village	City
1	Best	Divandareh
2	Azizabad	Divandareh
3	Biandareh	Baneh
4	Bademjan	Baneh
5	Khoshedare	Saqquez
6	Eshaghabad	Saqquez
7	Darezarat-sofla	Saqquez
8	Talejar	Saqquez
9	Hangechineh	Saqquez
10	Sardareh	Saqquez
11	Piromaran	Saqquez
12	Gheshlagh-pol	Saqquez
13	Kaniband	Saqquez
14	Ghahreman	Saqquez
15	Kaniniyaz	Saqquez
16	Gheshlagh-reza	Saqquez
17	Ghaplanto	Saqquez
18	Mahidar-olia	Saqquez
19	Khorkhoreh	Saqquez
20	Mirge-nakhshineh	Saqquez

ارزیابی کارایی مدل

نتایج ارزیابی عملکرد مدل در این پژوهش براساس دو گروه معیارهای مستقل از آستانه و وابسته به آستانه در جدول ۴ نشان داده شده است. با استفاده از آماره مساحت زیرمنحنی مشخصه عملکرد گیرنده (AUC) اثبات شد که دقت خروجی مدل ۰/۹۳۶ (۹۳/۶٪) بود. در این پژوهش، براساس طبقات عملکرد مدل بر اساس معیار ارزیابی AUC، مشاهده شد عملکرد مدل ناوی بیز در دسته‌بندی در طبقه عالی بود. بر پایه روش پژوهش، آماره‌های وابسته به آستانه به‌عنوان معیارهای ثانویه انتخاب شدند و صرفاً برای اطمینان از قابلیت پیش‌بینی مدل استفاده شدند. براساس آماره‌های وابسته به آستانه می‌توان گفت کارایی مدل یادگیری ماشینی بیز ساده براساس آماره Accuracy ۸۵/۷٪، براساس آماره Precision ۸۲/۶٪ و براساس آماره Recall ۹۰/۴٪ به‌دست آمد. افزون بر این،

آبخیزداری و مهار سیلاب تمام پهنه مد نظر را پوشش ندهد. از این‌رو، استفاده از مدل‌هایی که سطح کمتری را به‌عنوان پهنه بحرانی شناسایی می‌کنند، در شرایطی که دقت پیش‌بینی مدل مناسب باشد، امکان برنامه‌ریزی عملیات اجرایی براساس نقشه استعداد سیل‌گیری بهتر فراهم خواهد شد.

جدول ۲- مساحت پهنه‌های استعداد سیل‌گیری آبخیز زرینه‌رود.

Table 2- Area of flood susceptibility zones in the Zarineh-Rud Watershed.

No.	Flood susceptibility zone	Area (ha)
1	Very low	437565.78
2	Low	6598.98
3	Medium	2764.1
4	High	1296.36
5	Very high	297.18

براساس نقشه پهنه‌بندی استعداد سیل‌گیری، زمین مناطق روستایی که در پهنه‌های استعداد زیاد و خیلی‌زیاد بودند، شناسایی شدند (جدول ۳). برخی از زمین روستاهای بست و عزیزآباد از توابع شهرستان دیواندره، روستاهای بیاندره و بادمجان از توابع شهرستان بانه در طبقه‌های استعداد سیل‌گیری خیلی زیاد بودند. همچنین، برخی از زمین روستاهای خوشه‌دره، اسحاق‌آباد، دره‌زیارت‌سفلی، طاله‌جر، هنگه‌چینه، سردره، پیرعمران، قشلاق‌پل، کانی‌بند، قهرمان، کانی‌نیاز، قشلاق‌رضا، قیلانتو، ماهیدرعلیا، خورخوره و میرگه‌نخشینه از توابع شهرستان سقز نیز در طبقه استعداد سیل‌گیری خیلی‌زیاد بودند. در برخی شرایط، محل روستا نیز در پهنه استعداد سیل‌گیری زیاد یا خیلی‌زیاد بود که اهمیت اقدامات مهار سیلاب در این شرایط دوچندان خواهد بود. با مشخص شدن موقعیت اجرای برنامه‌های مهار سیلاب فعالیت دستگاه‌های اجرایی مرتبط با موضوع سیلاب به این بخش‌های آبخیز زرینه‌رود معطوف خواهد شد. برنامه‌های مهار سیلاب شامل دو گروه نرم و سخت هستند. گروه نرم عمدتاً مربوط به برنامه‌های مدیریتی مانند رعایت حریم رودها و مشارکت جامعه‌های محلی در افزایش پوشش گیاهی و جلوگیری از تغییر کاربری

تأیید شد و قابلیت آن برای پیش‌بینی استعداد سیل در مقیاس آبخیز زیاد بود.

عملکرد مدل یادگیری ماشینی بیز ساده براساس مهمترین آماره ارزیابی (مساحت زیرمنحنی مشخصه عملکرد گیرنده) و براساس آماره‌های ثانویه ارزیابی

جدول ۴- نتایج اعتبارسنجی مدل برای پیش‌بینی استعداد سیل‌گیری در آبخیز زرينه‌رود.

Table 4- Results of model validation for flood susceptibility prediction in the Zarineh-Rud Watershed.

Evaluation approach	Evaluation criteria	Value
Cut-off independent threshold	AUC	0.936
	Accuracy	0.857
Cut-off dependent threshold	Precision	0.826
	Recall	0.904

آماره‌های اعتبارسنجی مستقل به آستانه و وابسته به آستانه اثبات شد که قابلیت مدل بیز ساده برای پیش‌بینی استعداد سیل‌گیری آبخیزها زیاد است. عملکرد این مدل براساس آماره‌های نامبرده عالی ارزیابی شد و نقشه استعداد سیل‌گیری آبخیز زرينه‌رود تولید شد. براساس تحلیل‌های انجام‌شده، حدود ۱۵۹۳ هکتار از این آبخیز از دیدگاه سیل‌گیری در وضعیت بحرانی بودند. از این‌رو، تمام برنامه‌های عملیاتی و مدیریتی در راستای مهار سیل باید به محدوده ۱۵۹۳ هکتار معطوف شود تا از هدررفت منابع و اجرای اقدامات پیشگیرانه سیل در مناطق با استعداد سیل‌گیری کمتر جلوگیری شود.

بر اساس یک نتیجه‌گیری کلی مبتنی بر تحلیل‌های کمی و علمی، در شرایط کمبود داده آبخیزهای کشور با استفاده از مدل یادگیری ماشینی بیز ساده برای پیش‌بینی مناطق مستعد سیل می‌توان نتایج مفیدی را ارائه داد و در شرایط معطوف‌شدن عملیات اجرایی و برنامه‌های مدیریتی مهار سیل در بخش‌های با استعداد سیل زیاد و خیلی‌زیاد، مدیریت سیل امکان‌پذیرتر خواهد شد. از این‌رو، پیشنهاد می‌شود در پژوهش‌های آینده، عملکرد مدل یادگیری ماشینی بیز ساده در دیگر آبخیزهای کشور برای شناسایی مناطق مستعد سیل بیشتر بررسی شود و تا جایی که ممکن باشد نقاط ضعف احتمالی آن برطرف شود. همچنین، پیشنهاد می‌شود از این مدل برای تهیه نقشه استعداد

بر اساس یافته‌های این پژوهش، کارایی مدل یادگیری ماشینی بیز ساده برای پیش‌بینی استعداد سیل در مقیاس آبخیز مناسب بود که با نتایج چن و همکاران (۲۰۲۰) و فم و همکاران (۲۰۲۰) هماهنگی دارد. اگرچه عملکرد این مدل در زمینه پیش‌بینی استعداد سیل مناسب بود، اما پژوهشگران در دیگر آبخیزها باید قابلیت‌های آن را با انجام پژوهش‌های بیشتر، معرفی کنند. براساس نتایج پژوهش ویکراماسینگه و کالتوراجی (۲۰۲۱)، می‌توان از مدل بیز ساده به‌دلیل قابلیت‌های زیادش در زمینه‌های علوم پیچیده برای تجزیه و تحلیل داده‌های حجیم^{۲۱} استفاده کرد.

نتیجه‌گیری و پیشنهادها

هر ساله در کشور، اثرات زیان‌بار اجتماعی و اقتصادی سیلاب‌ها (مانند خسارت‌های مالی و تلفات جانی) مشاهده می‌شود و مشکلات پرشماری را به‌طور مستقیم و غیرمستقیم بر ساکنان آبخیزهای کشور تحمیل می‌کند. با توجه به کمبود داده‌های آب‌شناختی و آبی در سطح آبخیز، در این پژوهش مدل بیز ساده به‌عنوان یک مدل یادگیری ماشینی برای پیش‌بینی استعداد سیل‌گیری استفاده شد و عملکرد آن براساس داده‌های حقیقی سیل‌گیری و آماره‌های اعتبارسنجی ارزیابی شد. نتایج نشان داد که سیل یک پدیده پیچیده است و براساس مشارکت عامل‌های محیطی پرشماری رخ می‌دهد. با استفاده از

تضاد منافع نویسندگان

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

دسترسی به داده‌ها

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

مشارکت نویسندگان

نویسنده اول: مفهوم‌سازی، پردازش داده، مدل‌سازی، انجام تحلیل‌های آماری، نگارش نسخه اولیه مقاله
نویسنده دوم: جمع‌آوری داده، مفهوم‌سازی، ویرایش و بازبینی مقاله، بررسی نتایج

سیل‌گیری براساس داده‌های سیلاب‌های تاریخی در دیگر آبخیزهای کشور استفاده شود.

سیاس‌گذاری

از حمایت‌های اداره کل منابع طبیعی و آبخیزداری، شرکت آب منطقه‌ای، اداره کل راهداری و حمل و نقل جاده‌ای، دفتر مدیریت بحران استانداری و جمعیت هلال احمر استان کردستان برای تهیه و جمع‌آوری اطلاعات و داده‌های موقعیت سیل‌قردانی و تشکر می‌شود.

فهرست منابع

- Aalami M, Ardestani M, Malekmohammadi B. 2024. Flood potential modeling in Zarineh Rood watershed using artificial intelligence models. *Watershed Management Research*. 37(1): 2-17. (In Persian). DOI: 10.22092/WMRJ.2023.360973.1513
- Abedi R, Costache R, Shafizadeh-Moghadam H, Pham QB. 2022. Flash-flood susceptibility mapping based on XGBoost, random forest and boosted regression trees. *Geocarto International*. 37(19): 5479-5496. DOI: 10.1080/10106049.2021.1920636
- Arabameri A, Saha S, Mukherjee K, Blaschke T, Chen W, Ngo PTT, Band SS. 2020. Modeling spatial flood using novel ensemble artificial intelligence approaches in northern Iran. *Remote Sensing*. 12(20): p.3423. DOI: 10.3390/rs12203423
- Arabameri A, Seyed Danesh A, Santosh M, Cerda A, Chandra Pal S, Ghorbanzadeh O, Chowdhuri I. 2022. Flood susceptibility mapping using meta-heuristic algorithms. *Geomatics, Natural Hazards and Risk*. 13(1): 949-974. DOI: 10.1080/19475705.2022.2060138
- Barati GR, Bodagh Jamali J, Maleki N. 2012. Anticyclones and Heavy Rainfalls over Western Iran. *Physical Geography Research*. 44(2): 85-98. (In Persian). DOI: 10.22059/jphgr.2012.29208
- Berrar, D. 2025. Bayes' Theorem and Naive Bayes Classifier. *Encyclopedia of Bioinformatics and Computational Biology* (Second Edition). pp. 483-494. DOI: 10.1016/B978-0-323-95502-7.00118-4
- Chan JYL, Leow SMH, Bea KT, Cheng WK, Phoong SW, Hong ZW, Chen YL. 2022. Mitigating the multicollinearity problem and its machine learning approach: A review. *Mathematics*. p.1283. DOI:10.3390/math10081283
- Chowdhury ME, Islam AS, Zzaman RU, Khadem S. 2025. A machine learning-based approach for flash flood susceptibility mapping considering rainfall extremes in the northeast region of Bangladesh. *Advances in Space Research*. 75(2):1990-2017. DOI: 10.1016/j.asr.2024.10.047
- Chen J, Huang G, 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*. p.112810. DOI: 10.1016/j.jenvman.2021.112810
- Chen W, Li Y, Xue W, Shahabi H, Li S, Hong H, Wang X, Bian H, Zhang S, Pradhan B, Ahmad BB. 2020. Modeling flood susceptibility using data-driven approaches of naive bayes tree, alternating decision tree, and random forest methods. *Science of the Total Environment*.

- p.134979. DOI: 10.1016/j.scitotenv.2019.134979
- Choubin B, Hosseini FS, Rahmati O, Youshanloei MM. 2023. A step toward considering the return period in flood spatial modeling. *Natural Hazards*. 115(1): 431-460. DOI: 10.1007/s11069-022-05561-y
- Choubin B, Sajedi Hosseini F, Rahmati O. 2025. Spatial Analysis of Vulnerability in Zarrinehrood Watershed to Flood Occurrence. *Iranian Journal of Watershed Management Science* 18(67): 73-86. (In Persian). DOI: 10.22034/18.67.6
- Dutta P, Deka S. 2024. A novel approach to flood risk assessment: Synergizing with geospatial based MCDM-AHP model, multicollinearity, and sensitivity analysis in the Lower Brahmaputra Floodplain, Assam. *Journal of Cleaner Production*. p.142985. DOI: 10.1016/j.jclepro.2024.142985
- Fratini P, Crosta G, Carrara A. 2010. Techniques for evaluating the performance of landslide susceptibility models. *Engineering geology*. 111(1-4): 62-72. DOI: 10.1016/j.enggeo.2009.12.004
- Habibi A, Delavar MR, Sadeghian MS, Nazari B. 2023. Flood Susceptibility Mapping and Assessment Using Regularized Random Forest and NAÏVE Bayes Algorithms. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 10: 241-248. DOI: 10.5194/isprs-Annals-X-4-W1-2022-241-2023
- Hosseini SA, Ahmadi H, Houshyar M. 2021. Analysis of Rainfall System, Leading to Floods in July 2015 in Saghez City in Kurdistan Province. *Integrated Watershed Management*, 1(1): 45-62. (In Persian). DOI: 10.22034/iwm.2021.247942
- Khajehnejad R, Bahremand A, Mohammadrezaei M. 2024. An overview of integrated flood management in Iran: current status and challenge. *Journal of Aquifer and Qanat*. 5(1): 183-202. (In Persian). DOI: 10.22077/jaaq.2024.8149.1077
- Maswadi K, Ghani NA, Hamid S, Rasheed MB. 2021. Human activity classification using Decision Tree and Naïve Bayes classifiers. *Multimedia Tools and Applications*. 80: 21709-21726. DOI: 10.1007/s11042-020-10447-x
- Moazzam MFU, Lee BG, Rahman AU, Farid N, Rahman G. 2020. Spatio-statistical analysis of flood susceptibility assessment using bivariate model in the floodplain of river swat, district charsadda, Pakistan. *Journal of Geoscience and Environment Protection*. 8(05): 159. DOI: 10.4236/gep.2020.85010
- Pajila PB, Sheena BG, Gayathri A, Aswini J, Nalini M. 2023, September. A comprehensive survey on naive bayes algorithm: Advantages, limitations and applications. In 2023 4th International Conference on Smart Electronics and Communication (ICOSEC). pp. 1228-1234. DOI: 10.1109/ICOSEC58147.2023.10276274
- Pal S, Singha P. 2022. Analyzing sensitivity of flood susceptible model in a flood plain river basin. *Geocarto International*. 37(24):7186-7219. DOI: 10.1080/10106049.2021.1967464
- Pham BT, Phong TV, Nguyen HD, Qi C, Al-Ansari N, Amini A, Ho LS, Tuyen TT, Yen HPH, Ly HB, Prakash I. 2020. A comparative study of kernel logistic regression, radial basis function classifier, multinomial naïve bayes, and logistic model tree for flash flood susceptibility mapping. *Water*. p.239. DOI: 10.3390/w12010239
- Qi W, Ma C, Xu H, Chen Z, Zhao K, Han H. 2021. A review on applications of urban flood models in flood mitigation strategies. *Natural Hazards*. 108: 31-62. DOI: 10.1007/s11069-021-04715-8
- Rahmati O, Darabi H, Panahi M, Kalantari Z, Naghibi SA, Ferreira CSS, Kornejady A, Karimidastenaie Z, Mohammadi F, Stefanidis S, Tien Bui D. 2020. Development of novel hybridized models for urban flood susceptibility mapping. *Scientific Reports*. p.12937. DOI: 10.1038/s41598-020-69703-7
- Rahmati O, Kornejady A, Choubin B, Jaafari A, Amini A. 2024. Evaluating the performance of the flexible discriminant analysis model in predicting the flooding potential of the Zarrineh-Rood Watershed. *Water and Soil Management and Modelling*. 4(3): 269-284. DOI: 10.22098/mmws.2023.13102.1303
- Rahmati O, Kornejady A, Samadi M, Deo RC, Conoscenti C, Lombardo L, Dayal K, Taghizadeh-Mehrjardi R, Pourghasemi HR, Kumar S, Bui DT. 2019. PMT: New analytical framework for automated evaluation of geo-environmental modelling approaches. *Science of the Total Environment*. 664: 296-311. DOI: 10.1016/j.scitotenv.2019.02.017
- Rajabizadeh Y, Ayyoubzadeh SA, Zahiri A. 2019. Flood Survey of Golestan Province in 2018-2019 and Providing Solutions for Its Control and Management in the Future. *Journal of Ecohydrology*. 6(4): 921-942. (In Persian). DOI: 10.22059/ije.2019.283004.1137
- Samanta S, Pal DK, Palsamanta B. 2018. Flood susceptibility analysis through remote sensing, GIS and frequency ratio model. *Applied Water Science*. p. 66. DOI: 10.1007/s13201-018-0710-1

Seleem O, Ayzel G, de Souza ACT, Bronstert A, Heistermann M. 2022. Towards urban flood susceptibility mapping using data-driven models in Berlin, Germany. *Geomatics, Natural Hazards and Risk*. 13(1): 1640-1662. DOI: 10.1080/19475705.2022.2097131

Siahkamari S, Haghizadeh A, Zeinivand H, Tahmasebipour N, Rahmati O. 2018. Spatial prediction of flood-susceptible areas using frequency ratio and maximum entropy models. *Geocarto International*. 33(9): 927-941. DOI: 10.1080/10106049.2017.1316780

Sulaiman MS, Abood MM, Sinnakaudan SK, Shukor MR, You GQ, Chung XZ. 2021. Assessing and solving multicollinearity in sediment transport prediction models using principal component analysis. *ISH Journal of Hydraulic Engineering*. 27(sup1). pp. 343-353. DOI: 10.1080/09715010.2019.1653799

Tehrany MS, Kumar L. 2018. The application of a Dempster-Shafer-based evidential belief function in flood susceptibility mapping and comparison with frequency ratio and logistic regression methods. *Environmental Earth Sciences*. 77: 1-24. DOI: 10.1007/s12665-018-7667-0

Traoré K, Fowe T, Ouédraogo M, Zorom M, Bologo/Traoré M, Toé P, Karambiri H. 2024.

Mapping urban flood susceptibility in Ouagadougou, Burkina Faso. *Environmental Earth Sciences*. p. 561. DOI: 10.1007/s12665-024-11871-0

Thapa PS, Chaudhary S, Dasgupta P. 2022. Contribution of integrated watershed management (IWM) to disaster risk reduction and community development: Lessons from Nepal. *International Journal of Disaster Risk Reduction*. p.103029. DOI: 10.1016/j.ijdrr.2022.103029

Wang H, Wang H, Wu Z, Zhou Y. 2021. Using multi-factor analysis to predict urban flood depth based on Naive Bayes. *Water*. p. 432. DOI: 10.3390/w13040432

Wickramasinghe I, Kalutarage H. 2021. Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation. *Soft Computing*. 25(3):2277-2293. DOI: 10.1007/s00500-020-05297-6

Widya LK, Rezaie F, Lee W, Lee CW, Nurwatik N, Lee S. 2024. Flood susceptibility mapping of Cheongju, South Korea based on the integration of environmental factors using various machine learning approaches. *Journal of Environmental Management*. p.121291. DOI: 10.1016/j.jenvman.2024.121291



Spatial Prediction of Flood Susceptibility in the Zarineh-Rud Watershed using a Simple Bayes Machine-Learning Model

Farnoush Mohammadi ^{1*}, Aref Bahmani ²

1- Ph.D. Student, Department of Arid and Mountain Reclamation Engineering, Faculty of Natural Resources, University of Tehran, Karaj, Iran

2- Deputy of Watershed Management, General Directorate of Natural Resources and Watershed Management of Kurdistan Province, Sanandaj, Iran

Extended Abstract

Introduction and Goal

Floods cause numerous financial losses and loss of life every year, and have a devastating impact on the sustainable development of the country. However, generally, actual flood events have not been used in spatial analysis and modeling, and flood susceptibility maps have been prepared solely based on expert opinions and multi-criteria decision-making methods. On the other hand, management plans and actions of executive agencies are usually developed without considering the different flood-prone areas of watersheds. This study was conducted with the aim of utilizing observational data of flooding events and determining the efficiency of the Naive Bayes model in spatial prediction of flooding susceptibility in the Zarineh-Rud watershed.

Materials and Methods

The flood events database was prepared based on flood event information in the Zarineh-Rud watershed recorded by the Provincial Disaster Management Office and the Regional Water Company. Given that various environmental factors play a role in the formation of floods and inundation of lands adjacent to rivers, flood modeling is not possible without considering them. Therefore, after reviewing various sources, thirteen environmental factors affecting flooding were selected, including land elevation, slope direction, drainage density, land use, lithology, surface curvature, cross-sectional curvature, average annual rainfall, slope percentage, soil texture, flow power index, distance from the watercourse, and topographic moisture index. The multicollinearity of environmental factors was examined using the tolerance factor statistic. Raster layers of environmental factors were introduced as independent variables into the Naive Bayes model. The flood event locations were divided into two groups, training and validation, based on the spatial random method with a ratio of 30 and 70 %. After running the model, a flood susceptibility map of the Zarineh-Rud watershed was produced, in which each cell represents the probability of flooding in that area. The accuracy of the flood susceptibility map was evaluated using independent and threshold-dependent statistics and validation group data.

Article Type: Research Article

***Corresponding Author s' E-mail:** farnoushmohammadi@ut.ac.ir

Citation: Mohammadi, F., Bahmani, A. 2026 Spatial Prediction of Flood Susceptibility in the Zarineh-Rud Watershed using a Simple Bayes Machine-Learning Model. Watershed Management Research. 38(4): 108-127.

DOI: 10.22092/wmrj.2025.369572.1623

Received: 25 May 2025, **Received in revised form:** 01 June 2025, **Accepted:** 21 June 2025

Published online: 01 January 2026

Watershed Management Research, Vol.38, No.4, Ser. No: 149, Winter 2026, pp. 108-127.

Publisher: Fars Agricultural and Natural Resources and Education Center

©Author(s)



Results and Discussion

Based on the results of this study, the independent variables considered were without multicollinearity and could be used as predictors in the modeling process. The validation results based on the area under the receiver operating characteristic curve statistic showed that the flood susceptibility map has an accuracy of 93.6%. According to the threshold-dependent statistics, the efficiency of the Naive Bayes machine learning model was obtained as 85.7 based on the Accuracy statistic, 82.6% based on the Precision statistic, and 90.4% based on the Recall statistic.

Conclusion and Suggestions

The performance of the Naive Bayes machine learning model was suitable for spatial prediction of flood susceptibility at the watershed scale and various variables can be used for spatial analysis. The flood susceptibility map should be considered as the basis for planning river regulation operations (such as building coastal walls and removing bed layers), flood management (such as respecting river boundaries), and watershed management (such as building watershed management structures in the upstream of flood-prone areas) in the Zarineh-Rud watershed. Therefore, it is suggested that this model be used to prepare flood susceptibility maps based on historical flood data in other basins of the country.

Keywords: Crisis management, flood, Kurdistan Province, modeling, spatial analysis

Article Type: Research Article

Acknowledgement

We would like to appreciate the General Directorate of Natural Resources and Watershed Management, the Regional Water Company, the General Directorate of Road Administration and Road Transportation, the Governorate's Crisis Management Office, and the Kurdistan Province Red Crescent Society for providing information and data on the flood situation.

Conflicts of interest

The authors of this article declared no conflict of interest regarding the authorship or publication of this article.

Data Availability Statement

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

Authors' Contribution

Author 1: Writing - original draft preparation

Author 2: Resources, Software, Manuscript editing