"مقاله پژوهشی"

ارزیابی کارایی مدل‌های یادگیری ماشین در تهیه نقشه احتمال خطر سیل

محمدتقی آوند[۱]، سعید جانی‌زاده[۲] و فائزه جعفری[۳]

۱- دانشجوی دکتری آبخیزداری، دانشکده منابع طبیعی و علوم دریایی، دانشگاه تربیت مدرس، پردیس نور، ایران
(نویسنده مسوول: Mt.avand70@gmail.com)
۲- دانشجوی دکتری آبخیزداری، دانشکده منابع طبیعی و علوم دریایی، دانشگاه تربیت مدرس، پردیس نور، ایران
۳- دانشجوی دکتری آبخیزداری، دانشکده منابع طبیعی و علوم دریایی، دانشگاه تربیت مدرس، پردیس نور، ایران
تاریخ دریافت: ۱۳۹۹/۰۶/۱۵ تاریخ پذیرش: ۱۳۹۹/۰۷/۲۳
صفحه: ۱۹ تا ۳۲

چکیده

سیل یکی از مخرب‌ترین بلایای طبیعی است که هر ساله باعث تلفات مالی و جانی می‌شود. بنابراین تولید نقشه حساسیت برای مدیریت سیل و کاهش اثرات زیانبار آن ضروری است. پژوهش حاضر به منظور تهیه نقشه حساسیت به وقوع سیل با استفاده از مدل‌های داده‌کاوی شامل جنگل تصادفی (Random Forest) و ماشین گرادیان تقویتی (Gradient Boosting Machine) انجام گرفت. ابتدا ۲۷۵ موقعیت مکانی سیل و ۲۷۵ موقعیت مکانی غیرسیل در حوضه کمیجان استان مرکزی شناسایی شد. موقعیت‌های مکانی سیل‌گیر به صورت تصادفی به ۷۰ درصد (۱۹۰ موقعیت) و ۳۰ درصد (۸۲ موقعیت) به ترتیب برای مدلسازی و اعتبارسنجی تقسیم گردید. سپس ۱۲ فاکتور موثر بر وقوع سیل که شامل، شیب، جهت، ارتفاع، بارندگی، کاربری اراضی، فاصله از رودخانه، تراکم زهکشی، شکل شیب، انحنای شیب، سنگ شناسی، خاک و شاخص قدرت جریان می‌باشند، تعیین شدند. برای ارزیابی مدل‌های به کار رفته منحنی ROC مورد استفاده قرار گرفت. نتایج نشان داد که در مرحله اعتبارسنجی، سطح زیر منحنی برای مدل‌های RF و GBM به ترتیب ۰/۸۳ و ۰/۷۵ درصد بوده است که نشان‌دهنده صحت بیشتر مدل RF در تهیه نقشه حساسیت به وقوع سیل می‌باشد. مهم‌ترین فاکتورهای موثر در سیل در حوزه آبخیز کمیجان به ترتیب بارندگی، فاصله از رودخانه و ارتفاع می‌باشند.

واژه‌های کلیدی: پهنه‌بندی سیلاب، جنگل تصادفی، مدل‌های داده کاوی، منطقه کمیجان، GBM

مقدمه

هر ساله بلایای طبیعی مانند زمین لغزش، زلزله، سیلاب و غیره باعث تلفات فراوان جانی و مالی در سراسر جهان می گردد (۴۳،۲۵)، که سیلاب مخرب‌ترین آنها در نظر گرفته شده است (۴۵). سیل یکی از بلایای طبیعی است که خسارت ناشی از آن قابل شمارش نیست (۱۰). سیل در فواصل مختلف با مدت‌های مختلف رخ می‌دهند. سیل باعث خسارت جدی به محیط زیست، حمل و نقل، اقتصاد، کشاورزی و زندگی مردم می‌شود (۲۲،۴۵). بنابراین شناسایی مناطق حساس به وقوع سیل جهت فراهم کردن مدیریت لازم برای کاهش خسارت سیل ضروری می‌باشد (۲). تهیه نقشه حساسیت به وقوع سیل به عنوان مرحله ضروری برای جلوگیری و مدیریت سیل‌های آینده شناخته شده است (۱۸). هرچند که وقوع سیل دارای شرایط پیچیده‌ای می‌باشد که وقوع آن را برای یک پیش‌بینی قابل اطمینان، مشکل می‌سازد (۴۱).

علی‌رغم تلاش‌های متخصصان، تصمیم گیران، ذینفعان و ادارات دولتی در دهه‌های اخیر برای کاهش اثرات سیلاب، تعداد حوادث و تلفات اقتصادی و انسانی مرتبط با آن، در سراسر جهان در حال افزایش است. این پدیده نه تنها در کشورهای در حال توسعه، بلکه در تمام جهان شایع‌ترین مخاطره طبیعی است (۲۰). مسئله تهیه نقشه حساسیت به وقوع سیل توسط محققان مختلف بررسی شده است (۴۲،۳۵). دسترسی سریع به ماهواره بر پایه داده های سنجش از دور و بهبود روش‌های تجاری، استفاده از سیستم اطلاعات جغرافیایی را در تهیه نقشه حساسیت به وقوع سیل افزایش داده است. بنابراین GIS یک ابزار مفید برای بررسی وقایع چند بعدی مانند سیلاب می‌باشد (۴۱). دامنه وسیعی از تکنیک‌های مدلسازی در ارزیابی بلایای طبیعی پیشنهاد و مورد استفاده واقع شده است. محققان مختلفی ارزیابی نقشه‌های سیلاب توسط GIS انجام داده‌اند و اطلاعات مفیدی را در ارتباط با بعضی از روش‌های موجود جمع آوری نمودند (۶،۲۲). روش‌های مختلف تهیه نقشه حساسیت به وقوع سیل مانند مدل‌های آماری و احتمالاتی در تحقیق‌های مختلف و موارد مطالعاتی مختلفی انجام شده است (۱۹،۲۱). در سال‌های اخیر تکنیک‌های پیشرفته‌تری در ارزیابی نقشه‌های حساسیت سیلاب توسط روش‌های یادگیری ماشین و داده‌کاوی اجرا شده است.

به طور کلی، به دلیل اینکه حوضه‌ها به طور ذاتی پیچیده هستند، مدلسازی آنها با روش‌های هیدرولوژی ساده و خطی مطابقت ندارد (۳۷). به همین دلیل تکنیک‌های مختلفی برای بررسی سیل که رفتار چند بعدی دارد مورد استفاده قرار می‌گیرد. در میان تکنیک‌های مختلف، استفاده از GIS (۱۵،۲۹)، آنالیز سلسله مراتبی (۷)، نسبت فراوانی (۳۵)، رگرسیون لجستیک (۱۷)، منطق فازی (۳۱)، جنگل تصادفی (۵،۳۸)، شبکه عصبی مصنوعی (۲۶) و ماشین بردار پشتیبان (۸) مورد علاقه محققان می‌باشند. Kia و همکاران (۲۲)، از مدل ANN برای شبیه سازی مناطق مستعد سیل در حوضه رود جوهور، مالزی استفاده کردند. آنها اظهار داشتند که این تکنیک قادر است با عدم اطمینان‌های موجود در داده‌های ورودی مقابله کند و اطلاعات را از مجموعه داده‌های ناقص و متناقض استخراج کند. تهرانی و همکاران (۵۰)، به بررسی

[Downloaded from drnl.sanru.ac.ir on 2025-12-19]
[DOR: 20.1001.1.27174425.1399.1.1.4.3]

کارآیی ترکیب مدل SVM با WOE جهت تهیه نقشه حساسیت به وقوع سیل پرداختند. آنها بیان کردند که در ترکیب این مدلها با یکدیگر نتایج از صحت بالاتری برخوردار می‌باشد. در تابع WOE- RBF_SVM سطح زیرمنحنی برای نرخ موفقیت و نرخ پیش بینی به ترتیب برابر با ۹۶/۴۸ درصد و ۹۵/۶۷ درصد می‌باشد.

بویی و همکاران (۴)، از دو رویکرد جدید هوش مصنوعی مبتنی بر GIS سیستم استنتاج عصبی فازی تطبیقی شامل الگوریتم رقابتی امپریالیستی (ICA) و الگوریتم کرم شب‌تاب (FA) به منظور مدلسازی مکانی سیلاب در حوزه آبخیز هراز در استان مازندران استفاده کردند. به منظور اعتبارسنجی مدل‌ها از شاخص‌های آماری خطا RMSE و MSE، آزمون‌های آماری فریدمن و ویلکاکسون و سطح زیرمنحنی (AUC) استفاده شد. نتایج نشان‌دهنده دقت پیش‌بینی مناسب هر دو گروه بود، با این حال؛ مدل ANFIS-ICA (AUC=0.947) در مقایسه با دیگر مدل‌های مورد استفاده عملکرد بهتری داشت. بنابراین مدل برتر می‌تواند به عنوان یک روش امیدوارکننده برای مدیریت پایدار مناطق مستعد سیل معرفی شود.

خسروی و همکاران (۱۹)، به منظور تهیه نقشه حساسیت سیل در حوزه آبخیز هراز از چهار مدل یادگیری ماشین LMT، REPT، NBT و ADT استفاده کردند.برای این منظور از ۲۰۱ موقعیت مکانی سیلاب و ۱۱ عامل تاثیرگذار بر سیلاب به کار برده شد. همچنین از منحنی ROC و آزمون‌های رتبه‌بندی فریدمن و ویلکاکسون برای اعتبارسنجی و مقایسه قابلیت پیش‌بینی مدل‌ها استفاده شد. نتایج نشان داد که مدل ADT بیشترین قابلیت پیش‌بینی را برای ارزیابی حساسیت سیل دارد و مدل‌های LMT، NBT و REPT به ترتیب به دنبال آن بیشترین دقت را دارند.

یوسف و همکاران (۵۴) به پهنه‌بندی نقشه حساسیت به وقوع سیل شهر جده[1] در عربستان صعودی با روش‌های FR و LR و ترکیب آنها پرداختند. آنها بیان نمودند که روش ترکیبی، با سطح زیرمنحنی ۹۱/۳ درصد از روش FR با سطح زیر منحنی ۸۹/۶ درصد دارای صحت و اعتبار بیشتری می‌باشد. تحقیقی تحت عنوان نقشه حساسیت به وقوع سیل توسط ترکیب مدل‌های آماری دومتغیره و چند متغیره (نسبت فراوانی و رگرسیون لجستیک) در کشور کره جنوبی صورت گرفت. نتایج نرخ موفقیت و نرخ پیش‌بینی به ترتیب برابر با ۹۲/۷ و ۸۲/۳ درصد می‌باشند. آنها بیان نمودند که این روش کارایی لازم برای تهیه نقشه‌های حساسیت به وقوع سیل را دارا می‌باشد (۴۲). با هدف تهیه نقشه حساسیت سیل در حوزه آبخیز هراز از چهار مدل منفرد و ترکیبی نسبت فراوانی، شواهد وزنی، فرآیند تحلیل سلسله مراتبی و ترکیب تحلیل سلسله مراتبی و نسبت فراوانی (FR-AHP) استفاده شد. ۲۱۱ مکان سیل به دو بخش شامل ۷۰ درصد (۱۵۱ موقعیت برای مدلسازی) و ۳۰ درصد (۶۰ موقعیت برای اعتبار سنجی) تقسیم شدند. ۱۰ فاکتور تاثیرگذار بر سیلاب نیز انتخاب شدند. اعتبارسنجی نتایج نشان داد که مقدار AUC برای FR، AHP، WofE و FR-AHP به ترتیب ۹۷/۰۷، ۹۸/۹۶ و ۹۵/۹۱ و ۸۶,۱۹ درصد بود. بنابراین نتایج نشان داد مدل نسبت فراوانی بیشترین مقدار AUC را در بین مدل‌ها دارا می‌باشد (۱۸).

پهنه‌بندی مناطق حساس به وقوع سیل در مالزی توسط تکنیک SVM با ۴ تابع Linear, Polynomial, Radial Basis Function, Sigmoid مورد بررسی قرار گرفت. آنها بیان کردند که مساحت سطح زیر منحنی (AUC) به ترتیب برای توابع فوق برابر با ۸۴/۶۳، ۸۳/۹۲، ۸۴/۹۷ و ۸۱/۸۸ درصد است. آنها توسط شاخص Cohens Kappa بیان کردند که کل فاکتورهای در نظر گرفته شده به جز رواناب سطحی (که باعث کاهش صحت نتایج نهایی می‌شود) دارای تاثیر مثبت در سیلاب می‌باشند. طبق نتایج آنها شیب و طبقات ارتفاعی در همه توابع از موثرترین فاکتورها بودند (۴۱). فرانک و همکاران (۱۵)، به منظور اندازه گیری و مدل‌سازی بار رسوب در حوزه‌های با اندازه متوسط از منحنی سنجه، و مدل‌های خطی تعمیم یافته (GLM) و رگرسیون ناپارامتری و برای بررسی عدم قطعیت از مدل‌های جنگل تصادفی و جنگل رگرسیون چندک استفاده کرد. نتایج نشان داد روش‌های منحنی سنجه رسوب معمولی و مدلهای خطی تعمیم یافته با وجود در نظر گرفتن متغیرهایی مانند شدت بارش، پیشبینی مناسبی انجام ندادند اما دو روش QRF و RF توانستند بادقت زیادی رسوب را تخمین بزنند. همچنین RF و QRF عدم قطعیت مقدار رسوب برآورد شده را نیز محاسبه می‌کنند (۱۲).

خسروی و همکاران (۱۹)، به منظور مدلسازی حساسیت سیل در یکی از مهمترین مناطق مستعد سیل در چین یعنی حوزه آبخیز نینگدو با استفاده از روش های تصمیم گیری چند معیاره VIKOR، TOPSIS و SAW پرداختند. ۱۲ عامل موثر بر سیلاب به عنوان ورودی‌های مدل استفاده شدند. از منحنی ROC نیز به منظور ارزیابی مدل‌ها استفاده شد. علارغم دقت عالی (AUC>0.95) همه مدل‌ها در پیش‌بینی سیلاب، مدل NBT بهترین عملکرد را داشت (AUC=0.98).

فام و همکاران (۳۵)، در مطالعه‌ای به منظور تعیین مناطق مخاطره آمیز در برابر سیل در حوضه رودخانه اوجان چای جهت فراهم نمودن پایه‌هایی برای آنالیز ریسک و آسیب‌پذیری سیل در آینده انجام شد. برای این منظور ۱۶ لایه اطلاعاتی تهیه شد و به روش جمع جبری فازی روی‌هم گذاشته شد و به روش C Mean بر مبنای داده‌های ژئومورفولوژی مورد طبقه‌بندی فازی قرار گرفتند. نتایج این بررسی، صحت بیش از ۷۰ درصد (با ضریب کاپای ۰/۷۵) را در مقایسه با داده‌های مبنا نشان می‌دهد که نشانگر قابل قبول بودن استفاده از این شیوه در تهیه نقشه مخاطرات سیل است.

مطالعه حاضر با هدف پهنه‌بندی حساسیت به سیل در حوزه آبخیز کمیجان در مرز استان‌های مرکزی و همدان انجام گرفته است. تهیه نقشه حساسیت به سیل از جمله روش‌های جدیدی می‌باشد که در سال‌های اخیر مورد توجه محققان قرار گرفته است. در این پژوهش با استفاده از دو روش داده‌کاوی شامل مدل RF و GBM در محیط نرم‌افزار

1- Jaddah

[Downloaded from drnl.sanru.ac.ir on 2025-12-19]

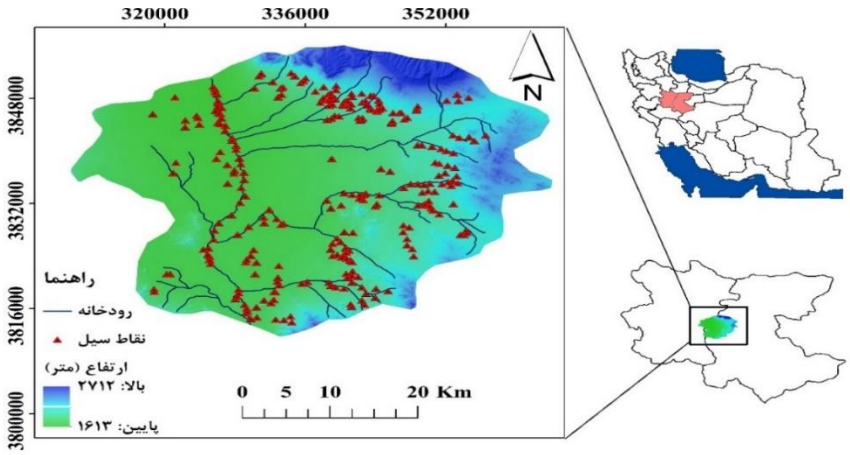
[DOR: 20.1001.1.27174425.1399.1.1.4.3]

R، برای اولین بار نقشه مناطق دارای احتمال رخداد سیل در حوزه آبخیز کمیجان تهیه شده است.

مواد و روش‌ها

منطقه مورد مطالعه

حوضه کمیجان در شمال غربی استان مرکزی واقع شده است که با مرکز استان (اراک) ۹۵ کیلومتر فاصله دارد. منطقه مورد مطالعه بین عرض جغرافیایی ۳۴ درجه و ۲۵ دقیقه تا ۳۴ درجه و ۴۸ دقیقه شمالی و طول جغرافیایی ۴۹ درجه و ۳۰ درجه تا ۳۸ درجه و ۳۰ دقیقه شرقی قرار دارد. مساحت منطقه مورد مطالعه تقریبا ۱۶۰ کیلومتر مربع می‌باشد. کمترین ارتفاع این منطقه ۱۶۱۳ متر و بیشترین ارتفاع ۲۷۱۲ متر از سطح دریا می‌باشد. منطقه مورد مطالعه با میانگین بارندگی ۲۲۵ میلی‌متر دارای اقلیم خشک تا نیمه خشک می‌باشد. ۷۱ درصد از بارندگی‌ها در ماه‌های دی و بهمن اتفاق می‌افتد. حداکثر بارش ۲۴ ساعته در این حوضه ۴۷ میلی‌متر است. اقلیم منطقه کمیجان در جنوب و غرب از منطقه سرد استان همدان تأثیر گرفته و در جنوب به جهت وجود دشت‌های هموار از هوای خنک‌تری بهره‌مند است. حداکثر درجه حرارت در تابستان ۳۵ درجه سانتی‌گراد است. پوشش غالب در این منطقه کشاورزی، بایر و مراتع می‌باشد که بیشترین مساحت مربوط به کشاورزی است.



شکل ۱- موقعیت منطقه مورد مطالعه
Figure 1. Location of the study area

تهیه نقشه نقاط سیل

نقاط سیل یک سطح مهم از رابطه بین رخداد سیل و عوامل به وجود آورنده آن می‌باشد. رویدادهای سیل تاریخی به عنوان مبنایی برای پیش‌بینی وقوع سیلاب در آینده به حساب می‌آیند. به طوری که مناطق نزدیک به رخدادهای گذشته حساسیت بالایی به سیل‌گیری دارند. بدین ترتیب، ۲۷۵ نقطه سیل در منطقه مورد مطالعه، توسط سازمان آب منطقه‌ای استان مرکزی ثبت شده است که ۷۰ درصد آن برای آموزش مدل و ۳۰ درصد برای اعتبار سنجی مدل کنار گذاشته شد (۳۳). ۲۷۵ نقطه غیرسیل نیز با استفاده از نقشه توپوگرافی و نرم‌افزار گوگل ارث (Google Earth) با توجه به مناطقی مانند تپه‌ها و کوه‌ها که سیلاب قادر به پیشروی در آنجا نیست، انتخاب گردید.

پارامترهای موثر در وقوع سیل

برای تهیه نقشه حساسیت به وقوع سیل و یا به طور کلی تولید مدلی برای ارزیابی در معرض آسیب بلایای طبیعی، مجموعه‌ای از فاکتورهای موثر باید تعریف گردد (۱۴،۳). نقشه رقومی ۱۲ پارامتر موثر بر خطر وقوع سیل شامل جهت، ارتفاع، فاصله از رودخانه، شیب، بارندگی، شاخص قدرت جریان، تراکم زهکشی، کاربری اراضی، خاک، سنگ‌شناسی، شکل شیب و همگرایی شیب با استفاده از نرم افزارهای ArcGIS 10.1، ENVI 5.1 و SAGAGIS 2 تهیه شدند و سپس به اندازه پیکسل ۱۲/۵ متر بر اساس DEM منطقه تبدیل شدند.

شیب زمین به دلیل تاثیر مستقیم بر رواناب سطحی و فرصت نفوذ، یکی از عوامل مهم در وقوع سیل حوزه‌های آبخیز به شمار می‌رود. به منظور تهیه نقشه شیب از مدل رقومی ارتفاع (با قدرت تفکیک ۱۲/۵ متر) و نرم‌افزار ArcGIS 10.1 استفاده شد. نقشه به دست آمده به ۶ کلاس (۰-۵، ۵-۸، ۸-۱۲، ۱۲-۲۰، ۲۰-۳۰ و ۳۰<) تقسیم‌بندی گردید. انحنای زمین و شکل شیب نیز بر وقوع سیل تاثیر زیادی دارد. این فاکتورها به سه کلاس محدب، مسطح و مقعر تقسیم‌بندی گردید. طبقات ارتفاعی فاکتور موثر دیگری بر وقوع سیل می‌باشد. این فاکتور تاثیر بالایی بر وقوع سیل دارد، زیرا که طبقات ارتفاعی پایین پتانسیل بالایی بر وقوع سیل دارند. بارندگی از جمله فاکتورهای بسیار با اهمیت در بررسی سیل می‌باشد. شدت و مدت بارندگی در زمان‌های مختلف می‌تواند سبب ایجاد سیل‌های بسیار شدید شود. در منطقه مورد مطالعه از بارندگی سالانه ایستگاه سینوپتیک کمیجان استفاده گردید. فاصله از رودخانه یکی از مهم‌ترین عوامل در سیل گرفتگی اراضی مجاور است. نقشه فاصله از رودخانه‌های مجاور بر اساس لایه رقومی شبکه جریان حوزه

[Downloaded from drnl.sanru.ac.ir on 2025-12-19]
[DOR: 20.1001.1.27174425.1399.1.1.4.3]

آبخیز کمیجان، در نرم‌افزار ArcGIS 10.1 تهیه گردید. SPI یا شاخص قدرت جریان یکی از شاخص‌های پارامترهای مهم در پتانسیل سیل‌خیزی حوضه‌های آبخیز به شمار می‌رود. شاخص توان آبراهه با توجه به رابطه زیر تعریف می‌گردد (۵)، برای تهیه این نقشه نیز از نرم افزار SAGA GIS 2 استفاده شد.

$$SPI= As \tan\beta \quad (۱)$$

زمین‌شناسی به دلیل تأثیر مستقیم بر میزان نفوذپذیری و رواناب سطحی، یکی از عوامل مهم در پدیده سیل حوزه‌های آبخیز است. نقشه زمین شناسی منطقه مذکور از نقشه کشوری جدا شد و با استفاده از نرم‌افزار ArcGIS تبدیل به نقشه رستری با اندازه پیکسل ۱۲/۵ متر شد و نقشه حاصل به ۱۱ طبقه تقسیم‌بندی گردید (جدول ۱). کاربری اراضی نتیجه روابط متقابل پارامترهای اجتماعی- فرهنگی و توان بالقوه سرزمین است. تغییرات در کاربری و پوشش اراضی نتایج چشمگیری در پتانسیل سیل‌خیزی حوزه‌های آبخیز دارد. نقشه کاربری اراضی حوزه کمیجان با استفاده از تصویر سنجنده ALOSPALSAR مربوط به ماهواره Landsat 8 تهیه شد. نقشه کاربری اراضی با استفاده از طبقه‌بندی نظارت شده در محیط نرم‌افزاری ENVI به پنج کلاس مرتع، زمین بایر، کشاورزی، باغ و مناطق شهری کلاس بندی گردید.

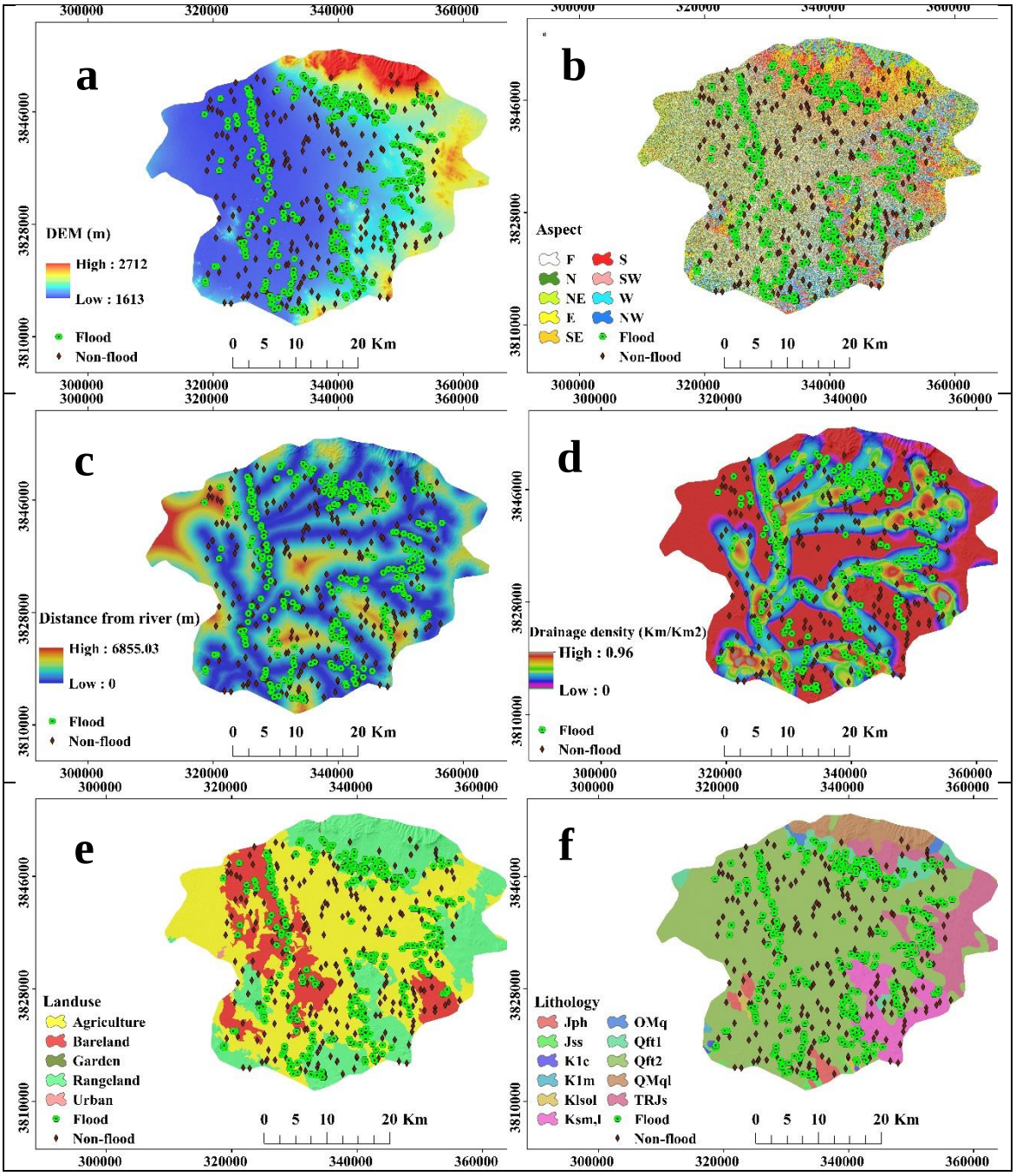
جدول ۱- کلاس‌های سنگ شناسی موجود در منطقه مورد مطالعه

Table 1. Lithological classes present in the study area

کد	سنگ شناسی	سن زمین شناسی
Jph	Dark grey shale and sandstone	ژوراسیک
Jss	Grey thick - bedded to massive orbitolina limestone	ژوراسیک
Klc	High level piedmont fan and vally terrace deposits	کرتازوئیک
Klm	Limestone, argillaceous limestone; tile red sandstone and gypsiferous marl	قبل از کرتازوئیک
Klsol	Limestone, marl, gypsiferous marl, sandy marl and sandstone	قبل از کرتازوئیک
Ksm,l	Low level pediment fan and valley terrace deposits	کرتازوئیک
OMq	Marl and calcareous shale with intercalations of limestone	الیگوسن-میوسن
Qft1	Massive to thick - bedded reefal limestone	کواترنری
Qft2	Phyllite, slate and meta-sandstone (Hamadan Phyllites)	کواترنری
QMq1	Red conglomerate and sandstone	الیگوسن-میوسن
TRJs	Sandstone	تریاسیک-ژوراسیک

[Downloaded from drnl.sanru.ac.ir on 2025-12-19]

[DOR: 20.1001.1.27174425.1399.1.1.4.3]

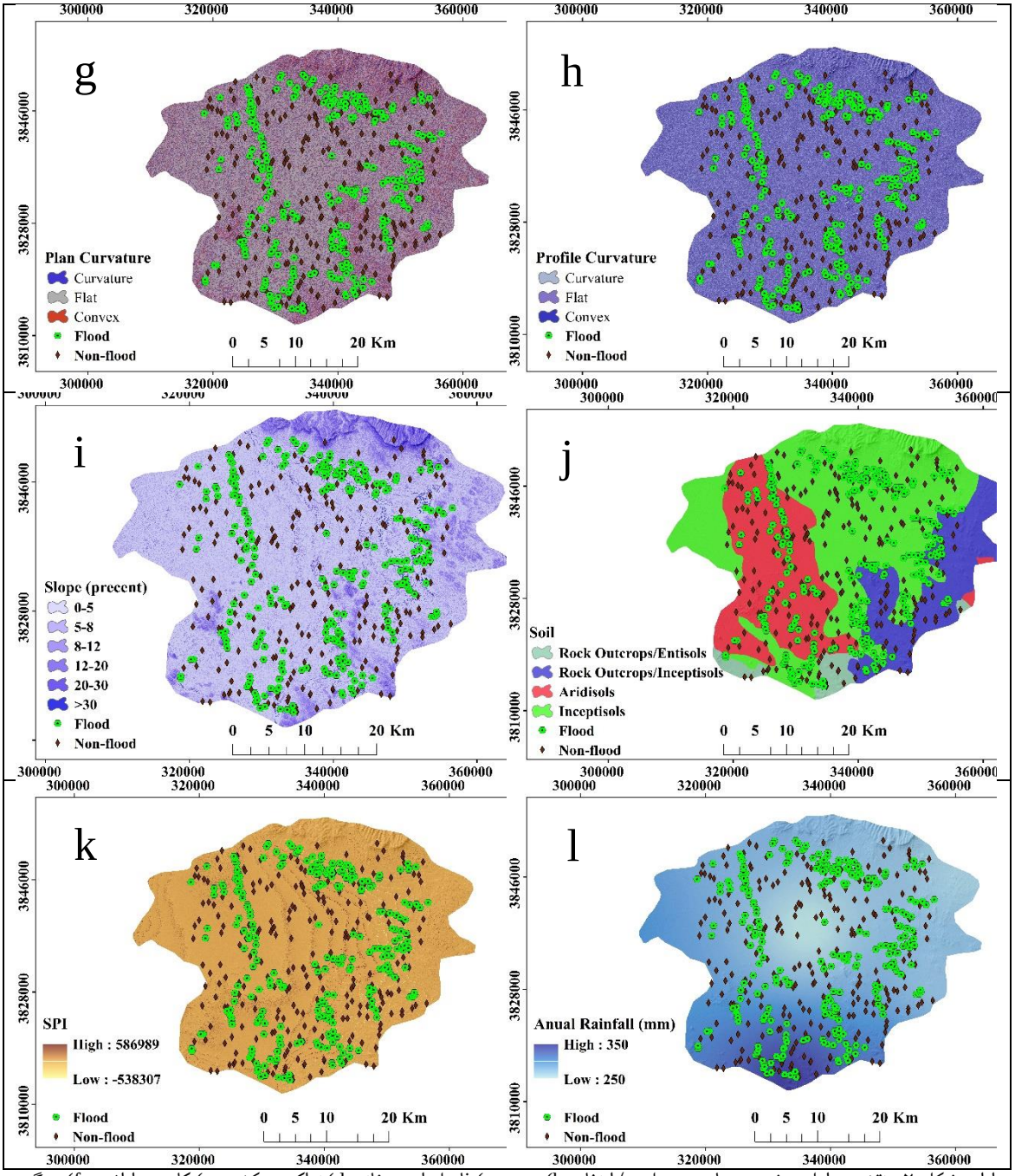


شکل ۲- نقشه عوامل موثر بر حساسیت سیل: a) ارتفاع، b) جهت، c) فاصله از رودخانه، d) تراکم زهکشی، e) کاربری اراضی، f) سنگ‌شناسی، g) شکل انحنا، h) نیمرخ انحنا، i) درصد شیب، j) بافت خاک، k) شاخص قدرت جریان، l) بارندگی

Figure 2. Factors affecting flood susceptibility: a) altitude, b aspect, c) distance from river, d) drainage density, e) land use, f) lithology, g) plan curvature, h) profile curvature, i) Slope, j) soil, k) stream power index, l) rainfall

[Downloaded from drnl.sanru.ac.ir on 2025-12-19]

[DOR: 20.1001.1.27174425.1399.1.1.4.3]



ادامه شکل ۲- نقشه عوامل موثر بر حساسیت سیل: a) ارتفاع، b) جهت، c) فاصله از رودخانه، d) تراکم زهکشی، e) کاربری اراضی، f) سنگ شناسی، g) شکل انحنا، h) نیمرخ انحنا، i) درصد شیب، j) بافت خاک، k) شاخص قدرت جریان، l) بارندگی

Continued Figure 2. Factors affecting flood susceptibility: a) altitude, b aspect, c) distance from river, d) drainage density, e) land use, f) lithology, g) plan curvature, h) profile curvature, i) Slope, j) soil, k) stream power index, l) rainfall

مدل‌های یادگیری ماشین تهیه نقشه حساسیت سیل

مدل جنگل تصادفی (RF)

جنگل تصادفی یک نوع مدرن از روش‌های درخت پایه است که شامل انبوهی از درخت‌های کلاس‌بندی و رگرسیونی می‌باشد همچنین یکی از روش‌های ناپارامتریک مناسب برای مدل سازی داده‌های پیوسته و گسسته روش درخت تصمیم است. از مشکلات این روش نوسانات بالای نتایج هر درخت است. به منظور کاهش این نوسانات و کاهش واریانس برآورد، روش جنگل تصادفی پیشنهاد شده است. جنگل تصادفی با استفاده از مجموعه‌ای از درخت‌ها با در نظر گرفتن n داده مشاهده‌ای مستقل ساخته می‌شود (۳۴).

$$(Y , X), i= 1,\ldots,n \qquad (۲)$$

این روش ترکیبی از چندین درخت تصمیم است که در ساخت آن چندین نمونه بوت استرپ از داده‌ها شرکت دارند و در ساخت هر درخت به طور تصادفی تعدادی از متغیرهای ورودی شرکت می‌کنند. با استفاده از روش بوت استرپ به

[Downloaded from drnl.sanru.ac.ir on 2025-12-19]

[DOR: 20.1001.1.27174425.1399.1.1.4.3]

تعداد زیاد (به عنوان مثال ۲۰۰۰ مرتبه) نمونه‌هایی n تایی از مجموعه داده‌های مشاهداتی اولیه، نمونه برداری همراه با جای‌گذاری می‌شوند. در طی فرآیند نمونه گیری حدود یک سوم از داده‌ها نمونه‌گیری نمی‌شوند و به عنوان نمونه خارج از کیسه (از این داده‌ها برای تعیین متغیرهای مهم و همچنین برآورد نااریب خطا استفاده می‌شود) در نظر گرفته می‌شوند. سپس بر روی هر نمونه بوت استرپ یک درخت گسترش داده می‌شود. در طی فرآیند ساخت درخت در هر شاخه، از بین تمام M متغیر مستقل به صورت تصادفی m متغیر برای تقسیم شدن انتخاب می‌شود. برای حالت رگرسیونی نسبت m/M برابر با یک سوم است و برای کلاسه بندی برابر با $m=\sqrt{M}$ پیشنهاد شده است.

پس از ساخت تمام درخت داده‌های تست به درخت معرفی شده و به تعداد درخت‌ها برای بردار ورودی یک خروجی به دست می‌آید. با میانگین‌گیری این خروجی‌ها، خروجی نهایی مدل و با در نظر گرفتن توزیع تجربی خروجی‌ها مقادیر صدک‌ها و دامنه عدم قطعیت محاسبه می‌شود. روش درخت رگرسیون جنگل تصادفی به ویژه هنگامی که تعداد مشاهدات در مقایسه با تعداد پیش‌بینی کننده‌ها نسبتاً کم باشد یک روش پیش بینی کارآمد است (۴۰). در این پژوهش محاسبات مدل جنگل تصادفی در محیط نرم‌افزار R انجام شد. متغیر اندازه گره (که نشان‌دهنده تعداد برگ‌ها در هر شاخه است) با آزمون و خطا تعیین شد.

مدل GBM

GBM یا (Gradient Boosting Machine) توسط فریدمن در سال ۲۰۰۱ معرفی شد. این مدل همچنین به نام‌های MART (Multiple Additive Regression Trees) و GBRT (Gradient Boosted Regression Trees) نیز شناخته می‌شود. این مدل درختان را در یک زمان می‌سازد، به طوری که هر درخت جدید به تصحیح خطاهای مربوط به درختان آموزشی قبلی کمک می‌کند. این الگوریتم با آموزش یک درخت تصمیم‌گیری آغاز می‌شود که برای هر یک از مشاهدات یک وزن برابر با آن تعیین می‌شود. پس از ارزیابی درخت اول، وزن مشاهداتی را که طبقه‌بندی آنها سخت است را افزایش می‌دهیم و وزن کمتر را به مشاهداتی می‌دهیم که طبقه‌بندی آنها راحت‌تر است. بنابراین درخت دوم بر این داده‌های وزنی رشد می‌کند. هدف از این کار این است که پیش‌بینی درخت اول را بهبود بخشید. سپس خطای طبقه‌بندی درخت دوم را محاسبه کرد و یک درخت جدید را برای پیش‌بینی بازماندهای اصلاح شده آماده می‌شود. این روند را تکرار می‌شود برای تعداد مشخصی از تکرارها. درخت‌های بعدی در طبقه‌بندی مشاهداتی که به وسیله درختان قبلی به خوبی طبقه‌بندی نشده‌اند به ما کمک می‌کنند (۱۳).

ارزیابی مدل با منحنی ROC

منحنی تشخیص عملکرد نسبی (ROC) روش مفیدی برای نمایش کیفیت شناسایی قطعی و احتمالی و نیز پیش‌بینی سیستم‌ها است. مساحت زیرمنحنی (AUC) کیفیت پیش‌بینی سیستم را به وسیله توصیف توانایی سیستم برای پیش‌بینی دقیق رخ دادن یا ندادن وقایع از پیش تعیین شده توصیف می‌کند. منحنی ROC نشان‌دهنده حساسیت مدل نسبت به درصد سلول‌ها یا واحدهای ناپایدار که به درستی توسط مدل پیش‌بینی شده‌اند، در مقابل درصد سلول‌های ناپایدار پیش‌بینی شده نسبت به کل است. این مقادیر توانایی مدل را نسبت به تمایز صحیح بین مشاهدات مثبت و منفی در نمونه اعتبارسنجی بیان می‌کند (۱،۴۸). حساسیت زیاد بیان‌کننده تعداد زیاد پیش‌بینی درست (مثبت‌های حقیقی) است، در حالی که ویژگی بالا (اختلاف Specificity- 1 کم) نشان‌دهنده عدد کم مثبت‌های کاذب است. در مساحت زیر منحنی، نرخ مثبت کاذب (1-Specificity) در محور X و نرخ مثبت حقیقی (Sensitivity) در محور Y نمایش داده می‌شود (رابطه).

$$X = 1 - Specificity = 1 - \left[\frac{TN}{TN+FP}\right] \quad (۳)$$

$$Y = Sensitivity = \left[\frac{TN}{TP-FN}\right] \quad (۴)$$

روش بررسی همخطی بین فاکتورها (VIF)

هم‌خطی (Collinearity) پدیده‌ای است که به بیان وضعیتی می‌پردازد که یک متغیر توصیفی (Explanatory Variable) در رگرسیون چندگانه (Multiple Regression) با یک یا چند متغیر دیگر دارای رابطه خطی است به طوری که می‌توان آن را به صورت ترکیب خطی از دیگر متغیرها در نظر گرفت. زمانی که همخطی در فاکتورهای مورد استفاده در یک مدل وجود داشته باشد، ضرایب مدل حاصل معتبر نیستند، زیرا اثر هر یک از متغیرها توصیفی روی متغیر پاسخ (Response Variable) شامل اثر متغیرهای دیگر در مدل نیز هست. بنابراین واریانس برآوردگرهای ضرایب رگرسیونی افزایش یافته و در عمل پیش‌گویی توسط مدل با خطای زیادی همراه خواهد بود. به این ترتیب با تغییری اندک روی داده‌های به کار رفته در مدل، ضرایب رگرسیونی به شدت تغییر خواهند کرد (۱،۱۸).

تعیین اهمیت فاکتورها

الگوریتم‌های مختلفی برای تعیین اهمیت متغیرهای موثر بر یک پدیده وجود دارد. در این تحقیق از روش جینی (Jini) و در پکیج Biomod2 در نرم‌افزار R برای تعیین اهمیت فاکتورها استفاده شده است. در این الگوریتم، بهبود معیار تقسیم اندازه‌گیری اهمیتی است که به متغیر تقسیم نسبت داده می‌شود و روی تمام کلاس‌ها به طور جداگانه برای هر متغیر جمع می‌شود. این معیار کاملاً مشابه R2 در رگرسیون مجموعه آموزشی است و عبارتست از مجموع توان ۲ احتمال همه کلاس‌ها منهای یک. که به شکل رابطه ۵ محاسبه می‌شود:

$$Gini = 1 - \sum_j p_j^2 \quad (۵)$$

نتایج و بحث

نتایج حاصل از آنالیز هم‌خطی بین متغیرهای مستقل در جدول ۲ نشان داده شده است. بر اساس نتایج بدست آمده مشاهده می‌گردد که متغیرهای مستقل انتخاب شده دارای هم‌خطی پایینی (کوچکتر از ۵) می‌باشند. بنابرین ۱۲ متغیر

[Downloaded from drnl.sanru.ac.ir on 2025-12-19]

[DOR: 20.1001.1.27174425.1399.1.1.4.3]

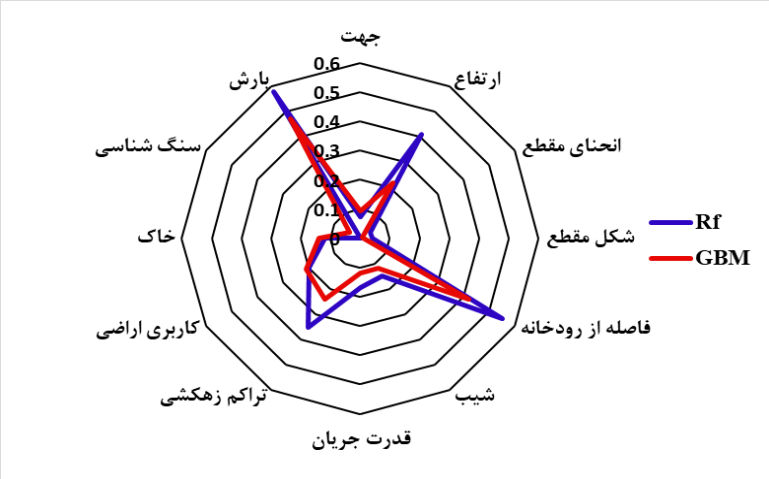
مستقل شناسایی شده جهت مدل‌سازی حساسیت سیل در منطقه مورد مطالعه استفاده شدند.

جدول ۲- نتایج آنالیز هم خطی چندگانه در متغیرهای مستقل

Table 1. Results of multi-collinearity analysis in independent variables

ردیف	فاکتورها	VIF
۱	جهت	۱/۰۱
۲	ارتفاع	۱/۸۰
۳	شکل انحنا	۱/۶۲
۴	نیمرخ انحنا	۱/۶۶
۵	فاصله از رودخانه	۱/۶۷
۶	شیب	۱/۲۶
۷	شاخص قدرت جریان	۱/۰۰
۸	تراکم زهکشی	۲/۷۵
۹	کاربری اراضی	۱/۸۰
۱۰	خاک	۱/۱۷
۱۱	سنگ شناسی	۱/۰۶
۱۲	بارندگی	۱/۰۱

مدل‌سازی حساسیت حوزه آبخیز به سیل با استفاده از مدل‌های GBM و RF با استفاده از نرم‌افزار R و پکیج biomod2 انجام شد. نتایج حاصل از تعیین مهم‌ترین متغیرهای تاثیرگذار در تعیین حساسیت منطقه به سیل در مدل‌های مورد استفاده در شکل ۳ نشان داده شده است. بر اساس نتایج بدست آمده مشاهده می‌گردد که متغیرهای بارندگی، فاصله از رودخانه، ارتفاع و تراکم زهکشی تاثیر بیش‌تری نسبت به سایر متغیرها در مدل‌سازی داشته‌اند در حالی که متغیرهای شکل مقطع و انحنای مقطع از اهمیت کمتری برخوردار بوده‌اند.



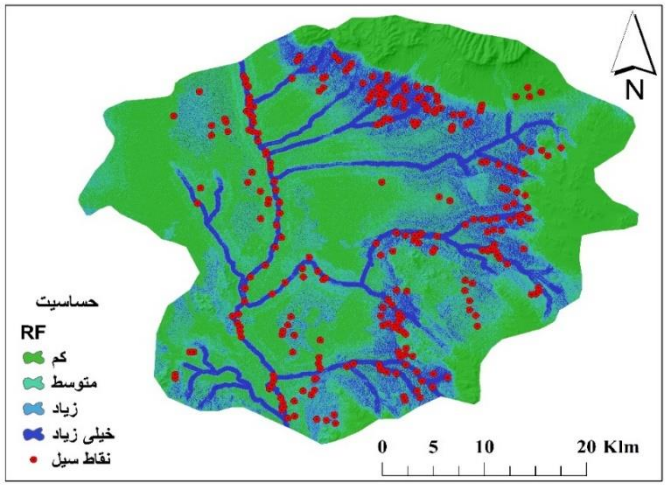
شکل ۳- اهمیت فاکتورهای مورد استفاده

Figure 3. Importance of the variables used

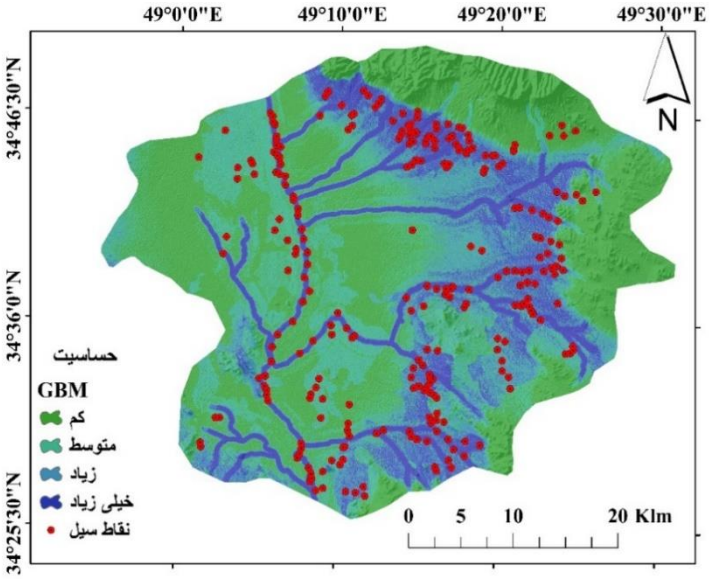
نقشه حساسیت به سیل در نرم‌افزار ArcGIS به چهار کلاس کم، متوسط، زیاد و خیلی زیاد با استفاده از روش Natural break به دلیل این‌که طبقه‌بندی براساس گروه‌بندی‌های طبیعیِ موجود در ذات داده‌ها انجام می‌شود و نقاط شکستگی بین دسته‌ها به گونه‌ای معین می‌شود که به بهترین وجه ممکن مقادیر مشابه در یک گروه جای داده شوند و تفاوت بین کلاس‌ها بیشینه شود. همچنین در این روش عوارض در کلاس‌هایی قرار می‌گیرند که مرز آنها بطوری تعیین شده است که در این مرزها تغییرات نسبی بزرگی در مقادیر داده‌ها صورت می‌گیرد (شکل های ۴ و ۵). براساس نتایج مدل GBM در صد مساحت‌های تحت پوشش کلاس‌های کم، متوسط، زیاد و خیلی زیاد به ترتیب ۴۰/۵۵، ۲۶/۰۸، ۱۵/۲۶ و ۱۸/۱۲ درصد می‌باشد. براساس نتایج مدل RF درصد مساحت‌های تحت پوشش کلاس‌های کم، متوسط، زیاد و خیلی زیاد به ترتیب ۴۷/۲۵، ۲۱/۱۷، ۱۴/۹۷ و ۱۶/۶۱ درصد می‌باشد.

[Downloaded from drnl.sanru.ac.ir on 2025-12-19]

[DOR: 20.1001.1.27174425.1399.1.1.4.3]



شکل ۴- نقشه حساسیت به وقوع سیل حوزه آبخیز کمیجان توسط مدل RF
Figure 4. Flood susceptibility map of Komijan watershed by RF model



شکل ۵- نقشه حساسیت به وقوع سیل حوزه آبخیز کمیجان توسط مدل GBM
Figure 5. Flood susceptibility map of Komijan watershed by GBM model

جدول ۳- مساحت کلاس‌های حساسیت سیل مدل‌های RF و GBM
Table 3. Flood susceptibility classes of RF and GBM models

	RF	GBM
کم	۴۷/۲۵	۴۰/۵۵
متوسط	۲۱/۱۷	۲۶/۰۸
زیاد	۱۴/۹۷	۱۵/۲۶
خیلی زیاد	۱۶/۶۱	۱۸/۱۲

به منظور تعیین دقت نقشه حساسیت به سیل در این پژوهش، مقدار Receiver Operating Characteristic (ROC)، True Skill Statistic (TSS) و KAPPA مورد استفاده قرار گرفت (۳۲). مساحت زیر منحنی ROC کیفیت پیش‌بینی سیستم را با استفاده از ارائه توانایی در مدل کردن وقوع و عدم وقوع وقایع از پیش تعیین شده، نشان می‌دهد (۹،۲۴). نتایج حاصل از ارزیابی مدل‌های مورد استفاده در جدول ۴ نشان داده شده است. براساس نتایج به دست آمده مدل‌های GBM و RF به ترتیب داری ضریب کارایی ۰/۷۵ و ۰/۸۳ بر اساس معیار ROC می‌باشند.

نتایج حاصل از ارزیابی مدل بر اساس معیار ROC نشان‌دهنده این است که مدل‌های مورد استفاده دارای دقت خوبی در تعیین حساسیت منطقه مورد مطالعه به سیل می‌باشند. هم‌چنین مقایسه مدل‌ها نشان داد که مدل RF با ضریب کارایی ROC ۰/۸۳ کارایی بهتری داشته است.

[Downloaded from drnl.sanru.ac.ir on 2025-12-19]
[DOR: 20.1001.1.27174425.1399.1.1.4.3]

جدول ۴- ارزیابی مدل RF و GBM با استفاده از شاخص‌های مختلف

Table 4. Evaluation RF and GBM models with different criteria

مدل	GBM			RF		
شاخص	Testing. data	Sensitivity	Specificity	Testing. data	Sensitivity	Specificity
ROC	۰/۷۵۱	۶۵/۴۵۵	۷۸/۱۸۲	۰/۸۳۴	۶۳/۴۵۵	۷۸/۱۸۲
TSS	۰/۶۴۷	۶۵/۴۵۵	۷۸/۱۸۲	۰/۶۷۲	۶۳/۴۵۵	۷۸/۱۸۲
KAPPA	۰/۶۳۶	۶۵/۴۵۵	۷۸/۱۸۲	۰/۷۲۴	۶۳/۴۵۵	۷۸/۱۸۲

براساس نتایج بدست آمده، مدل RF نسبت به مدل GBM کارایی بهتری در تعیین حساسیت سیل در منطقه مورد مطالعه دارد. عملکرد بهتر RF می‌تواند به دلیل توانایی آن در مدل‌سازی پایگاه داده‌های بزرگ و توانایی ادغام متغیرهای ورودی زیاد بدون تغییر متغیر باشد، همچنین مطالعات بسیار زیادی وجود دارند که به این نتیجه رسیده‌اند که مدل جنگل تصادفی توانایی بسیاری زیادی در تهیه نقشه مناطق حساس به سیل دارد که از جمله این مطالعات شامل ژائو و همکاران (۵۵)، چن و همکاران (۹)، تانگ و همکاران (۴۹) می‌باشد. جنگل‌های تصادفی از واریانس بالا در میان درختان فردی استفاده می‌کنند که هر یک از درخت‌ها را برای عضویت در کلاس می‌پذیرد (۳۹). سپس، RF بر اساس بیشترین تعداد آراء، طبقه مربوطه را تعیین می‌کند. علاوه بر این، RF می‌تواند با تعاملات و ارتباطات غیر خطی بین عوامل موثر را شناسایی و پیش‌بینی نماید (۱،۴). همچنین، در مطالعات مختلف در زمینه‌های آتش‌سوزی، محیط زیست، نقشه پتانسیل آب زیرزمینی و حساسیت به زمین لغزش مشخص شده است که مدل RF توانایی خوبی در مدل‌سازی و پیش‌بینی این عوامل دارد (۴،۲۷، ۲۸،۴۴).

نتایج پیش‌بینی نشان داد که پارامترهای بارندگی، فاصله از رودخانه، ارتفاع و تراکم زهکشی تاثیر بیش‌تری بر پتانسیل و استعداد سیل گرفتگی اراضی داشته و به‌کارگیری آن‌ها در مدل‌های مورد استفاده برای ارزیابی پتانسیل سیل مفید می‌باشد که با یافته‌های (۴۲) و (۲۰) هم‌خوانی دارد. براساس نتایج بدست آمده مشخص گردید که در پارامتر فاصله از آبراهه حساسیت به سیل در مناطق نزدیک آبراهه بیش‌تر است که با نتایج (۱۱) مطابقت دارد آنها نیز بیان داشتند حساس‌ترین مناطق هنگام وقوع سیل‌گیری، مناطق نزدیک آبراهه است. بین بارندگی و وقوع سیل‌گیری در منطقه مورد مطالعه رابطه مستقیمی وجود دارد به گونه‌ای که با افزایش بارندگی تعداد سیل‌گیری و وزن طبقات حساس به سیل نیز افزایش پیدا می‌کند که این نتایج با یافته‌های فام و همکاران (۳۵) و رضوی‌ترمه و همکاران (۴۳) مبنی بر وجود رابطه مستقیم بین بارندگی و سیل‌گیری در یک منطقه مطابقت دارد. در خصوص عامل تراکم زهکشی مناطقی که تراکم زهکشی بیش‌تری دارند به دلیل تجمع سریع رواناب نسبت به وقوع سیل‌گیری حساس‌تر هستند. یکی دیگر از عوامل تاثیرگذار شناخته شده در این تحقیق عامل ارتفاع می‌باشد، بررسی این عامل نشان‌دهنده این امر است که با افزایش ارتفاع، حساسیت طبقات نسبت به سیل‌گیری کاهش می‌یابد و حداکثر حساسیت سیل در طبقات مربوط به طبقات با دامنه های ارتفاعی کم است، که دلیل این امر را می‌توان در تجمع آب باران و وقوع سیل در این مکان‌ها دانست که اغلب مطالعات قبلی در این زمینه از قبیل مطالعات خسروی و همکاران (۱۸) و چاپی و همکاران (۷) تاییدکننده افزایش سیل‌گیری در مناطق پایین دست می‌باشند. تجمع سیلاب‌ها عمدتا در ارتفاعات پایین رخ می‌دهد و سیلاب‌های بزرگ در مناطقی با ارتفاع کم اتفاق| می‌افتند. با توجه به نتایج حاصل از حساسیت سیل در منطقه مطالعه مشخص گردید که حجم سیلاب و توزیع سیل در نزدیکی رودخانه‌ها و نزدیکی آنها که ارتفاع کم‌تر است قرار دارد و وقوع سیل در این مناطق افزایش می‌یابد. به طور کلی، طبق نظر (۱۵) و (۳۶) مناطقی که بیشترین حساس به سیل را دارند، مناطق دارای ارتفاع کم، حداقل شیب، مساحت مسطح و نزدیک به رودخانه‌ها است.

نتیجه‌گیری

سیل‌گیری یکی از پدیده‌های مهم مصیبت‌بار و فاجعه انگیز در نواحی مختلف محسوب می‌شود. از طریق آنالیز منطقه‌ای سیل‌گیری‌ها می‌توان مناطق حساس به سیل‌گیری را شناسایی و بنابرین از این طریق خسارات ناشی از آن را کاهش داد. هدف از تحقیق حاضر در واقع کاربرد برخی متغیرهای موثر بر وقوع سیل که این پارامترها برآیند عوامل مختلف محیطی و انسانی هستند، به منظور تعیین مناطق حساس به سیل با استفاده از مدل‌های GBM و RF بود. بدین منظور از ۱۲ متغیر محیطی و ۲۷۵ نقطه سیل استفاده گردید. نتایج حاصل از تعیین مناطق سیل‌گیر نشان داد که متغیرهای بارندگی، فاصله از رودخانه، ارتفاع و تراکم زهکشی اهمیت بیش‌تری در مدل‌سازی دارند. هم چنین نتایج حاصل از مدل‌سازی نشان داد که مدل RF کارایی بالاتری نسبت به مدل GBM دارد. بر اساس نقشه پیش‌بینی خطر سیل ارائه شده می‌توان اقدامات مدیریتی مناسبی جهت کاهش خسارت‌ها و تلفات ناشی از سیل انجام داد. به کارگیری تکنیک داده‌کاوی و سیستم اطلاعات جغرافیایی به منظور بررسی پتانسیل و استعداد سیل، مخصوصاً در کشورهای در حال توسعه که دسترسی به اطلاعات و داده‌های هیدروژئولوژیکی و ادافیکی با مشکل و محدودیت مواجه هستند، می‌تواند مفید باشد.

[Downloaded from drnl.sanru.ac.ir on 2025-12-19]

[DOR: 20.1001.1.27174425.1399.1.1.4.3]

منابع

1. Arab Ameri, A., H.R. Pourghasemi and K. Shirani. 2017. Zoning of flood sensitivity using a new ensemble method of Bayesian theory Hierarchical analysis process (Case study: Neka watershed - Mazandaran province). Eco–Hydrology, 4(2): 447-462.
2. Avand, M., S. Janizadeh. D.T. Bui, V.H. Pham. P.T.T. Ngo and V. Nhu. 2020. A tree-based intelligence ensemble approach for spatial prediction of potential groundwater. Int. Journal Digital Earth 0, 1–22. https://doi.org/10.1080/17538947.2020.1718785.
3. Bubeck, P., W.J.W. Botzen and J.C.J.H. Aerts. 2012. A Review of Risk Perceptions and Other Factors that Influence Flood Mitigation Behavior. Risk Analysis. 32, 1481-1495. https://doi.org/10.1111/j.1539-6924.2011.01783.x.
4. Bui, D.T., M. Panahi, H. Shahabi, V.P. Singh, A. Shirzadi, K. Chapi and B.B. Ahmad. 2018. Novel hybrid evolutionary algorithms for spatial prediction of floods. Scientific Reports, 8(1): 1-14.
5. Bui, D.T., B. Pradhan, H. Nampak, Q.T. Bui, Q.A. Tran and Q.P. Nguyen. 2016. Hybrid artificial intelligence approach based on neural fuzzy inference model and metaheuristic optimization for flood susceptibilitgy modeling in a high-frequency tropical cyclone area using GIS. Journal of Hydrology, 540: 317-330.
6. Catani, F., D. Lagomarsino, S. Segoni and V. Tofani. 2013. Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. Nat. Hazards Earth System Sciences, 13: 2815-2831.
7. Chapi, K., V.P. Singh, A. Shirzadi, H. Shahabi, D.T. Bui, B.T. Pham and K. Khosravi. 2017. A novel hybrid artificial intelligence approach for flood susceptibility assessment. Environmental modelling & software, 95: 229-245.
8. Chau, K.W., C.L. Wu and Y.S. Li. 2005. Comparison of several flood forecasting models in Yangtze River. Journal of Hydrologic Engineering, 10: 485-491.
9. Chen, W., Y. Li, W. Xue, H. Shahabi, S. Li, H. Hong and B.B. Ahmad. 2020. Modeling flood susceptibility using data-driven approaches of naïve bayes tree, alternating decision tree and random forest methods. Science of The Total Environment, 701: 134979.
10. Chen, W., M. Panahi and H.R. Pourghasemi, 2017a. Performance evaluation of GIS-based new ensemble data mining techniques of adaptive neuro-fuzzy inference system (ANFIS) with genetic algorithm (GA), differential evolution (DE), and particle swarm optimization (PSO) for landslide spatial modelling. Catena 157, 310-324. https://doi.org/10.1016/j.catena.2017.05.034.
11. Chen, W., H.R. Pourghasemi, A. Kornejady and N. Zhang. 2017b. Landslide spatial modeling: introducing new ensembles of ANN, MaxEnt, and SVM machine learning techniques. Geoderma, 305: 314-327.
12. Costache, R., Q.B. Pham, M. Avand, N.T.T. Linh, M. Vojtek, J. Vojteková, S. Lee, D.N. Khoi, P.T.T. Nhi and T.D. Dung. 2020. Novel hybrid models between bivariate statistics, artificial neural networks and boosting algorithms for flood susceptibility assessment. Journal of Environmental Management, 265: 110485.
13. Du, J., J. Fang, W. Xu and P. Shi. 2013. Analysis of dry/wet conditions using the standardized precipitation index and its potential usefulness for drought/flood monitoring in Hunan Province, China. Stochastic Environmental Research and Risk Assessment, 27: 377-387. https://doi.org/10.1007/s00477-012-0589-6.
14. Fernández, D.S. and M.A. Lutz. 2010. Urban flood hazard zoning in Tucumán Province, Argentina, using GIS and multicriteria decision analysis. Engineering Geology, 111: 90-98.
15. Francke, T. 2009. Measurement and modelling of water and sediment fluxes in meso-scale dryland catchments. Unpubl. PhD Thesis, Univ. Potsdam, Potsdam Google Sch.
16. Friedman, J.H. 1991. Estimating functions of mixed ordinal and categorical variables using adaptive splines.
17. Janizadeh, S., M. Avand, A. Jaafari, T.Van. Phong, M. Bayat, E. Ahmadisharaf, I. Prakash, B.T. Pham and S. Lee. 2019. Prediction Success of Machine Learning Methods for Flash Flood Susceptibility Mapping in the Tafresh Watershed, Iran. Sustainability, 11: 5426.
18. Khosravi, K., E. Nohani, E. Maroufinia and H.R. Pourghasemi. 2016. A GIS-based flood susceptibility assessment and its mapping in Iran: a comparison between frequency ratio and weights-of-evidence bivariate statistical models with multi-criteria decision-making technique. Natural Hazards, 83(2): 947-987.
19. Khosravi, K., B.T. Pham, K. Chapi, A. Shirzadi, H. Shahabi, I. Revhaug and D.T. Bui. 2018. A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran. Science of the Total Environment, 627: 744-755.
20. Khosravi, K., H.R. Pourghasemi, K. Chapi and M. Bahri. 2016. Flash flood susceptibility analysis and its mapping using different bivariate models in Iran: a comparison between Shannon's entropy, statistical index and weighting factor models. Environmental Monitoring and Assessment, 188(12): 656. https://doi.org/10.1007/s10661-016-5665-9.

[Downloaded from drnl.sanru.ac.ir on 2025-12-19]

[DOR: 20.1001.1.27174425.1399.1.1.4.3]

21. Khosravi, K., H. Shahabi, B.T. Pham, J. Adamowski, A. Shirzadi, B. Pradhan and H. Hong. 2019. A comparative assessment of flood susceptibility modeling using Multi-Criteria Decision-Making Analysis and Machine Learning Methods. Journal of Hydrology, 573: 311-323.
22. Kia, M.B., S. Pirasteh, B. Pradhan, A.R. Mahmud, W.N.A. Sulaiman, A. Moradi, W. Nor, A. Sulaiman and A. Moradi. 2012. An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia. Environmental Earth Sciences, 67: 251-264. https://doi.org/10.1007/s12665-011-1504-z.
23. Kornejady, A., K. Heidary, M. Sarparast, G. Khosravi and M. Mombeini. 2014. Performance Assessment of Two "LNRF" and "AHP-Area Density" Models in landslide Susceptibility Zonation. Journal of Life Science and Biomedicine, 4: 169-176.
24. Kourgialas, N.N. and G.P. Karatzas. 2011. Gestion des inondations et méthode de modélisation sous SIG pour évaluer les zones d'aléa inondation-une étude de cas. Hydrological Sciences Journal, 56: 212-225. https://doi.org/10.1080/02626667.2011.555836.
25. Lee, S., I. Park and J.K. Choi. 2012. Spatial prediction of ground subsidence susceptibility using an artificial neural network. Environmental Management, 49: 347-358. https://doi.org/10.1007/s00267-011-9766-5.
26. Leskens, J.G., M. Brugnach, A.Y. Hoekstra and W. Schuurmans. 2014. Environmental Modelling & Software Why are decisions in fl ood disaster management so poorly supported by information from fl ood models ? Environemental Modeling Software, 53: 53-61. https://doi.org/10.1016/j.envsoft.2013.11.003.
27. Levy, J.K., J. Hartmann, K.W. Li, Y. An and A. Asgary. 2007. Multi-Criteria Decision Support Systems for Flood Hazard Mitigation and Emergency Response in Urban Watersheds 1. JAWRA Journal of the American Water Resources Association, 43: 346-358.
28. Mukerji, A., C. Chatterjee and N.S. Raghuwanshi. 2009. Flood forecasting using ANN, neuro-fuzzy, and neuro-GA models. Journal of Hydrologic Engineering, 14(6): 647-652.
29. Naghibi, S.A., K. Ahmadi and A. Daneshi. 2017a. Application of Support Vector Machine, Random Forest, and Genetic Algorithm Optimized Random Forest Models in Groundwater Potential Mapping. Water Resources Management, 31: 2761-2775. https://doi.org/10.1007/s11269-017-1660-3.
30. Naghibi, S.A., D.D. Moghaddam, B. Kalantar, B. Pradhan and O. Kisi. 2017b. A comparative assessment of GIS-based data mining models and a novel ensemble model in groundwater well potential mapping. Journal of Hydrology, 548: 471-483. https://doi.org/10.1016/j.jhydrol.2017.03.020.
31. Nguyen, P.T., D.H. Ha, M. Avand, A. Jaafari, H.D. Nguyen, N. Al-Ansari and L.S. Ho. 2020. Soft Computing Ensemble Models Based on Logistic Regression for Groundwater Potential Mapping. Applied Sciences, 10(7): 2469. https://doi.org/10.3390/app10072469.
32. Nikoo, M.M., F. Ramezani, M. Hadzima-Nyarko, E.K. Nyarko and M.M. Nikoo. 2016. Flood-routing modeling with neural network optimized by social-based algorithm. Natural hazards, 82: 1-24. https://doi.org/10.1007/s11069-016-2176-5.
33. Oliveira, S., F. Oehler, J. San-Miguel-Ayanz, A. Camia and J.M.C. Pereira. 2012. Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest. Forest Ecology and Management, 275: 117-129. https://doi.org/10.1016/j.foreco.2012.03.003.
34. Peters, J., B. De Baets, N.E.C. Verhoest, R. Samson, S. Degroeve, P.De. Becker and W. Huybrechts. 2007. Random forests as a tool for ecohydrological distribution modelling. Ecological Modelling, 207: 304-318. https://doi.org/10.1016/j.ecolmodel.2007.05.011.
35. Pham, B.T., M. Avand, S. Janizadeh, T.V. Phong, N. Al-Ansari, L.S. Ho and F. Jafari. 2020. GIS based hybrid computational approaches for flash flood susceptibility assessment. Water, 12(3): 683. https://doi.org/10.3390/w12030683.
36. Pham, B.T., A. Jaafari, M. Avand, N.Al-Ansari, T. Dinh Du, H.P.H. Yen, T. Van Phong, D.H. Nguyen, H. Van Le, D. Mafi-Gholami. 2020. Performance Evaluation of Machine Learning Methods for Forest Fire Modeling and Prediction. Symmetry (Basel). 12: 1022.
37. Pierdicca, N., L. Pulvirenti, M. Chini, L. Guerriero and P. Ferrazzoli. 2010. A FUZZY-LOGIC-BASED APPROACH FOR FLOOD DETECTION FROM COSMO- SKYMED DATA Dept. Information , Electronic and Telecommunications Engineering , Sapienza University of Rome, Istituto Nazionale di Geofisica e Vulcanologia , Rome , Italy Dept . Computer Sciences,4796-4798.
38. Pourghasemi, H., S. Youse, A. Kornejady, A. Cerdà. 2017. Performance assessment of individual and ensemble data-mining techniques for gully erosion modeling, Science of the Total Environment, 609: 764-775. https://doi.org/10.1016/j.scitotenv.2017.07.198.
39. Pourghasemi, H.R. and M. Mohammadi. 2016. Presenting a new ensemble method of Bayesian models and logistic regression in landslide susceptibility assessment in Khalkhal city. Journal of Environmental Erosion Research, 6(2): 16-30.
40. Pourghasemi, H.R. and M. Beheshtirad. 2015. Assessment of a data-driven evidential belief function model and GIS for groundwater potential mapping in the Koohrang Watershed, Iran. Geocarto International, 30: 662-685.

[Downloaded from drnl.sanru.ac.ir on 2025-12-19]

[DOR: 20.1001.1.27174425.1399.1.1.4.3]

41. Pourghasemi, H.R. and N. Kerle. 2016. Random forests and evidential belief function-based landslide susceptibility assessment in Western Mazandaran Province, Iran. Environemental Earth Sciences, 75: 185.
42. Rahmati, O., H.R. Pourghasemi and H. Zeinivand. 2016. Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the Golastan Province, Iran. Geocarto International, 31: 42-70. https://doi.org/10.1080/10106049.2015.1041559.
43. Razavi Termeh, S.V., A. Kornejady, H.R. Pourghasemi and S. Keesstra. 2018. Flood susceptibility mapping using novel ensembles of adaptive neuro fuzzy inference system and metaheuristic algorithms. Science of the Total Environemental, 615: 438-451. https://doi.org/10.1016/j.scitotenv.2017.09.262.
44. Sahoo, G.B., S.G. Schladow and J.E. Reuter. 2009. Forecasting stream water temperature using regression analysis, artificial neural network, and chaotic non-linear dynamic models. Journal of Hydrology, 378: 325-342. https://doi.org/10.1016/j.jhydrol.2009.09.037.
45. Shahabi, H., B. Jarihani, S. Tavakkoli piralilou, D. Chittleborough, M. Avand and O. Ghorbanzadeh, 2019. A Semi-Automated Object-Based Gully Networks Detection Using Di ff erent Machine Learning Models: sensor, 19: 1-21.
46. Stumpf, A. and N. Kerle. 2011. Object-oriented mapping of landslides using Random Forests. Remote Sensing of Environment, 115: 2564-2577. https://doi.org/10.1016/j.rse.2011.05.013.
47. Svetnik, V., A. Liaw, C. Tong, J.C. Culberson, R.P. Sheridan and B.P. Feuston. 2003. Random forest: a classification and regression tool for compound classification and QSAR modeling. Journal of Chemical Information and Computer Sciences, 43: 1947-1958.
48. Talukdar, S., B. Ghose, R. Salam, S. Mahato, Q.B. Pham, N.T.T. Linh and M. Avand. 2020. Flood susceptibility modeling in Teesta River basin, Bangladesh using novel ensembles of bagging algorithms. Stochastic Environmental Research and Risk Assessment, 1-24.
49. Tang, X., J. Li, M. Liu, W. Liu and H. Hong. 2020. Flood susceptibility assessment based on a novel random Naïve Bayes method: A comparison between different factor discretization methods. Catena, 190: 104536.
50. Tehrany, M.S., B. Pradhan, M.N. Jebur. 2014. Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. Journal of Hydrology, 512: 332-343. https://doi.org/10.1016/j.jhydrol.2014.03.008.
51. Tehrany, M.S., B. Pradhan and M.N. Jebur. 2015. Flood susceptibility analysis and its verification using a novel ensemble support vector machine and frequency ratio method. Stochastic Environmental Research and Risk Assessment, 29: 1149-1165. https://doi.org/10.1007/s00477-015-1021-9.
52. Tierney, K.J., M.K. Lindell and R.W. Perry. 2002. Facing the unexpected: disaster preparedness and response in the United States. Disaster Prevention and Management: An International Journal, 11: 222.
53. Vorpahl, P., H. Elsenbeer, M. Märker and B. Schröder. 2012. How can statistical models help to determine driving factors of landslides? Ecological Modelling, 239: 27-39. https://doi.org/10.1016/j.ecolmodel.2011.12.007.
54. Youssef, A.M., B Pradhan and A.M. Hassan. 2011. Flash flood risk estimation along the St. Katherine road, southern Sinai, Egypt using GIS based morphometry and satellite imagery. Environmental Earth Sciences, 62: 611-623. https://doi.org/10.1007/s12665-010-0551-1.
55. Zhao, G., B. Pang, Z. Xu, J. Yue and T. Tu. 2018. Mapping flood susceptibility in mountainous areas on a national scale in China. Science of the Total Environment, 615: 1133-1142.

[Downloaded from drnl.sanru.ac.ir on 2025-12-19]

[DOR: 20.1001.1.27174425.1399.1.1.4.3]

Evaluating the Efficiency of Machine Learning Models in Preparing Flood Probability Mapping

Mohammadtaghi Avand[1], Saeid Janizadeh[2] and Faeze Jafari[3]

1- PhD Student in Watershed Management, Faculty of Natural Resources and Marine Sciences, Tarbiat Modares University, Noor, Iran, (Corresponding author: Mt.avand70@gmail.com)
2- PhD Student in Watershed Management, Faculty of Natural Resources and Marine Sciences, Tarbiat Modares University, Noor, Iran
3- PhD Student in Watershed Management, Faculty of Natural Resources and Marine Sciences, Tarbiat Modares University, Noor, Iran
Received: September 5, 2020 Accepted: October 14, 2020

Abstract

Flood is one of the most devastating natural disasters that annually causes financial and life losses. Therefore, developing a susceptibility map for flood management and reducing its harmful effects is essential. The present study was conducted to prepare a flood susceptibility map using data mining models including Random Forest (RF) and Gradient Boosting Machine (GBM). At first, 275 flooding locations flood and 275 non-flood locations were identified in the Komijan watershed of Markazi province. Spatial locations were randomly divided to 70% (190 location) and 30% (82 location) for modeling and validation, respectively. Then, 12 factors affecting the occurrence of flood including slope, aspect, altitude, rainfall, land use, distance from river, drainage density, plan curvature, profile curvature, lithology, soil and stream power index were determined. The ROC curve was used to evaluate the models used. The results showed that in the validation stage, the under curve for RF and GBM models was 0.83 and 0.75%, respectively, which indicates that the RF model is more accurate in producing a flood susceptibility map. The most important factors affecting the flood are rainfall, distance from river and altitude.

Keywords: Data mining models, Flood zoning, GBM, Komijan watershed, Random forest

[Downloaded from drnl.sanru.ac.ir on 2025-12-19]

[DOR: 20.1001.1.27174425.1399.1.1.4.3]