

پهنه‌بندی احتمال تولید سیل‌های ناگهانی با استفاده از روش‌های یادگیری ماشین (مطالعه موردی: حوزه آبخیز سد کرج)

عبدالواحد کردی^۱، امیر سعدالدین^{۲*}، جهانگیر محمدی^۳، واحد بردی شیخ^۴، امید اسدی نلیوان^۵

تاریخ دریافت: ۱۴۰۴/۰۱/۲۸

تاریخ پذیرش: ۱۴۰۴/۰۴/۲۲

چکیده

حوزه آبخیز سد کرج در استان البرز به دلیل شرایط خاص پستی و بلندی، زمین‌شناختی، اقلیمی و عوامل انسان‌ساخت، از پتانسیل سیل‌خیزی زیادی دارد و صدمات ناشی از سیل‌های ناگهانی، همواره مشکلات عدیده‌ای را برای مسافران، گردشگران و ساکنان محلی ایجاد نموده است. بررسی فرایند تشکیل و نیز ویژگی‌های رخداد‌های سیل در حوزه آبخیز مذکور، بیانگر وقوع سیل‌های ناگهانی به عنوان تیپ غالب سیل در منطقه است. به منظور ارزیابی احتمال مکانی تولید سیل، ابتدا نقشه شواهد سیل‌های پیشین رخ داده طی یک دوره ۲۰ ساله، با استفاده از بانک اطلاعات سیل دستگاه‌های مرتبط، منابع محلی و بازدیدهای میدانی تهیه شد. پس از تفکیک انواع سیل، ۱۴ عامل مؤثر بر سیل‌خیزی براساس مرور منابع متعدد، انتخاب و هم‌خطی عوامل با آزمون فاکتور تورم واریانس بررسی شد. در ادامه، رخداد‌های سیل، به داده‌های آموزشی (۷۰ درصد) و اعتبارسنجی (۳۰ درصد) دسته‌بندی شدند و سه تکنیک داده‌کاوی پیشینه آنتروپی، جنگل تصادفی و ماشین بردار پشتیبان برای مدل‌سازی مکانی تولید سیل به کار رفت. عملکرد مدل‌های مذکور، با استفاده از معیار منحنی تشخیص عملکرد نسبی (ROC) و شاخص کاپا ارزیابی شد. مدل ماشین بردار پشتیبان با مقدار مساحت زیرمنحنی ۰/۸۸۵ و ضریب کاپا ۰/۶۸، به عنوان مدل برتر پیش‌بینی مکانی وقوع تولید سیل انتخاب و نقشه احتمال وقوع تولید سیل بر این اساس تهیه و در پنج کلاس طبقه‌بندی شد. طبق پهنه‌بندی انجام شده، ۵۳ درصد محدوده مورد مطالعه (حدود ۴۶۲۰۰ هکتار) دارای پتانسیل سیل‌خیزی متوسط، زیاد و خیلی زیاد قرار دارد. همچنین براساس نتایج حاصل، عوامل ارتفاع از سطح دریا، تراکم زهکشی، شیب و میانگین بارندگی سالیانه به ترتیب بیشترین نقش را در وقوع تولید سیلاب‌های ناگهانی دارند. یافته‌های تحقیق حاضر، برای شناسایی میزان حساسیت مناطق سیل‌خیز دارای اولویت به منظور تدوین برنامه‌های مدیریتی از قبیل اجتناب از مناطق بسیار حساس، افزایش ظرفیت‌های سازگاری و مدارا جوامع محلی کاربرد دارد و می‌تواند مبنای تصمیم‌گیری و برنامه‌ریزی صحیح اقدامات لازم از سوی دستگاه‌ها و سازمان‌های متولی امر مدیریت سیل، و نیز هزینه‌کرد صحیح اعتبارات با توجه به محدودیت‌های موجود قرار گیرد.

کلیدواژه‌ها: احتمال وقوع تولید سیل، پهنه‌بندی سیل، داده‌کاوی، سیل‌خیزی.

۱. دانشجوی دکتری آبخیزداری، دانشگاه علوم کشاورزی و منابع طبیعی گرگان، ایران، vahed.kordi@yahoo.com

۲. نویسنده مسئول، استاد گروه آبخیزداری، دانشگاه علوم کشاورزی و منابع طبیعی گرگان، ایران، نویسنده مسئول، amir.sadoddin@gau.ac.ir

۳. دانشیار گروه جنگل، دانشگاه علوم کشاورزی و منابع طبیعی گرگان، ایران، mohamadi.jahangir@gau.ac.ir

۴. استاد گروه آبخیزداری، دانشگاه علوم کشاورزی و منابع طبیعی گرگان، ایران، v.sheikh@yahoo.com

۵. استادیار گروه مهندسی منابع آب، دانشکده کشاورزی، دانشگاه مراغه، ایران، omid.asadi@ut.ac.ir

مقدمه

سیل یکی از رویدادهای حادی هیدرواقليمی و از جدی‌ترین بلایای طبیعی است که جوامع بشری را مورد تهدید قرار می‌دهد. تغییرات اقلیمی، رشد سریع جمعیت و توسعه بی‌رویه در مناطق شهری از جمله عواملی هستند که منجر به افزایش فراوانی و شدت سیل‌ها شده‌اند (کلدسن،^۱ ۲۰۱۰). بررسی‌ها نشان می‌دهد که از سال ۱۹۸۸ تا ۱۹۹۷، سیلاب عامل بیش از ۵۸ درصد از تلفات انسانی ناشی از بلایای طبیعی در جهان بوده است (ساین،^۲ ۲۰۱۰). این خسارات علاوه بر تلفات جانی، خسارت‌های اقتصادی عظیمی را نیز شامل می‌شود، به‌طوری که در این بازه زمانی تنها در دو دهه اخیر، خسارات مالی ناشی از سیلاب بیش از ۷۰۰ میلیارد دلار برآورد شده است (میلر^۳ و همکاران، ۱۹۹۰). بااینکه قسمت اعظم کشور ایران در اقلیم خشک و نیمه‌خشک قرار دارد، در طول سالیان متمادی، سیل صدمه بیشتری از خشکسالی بر کشور تحمیل نموده است. نمونه روشن این موضوع فراوانی وقوع سیل در چند دهه اخیر است که اکثر مناطق کشور حتی ناحیه خزری را که پوشیده از جنگل‌های انبوه و خاک‌های عمیق است در بر گرفته و به‌نوبه خود منجر به تلفات و خسارات مختلف ازجمله تخریب زیرساخت‌ها و سکونتگاه‌ها، فرسایش خاک و تخریب سرزمین، نابودی و یا کاهش کیفیت زیستگاه‌ها و تبعات غیرمستقیم و ناملموس بسیاری شده است (اکبری و همکاران، ۲۰۱۶؛ هیئت ویژه گزارش ملی سیلاب‌ها، ۲۰۲۰). سیل بیش از آنکه یک پدیده طبیعی تلقی شود، ناشی از عوارض به هم خوردن تعادل طبیعی و نیز افزایش در معرض بودن و افزایش آسیب‌پذیری مناطق است؛ به‌طوری‌که امروزه بارش‌های معمولی نیز موجب جاری شدن سیلاب می‌گردد (اسدی نلیوان و همکاران، ۲۰۱۸). با توجه به اینکه فرایندهای ایجاد سیل متعدد بوده و پدیده سیل ماهیت چندبعدی و پویا دارد، قطعاً تمامی عوامل طبیعی، انسانی و سازمانی - مدیریتی در کنار یکدیگر بر بزرگی، شدت، گستره و تداوم سیل تأثیر خواهند داشت (بای و همکاران، ۲۰۲۲). شناخت ابعاد و پهنه‌های سیل‌گیر و بررسی ویژگی‌های مکانی

آن ازجمله فعالیت‌هایی است که می‌تواند در بحث مدیریت بهینه مناطق در معرض خطر سیل، مؤثر و مفید باشد. با تحلیل‌های مکانی و بررسی پهنه‌های طبیعی وقوع سیل، می‌توان مدیریت خطر سیل را به‌عنوان یک چارچوب منطقی ارزیابی و سنگ بنای اقدامات پیشگیرانه معرفی نمود.

بنابراین برای پیشگیری و مهار سیل باید با بهره‌گیری از روش‌های علمی صحیح، عوامل ایجاد و تشدید سیل شناسایی و مناطقی که پتانسیل بالایی در تولید سیل دارند تعیین شوند؛ زیرا به‌دلیل وسعت زیاد و گستردگی حوزه‌های آبخیز، انجام عملیات اجرایی و اصلاحی در سراسر حوزه‌های آبخیز امکان‌پذیر نبوده و حتی در صورت عدم بررسی دقیق می‌تواند امکان تشدید دبی اوج را با تشدید هم‌زمانی دبی‌های اوج زیرحوزه‌ها سبب گردد (اسدی نلیوان و همکاران، ۲۰۱۸). ازاین‌رو باید به طریقی مناطقی که پتانسیل بالایی در تولید سیل دارند، شناسایی شوند تا امکان انجام عملیات اصلاحی در سطوح کوچک‌تر و خطرساز فراهم شود. درزمینه احتمال وقوع مکانی سیل مطالعات متعددی در سطح جهان انجام گرفته است. تهرانی و همکاران (۲۰۱۴a) تهیه نقشه حساسیت^۴ سیل را با استفاده از مدل‌های ماشین بردار پشتیبان و وزن واقعه در کشور مالزی بررسی نمودند. این عوامل با استفاده از مدل وزن واقعه وزندهی شده و به‌عنوان ورودی در مدل ماشین بردار پشتیبان استفاده شدند تا همبستگی بین وقوع سیل و هریک از عوامل ارزیابی گردد. نتایج حاکی از آن بود که بهترین نتیجه از تابع کرنل شعاعی حاصل می‌گردد؛ به‌طوری‌که دقت مدل برابر با ۹۶/۴۸ درصد بوده که حاکی از کارایی بالای مدل است. تهرانی و همکاران (۲۰۱۴b) به تهیه نقشه حساسیت وقوع سیل توسط ترکیب مدل‌های آماری دومتغیره و چندمتغیره در کشور کره جنوبی پرداختند. نتایج آن‌ها حاکی از نرخ یادگیری و پیش‌بینی برابر با ۹۲/۷ و ۸۲/۳ درصد است. آن‌ها همچنین بیان کردند که این روش کارایی لازم برای تهیه نقشه‌های حساسیت به وقوع سیل را دارا می‌باشد. ونگ^۵ و همکاران (۲۰۱۵)، حساسیت سیل را با استفاده از مدل جنگل

1. Kjeldsen
2. Singh
3. Miller

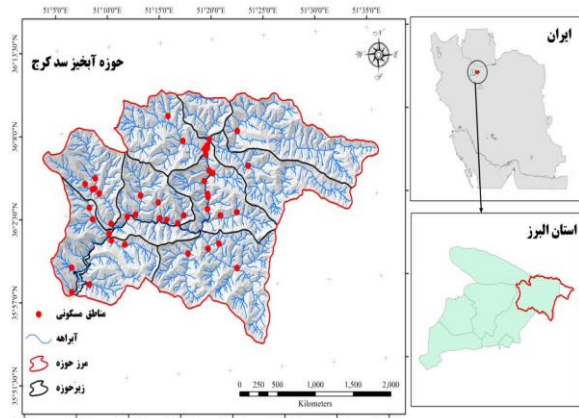
4. Susceptibility
5. Wang

نقشه‌ها از ۱۱ عامل مؤثر بر سیل شامل ارتفاع، شیب، جهت، شاخص رطوبت توپوگرافی، شاخص توان جریان، شاخص پوشش گیاهی، زمین‌شناسی، بارندگی، کاربری اراضی، فاصله تا جاده‌ها و فاصله تا شبکه آبراه‌ها استفاده کردند. نتایج نشان داد روش‌های یادگیری ماشین عملکرد بهتری نسبت به روش‌های تصمیم‌گیری چندمعیاره دارند. رحمتی و همکاران (۲۰۱۶) با استفاده از مدل تحلیل تصمیم‌گیری چندمعیاره (MCDM)، نقشه‌های حساسیت سیلاب در شهر یاسوج را تهیه کردند و نقش متغیرهایی نظیر کاربری اراضی و شیب زمین را در وقوع سیلاب برجسته ساختند. حق‌زاده و همکاران (۲۰۱۷) با استفاده از مدل آنتروپی شانون، نشان دادند که عوامل نظیر ارتفاع و تراکم زهکشی تأثیر بسزایی در توزیع مکانی سیلاب دارند. در دهه اخیر، الگوریتم‌های یادگیری ماشین به دلیل توانایی بالا در تحلیل داده‌های بزرگ و پیچیده، مورد توجه گسترده قرار گرفته‌اند. این روش‌ها شامل الگوریتم‌هایی نظیر جنگل تصادفی، ماشین بردار پشتیبان و شبکه‌های عصبی مصنوعی، توانسته‌اند دقت بالاتری در پیش‌بینی وقوع سیلاب ارائه دهند. برای مثال، جوبین^۶ و همکاران (۲۰۱۹) با استفاده از مدل جنگل تصادفی، نقشه سیل‌خیزی در حوزه آبخیز خیاوچای را تهیه کرده و دقت مدل را موفقیت‌آمیز گزارش کردند. جمع‌بندی بررسی سوابق تحقیق گویای آن است که در سطح دنیا، مطالعات گسترده‌ای برای ارزیابی و تهیه نقشه حساسیت و پتانسیل سیل‌خیزی در مناطق مختلف با شرایط محیطی متنوع و با به‌کارگیری روش‌های متعدد صورت گرفته است. پژوهش حاضر نیز در همین راستا و با هدف تعیین میزان اثرگذاری عوامل مؤثر بر تولید سیل‌های ناگهانی، شناسایی و مکان‌یابی پهنه‌های سیل‌خیز و به نوعی کلاس‌بندی مناطق حساس به‌منظور مدیریت مؤثر سیل‌های محتمل آتی انجام شده است. همچنین نتایج حاصل به‌عنوان ابزاری کاربردی برای بهره‌مندی آبخیزنشینان، برنامه‌ریزان و تصمیم‌گیران امر به‌ویژه مسئولان ارگان‌ها و سازمان‌های راه و شهرسازی، جهاد کشاورزی، بنیاد مسکن با هدف انجام اقدامات پیشگیرانه، فنی، نظارتی و کنترلی لازم محسوب می‌شود.

تصادفی در حوضه رودخانه دنجیانگ^۱ کشور چین ارزیابی نمودند. نتایج تحقیق آن‌ها نشان داد که: ۱. میزان خطای این مدل با افزایش اندازه نمونه و تعداد درختان طبقه‌بندی‌شده کاهش می‌یابد؛ ۲. مناطق با خطر زیاد و خیلی زیاد تقریباً ۱۹/۹ درصد کل منطقه را اشغال می‌کند؛ ۳. شاخص حداکثر بارش سه‌روزه (M3PD)، عمق رواناب (RD)، فرکانس طوفان (TF)^۲، مدل رقومی ارتفاعی (DEM) و شاخص رطوبت توپوگرافی (TWI) پنج شاخص مهم از ۱۱ شاخص انتخاب‌شده در فرایند مدل‌سازی حساسیت سیل هستند. تهرانی و همکاران (۲۰۱۵) پهنه‌بندی مناطق حساس به وقوع سیل را در مالزی توسط روش داده‌کاوی ماشین بردار پشتیبان با چهار تابع خطی، چندجمله‌ای، شعاعی و سیگموئید مورد بررسی قرار دادند. نتایج بیانگر آن است که شاخص سطح زیرمنحنی (AUC) به‌ترتیب برای توابع فوق برابر با ۸۴/۶۳، ۸۳/۹۲، ۸۴/۹۷ و ۸۱/۸۸ درصد است که نشان از کارایی بالاتر تابع شعاعی می‌باشد. طبق نتایج ارائه‌شده، شیب و طبقات ارتفاعی در کلیه توابع استفاده‌شده از مؤثرترین فاکتورها بوده‌اند. سادلر^۳ و همکاران (۲۰۱۸) فراوانی وقوع سیلاب ساحلی را با استفاده از مدل‌های پواسون و جنگل تصادفی مورد بررسی قرار دادند. نتایج حاصل از تحقیق آن‌ها نشان داد که مدل جنگل تصادفی بهتر از مدل پواسون در پیش‌بینی تعداد گزارش‌های سیل مؤثر بوده و نرخ منفی کاذب کمتری داشته است. شائو^۴ و همکاران (۲۰۱۹) ارزیابی حساسیت سیلاب شهری را با استفاده از مدل‌های یادگیری ماشین بررسی نمودند و نتایج حاصل از تحقیق آن‌ها حاکی از آن است که مدل ماشین بردار پشتیبان می‌تواند از اطلاعات مکانی بهتر استفاده کند و از تمام مدل‌های مورد مقایسه بهتر عمل می‌کند. بنابراین نقشه حساسیت با دقت بالا برای مدیریت مؤثر سیلاب شهری قابل ارائه است. ناچاپا^۵ و همکاران (۲۰۲۰) برای تهیه نقشه پهنه‌بندی سیل در استان سالزبورگ کشور اتریش از روش‌های تصمیم‌گیری چندمعیاره، یادگیری ماشین و مقایسه آن‌ها استفاده کرده‌اند. آن‌ها برای تهیه

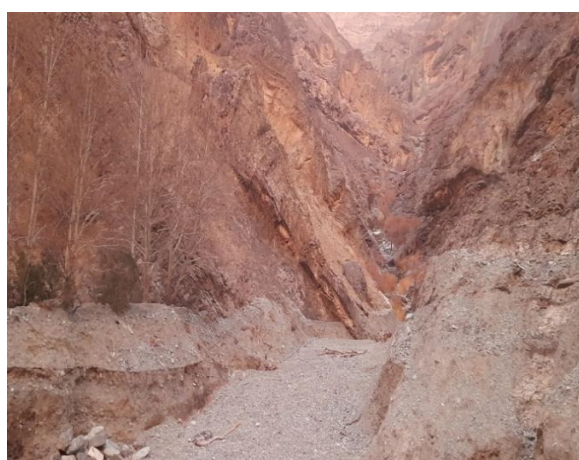
1. Donjiang
2. Typhoon frequency
3. Sadler
4. Shao
5. Nachappa

مواد و روش‌ها



شکل (۱): موقعیت جغرافیایی حوزه آبخیز سد کرج

Figure (1): Location of Karaj Dam Watershed



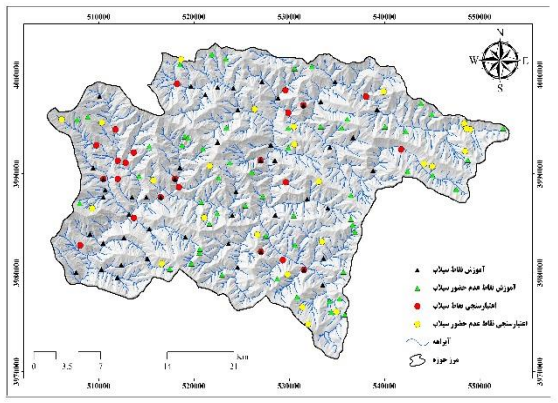
شکل (۲): نمونه‌ای تپیک از شرایط طبیعی حوزه آبخیز سد کرج

Figure (2): A typical sample of natural conditions of Karaj Dam Watershed

ارزیابی احتمال وقوع تولید سیل

برای شبیه‌سازی فرایند رواناب و جریان سیلاب در حوزه‌های آبخیز مدل‌های مختلفی مورد استفاده قرار گرفته‌اند. این مدل‌ها با تمرکز بر فرایندهای فیزیکی و هیدرولوژیکی، به شبیه‌سازی رفتار آب در سیستم‌های طبیعی پرداخته و ابزاری برای برنامه‌ریزی مدیریت منابع آب فراهم می‌کنند (کوریگالاس و کاراتزاس،^۱ ۲۰۱۱). با این حال، پیچیدگی‌های زمانی و مکانی داده‌ها و نیاز به حجم بالای اطلاعات ورودی، از محدودیت‌های عمده این مدل‌ها به شمار می‌رود (برونسترت،^۲ ۲۰۰۳). در این تحقیق پس از بررسی منابع علمی داخلی و خارجی، مدل‌های پیشینه آنتروپی، جنگل تصادفی و ماشین

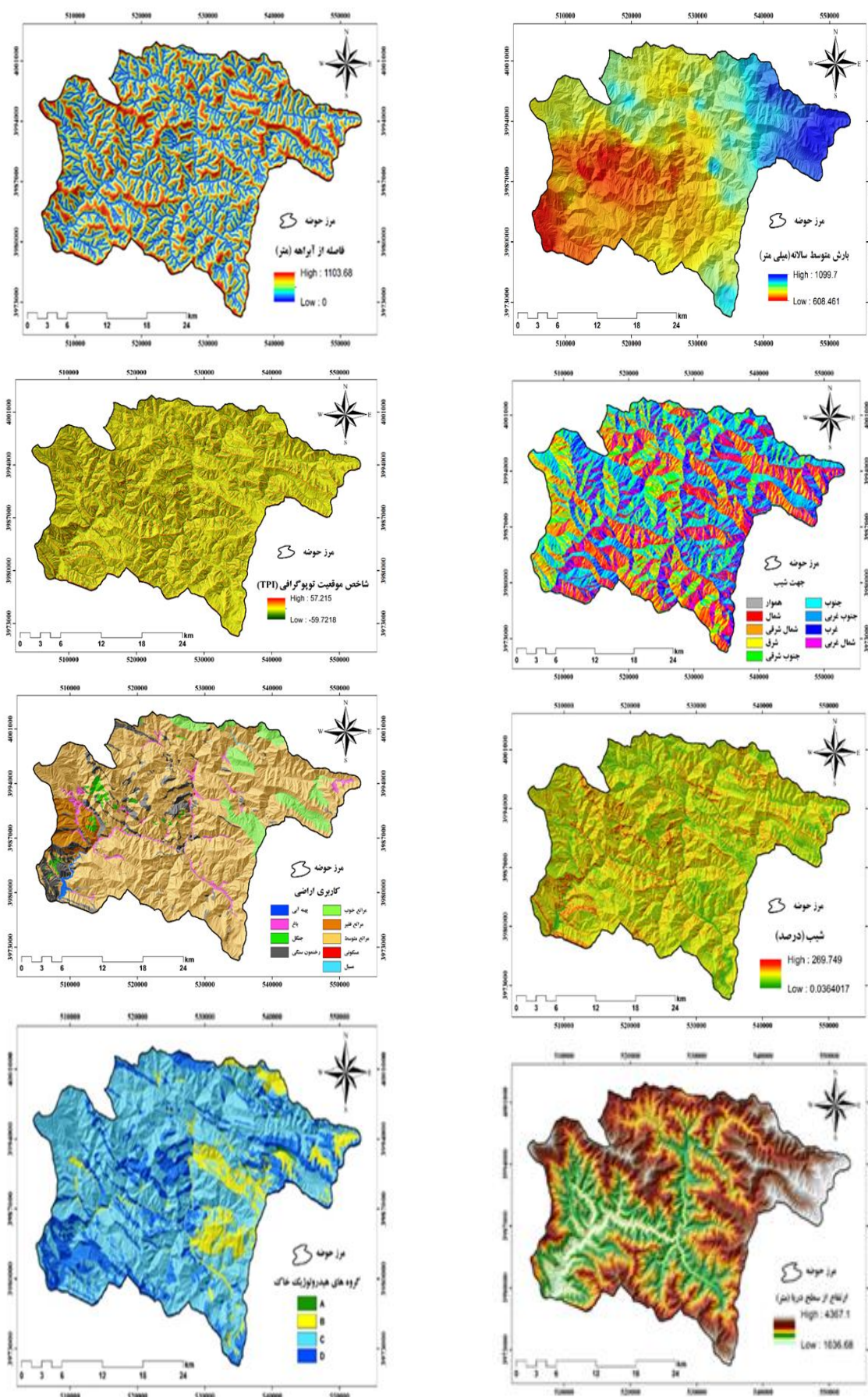
حوزه آبخیز سد کرج در استان البرز، به عنوان محدوده مطالعاتی این پژوهش، به دلیل شرایط خاص طبیعی، اقلیمی و موقعیت جغرافیایی، از حساس‌ترین مناطق سیل‌خیز کشور محسوب می‌شود و هرساله شاهد وقوع سیل‌های سهمگین در این منطقه هستیم که در مواردی علاوه بر خسارت‌های مالی گسترده، تلفات جانی نیز به همراه دارد. همچنین قرارگیری محور مواصلاتی پرتردد و بااهمیت کرج-چالوس در این منطقه، بر اهمیت و ضرورت مطالعه سیل‌خیزی و اولویت‌بندی پتانسیل سیل و یا به عبارتی نقشه‌بندی احتمال وقوع تولید سیل به عنوان اولین اقدام لازم به منظور بررسی ریسک سیل در این خطه از میهن می‌افزاید. این محدوده در شرق استان البرز و در دامنه مختصات جغرافیایی ۳۵ درجه و ۵۱ دقیقه تا ۳۶ درجه و ۱۳ دقیقه عرض شمالی و ۵۱ درجه و پنج دقیقه تا ۵۱ درجه و ۳۵ دقیقه طول شرقی واقع شده است. همچنین مشتمل بر ۱۲ زیرحوزه فرعی است که در تقسیم‌بندی حوزه‌های آبخیز کشور جزء بزرگ‌ترین حوزه‌های دریاچه نمک محسوب می‌شود (شکل ۱). حوزه آبخیز مورد مطالعه از نظر توپوگرافی، کوهستانی بوده و از دامنه‌های پرشیب و سنگلاخی تشکیل شده است (شکل ۲). متوسط ارتفاع این آبخیز ۲۸۲۷ متر و مساحت آن ۸۵۰۶۸ هکتار است. بیشترین و کمترین مقدار بارندگی متوسط سالانه در ایستگاه‌های نسا و سد امیرکبیر با ارتفاعات ۲۸۰۰ و ۱۵۸۸ متر از سطح دریا، به ترتیب معادل ۶۶۷ و ۴۱۸ میلی‌متر اندازه‌گیری شده است (اداره کل منابع طبیعی و آبخیزداری استان البرز، ۱۴۰۰). روش تحقیق انجام‌شده مشتمل بر بخش‌های ارزیابی احتمال وقوع تولید سیل، ارزیابی کارایی مدل‌های یادگیری ماشین، و در نهایت استنتاج و آزمون فرضیه‌ها است. از مصادیق بارز برتری این تحقیق، تفکیک فرایند-محور رخداد‌های سیل و براساس آن تمرکز بر فراوانی وقوع تولید سیل‌های ناگهانی به عنوان تیپ غالب سیل‌های منطقه است که بیشتر در بهار، اوایل تابستان و پس از وقوع بارش‌های همرفتی (جابه‌جایی) رخ داده‌اند.

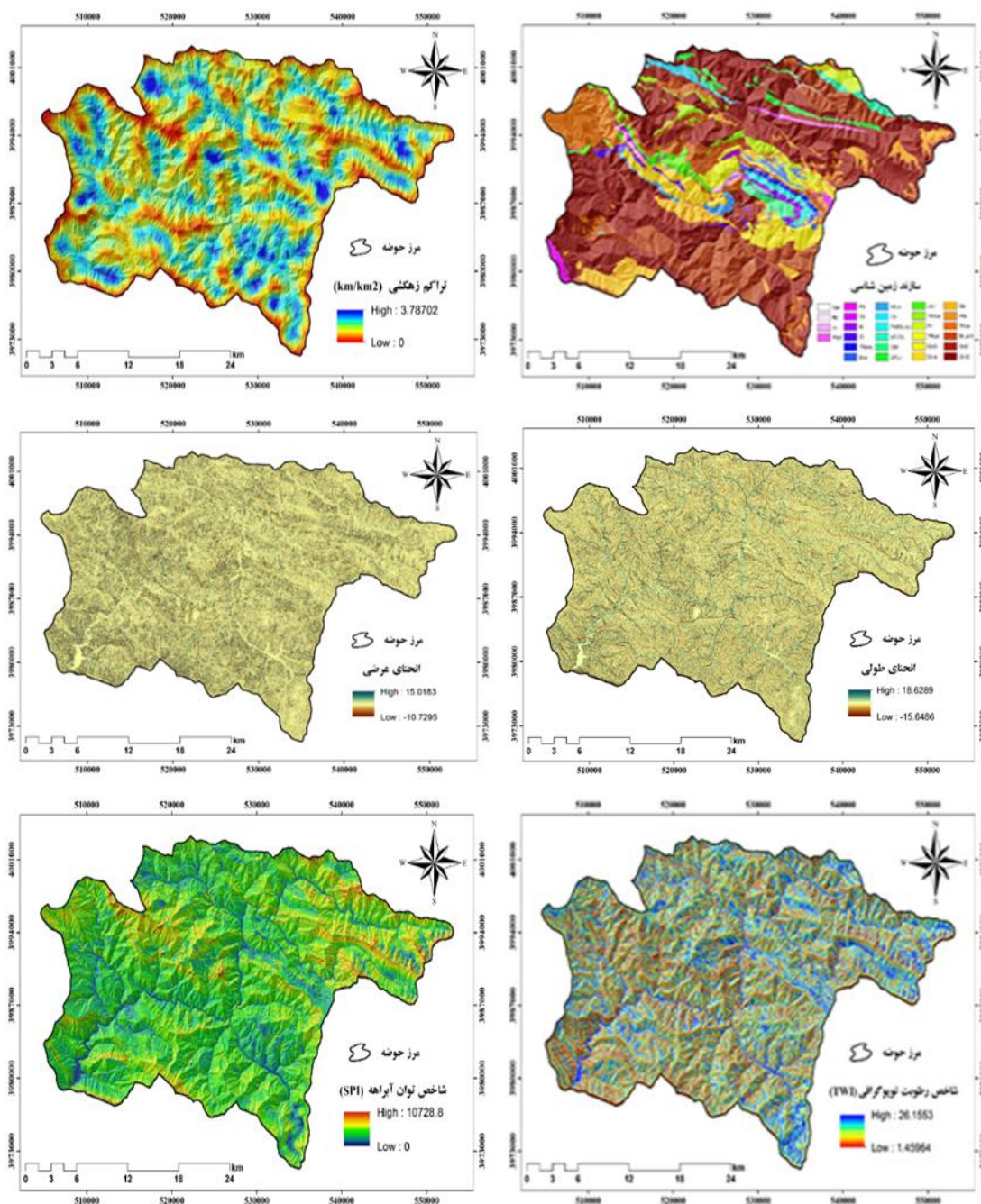


شکل (۳): توزیع مکانی نقاط وقوع و عدم وقوع سیل
Figure (3): Spatial distribution of flood occurrence and non-occurrence places

سپس نسبت به انجام آزمون غربالگری VIF به‌منظور شناخت و حذف عوامل ایجادکننده هم‌خطی بالا (بالاتر از عدد ۱۰) اقدام گردید. از پایگاه داده‌های راداری آلاسکا و از ماهواره Alos Palsar، نقشه مدل رقومی ارتفاع (DEM) با قدرت تفکیک ۱۲/۵ متر تهیه شده و سایر مشتقات نظیر شیب، جهت شیب، انحناهای عرضی و طولی دامنه، شاخص‌های TWI و SPI در نرم‌افزار SAGA-GIS تهیه گردید. نقشه زمین‌شناسی از اندکس‌های ۱:۱۰۰۰۰۰ سازمان زمین‌شناسی و اکتشافات معدنی کشور استخراج شد. نقشه کاربری اراضی از بانک داده اداره کل منابع طبیعی و آبخیزداری استان البرز دریافت و با جدیدترین تصاویر گوگل ارث^۱ تدقیق گردید. همچنین، داده‌های خطوط همباران، آبراهه‌ها و بافت خاک نیز از اداره کل منابع طبیعی و آبخیزداری استان البرز تهیه شد. نقشه گروه‌های هیدرولوژیکی خاک نیز از نقشه بافت خاک تهیه شد. در ادامه، اطلاعات به محیط سامانه اطلاعات جغرافیایی (ArcGIS 10.8) انتقال یافت و نقشه‌بندی عوامل و شواهد سیل در مقیاس ۱:۲۵۰۰ و واحد سلولی در ابعاد ۱۲/۵ متری انجام شد.

بردار پشتیبان براساس قرابت و تشابه بیشتر مناطق هدف تحقیقات گذشته و شرایط محیطی محدوده مورد مطالعه، دقت نتایج و امکان سنجش کارایی مدل‌ها، برای تعیین احتمال مکانی وقوع سیل انتخاب شدند. به‌منظور هرگونه مدل‌سازی مکانی، تهیه دو دسته لایه اطلاعات شواهد میدانی رخداد سیل‌های به‌وقوع‌پیوسته گذشته و عوامل مؤثر در وقوع پدیده سیل، ضروری است. بدین منظور، ابتدا شواهد سیل‌های موجود با استفاده از بازدیدهای میدانی، منابع محلی و سازمانی بررسی و شناسایی شدند. نقشه شواهد رخداد سیل‌های ناگهانی (انتخاب ۸۲ نقطه سیل ناگهانی بعد از غربال ۱۰۶ نقطه وقوع سیل)، به دو دسته تعلیمی برای آموزش مدل‌های یادگیری ماشین (۷۰ درصد از کل داده‌های حضور پدیده معادل ۵۷ نقطه) و اعتبارسنجی برای سنجش کارایی نتایج مدل‌های سه‌گانه پیش‌بینی مکانی (۳۰ درصد از کل داده‌های حضور پدیده معادل ۲۵ نقطه) با استفاده از روش تقسیم‌بندی متقابل ۱۰ تایی تفکیک گردید، که به همراه نقاط عدم حضور پدیده سیل با تعداد برابر (۸۲ نقطه) در نقشه ارائه شده است (شکل ۳). در این روش، طی ۱۰ مرحله، شواهد سیل با جایگشت‌های مختلف به دو دسته آموزشی و اعتبارسنجی تقسیم شدند. این روش به افزایش قدرت یادگیری، قدرت پیش‌بینی، ظرفیت تعمیم نتایج، انتقال‌پذیری مکانی نتایج و افزایش پایداری مدل کمک شایان توجهی می‌کند. در ادامه و مبتنی بر مرور سوابق تحقیقات پیشین، ۱۴ عامل مؤثر بر پتانسیل سیل‌خیزی شامل ارتفاع (متر)، شیب (درصد)، جهت شیب، فاصله از آبراهه (متر)، تراکم زهکشی (کیلومتر بر کیلومتر مربع)، سنگ‌شناسی، بارندگی، کاربری اراضی، شاخص توان آبراهه (SPI)، شاخص موقعیت توپوگرافیکی (TPI)، شاخص رطوبت توپوگرافی (TWI)، گروه‌های هیدرولوژیکی خاک و انحناهای عرضی و طولی دامنه انتخاب شدند.





شکل (۴): نقشه‌های عوامل مؤثر بر تولید سیل در حوزه آبخیز سد کرج

Figure (4): Maps of the factors affecting flood generation in the Karaj Dam Watershed

نیاز ندارد، بلکه از عوامل تأثیرگذار و همچنین نقاط حضور آن پدیده (پهنه‌های سیل برای مدل‌سازی) استفاده می‌کند. در این روش همبستگی عوامل مؤثر در وقوع سیل با شواهد سیل محاسبه شده و با همبستگی متغیرها با ۱۰۰۰۰ نقطه تصادفی از منطقه به عنوان نقاط زمینه (مناطق شبه عدم حضور سیل) مقایسه می‌گردد. در نهایت مدلی به دست می‌آید که دارای بیشترین قدرت تشخیص مناطق مستعد وقوع سیل است. نقشه نهایی

مدل بیشینه آنتروپی^۱

این روش برای اولین بار برای پیش‌بینی پراکنش گونه‌های جانوری طراحی شده و به مرور در زمینه‌های دیگر نظیر پیش‌بینی استعداد زمین‌لغزش استفاده شده است (کرنزادی و همکاران، ۲۰۱۷). مزیت این روش آن است که برای پیش‌بینی رفتار یک گونه یا پدیده به نقاط عدم حضور آن گونه یا پدیده

1. Maximum Entropy model

روی دریا و منازل مسکونی در خشکی شد (جاکولا، ۲۰۰۶^۷). این مدل با شناسایی مجموعه‌ای از نقاط نمونه که به عنوان حیاتی‌ترین عناصر تشخیص الگو در نظر گرفته می‌شوند و حاوی بیشترین اطلاعات هستند، به ترسیم یک خط جداکننده به نام Hyperplane اقدام می‌کنند که در اینجا جداکننده الگوهای حضور و عدم حضور سیل است (کریستیانینی و شل‌کوف، ۲۰۰۲^۸). یکی از مزایای مدل ماشین بردار پشتیبان قابلیت بررسی داده‌ها در ابعاد مختلف است، به طوری که اگر تنها دو دسته ورودی داشته باشیم، مدل به ترسیم یک خط جداکننده اقدام می‌کند. در صورتی که تعداد ورودی‌ها افزایش یابد و به اصطلاح مدل وارد فضای n بعدی شود، مدل با توابعی به نام کرنل^۹ (پایه شعاعی، چندجمله‌ای، سیگموئید و خطی)، ورودی‌ها را به فضای چندبعدی انتقال داده و این بار به جای یک خط، به ترسیم صفحه جداکننده^{۱۰} اقدام کرده و به بهترین تفکیک در الگوی مورد ارزیابی خود دست می‌یابد. این مدل در نرم‌افزار ModEco اجرا می‌شود. در اجرای الگوریتم ماشین بردار پشتیبان نیاز به تعیین مقادیر بهینه^{۱۱} γ و C است. یکی از روش‌های مناسب برای پیدا کردن پارامترهای بهینه استفاده از روش اعتبارسنجی متقابل^{۱۲} و جست‌وجوی خودکار این مقادیر است که در این تحقیق انجام خواهد شد. گاما ضریب کرنل برای poly sigmoid rbf است. هرچه مقدار گاما بیشتر باشد، الگوریتم تلاش می‌کند برازش را دقیقاً براساس مجموعه داده‌های تمرینی انجام دهد و این امر موجب تعمیم یافتن خطا و وقوع مشکل بیش‌برازش^{۱۳} می‌شود. پارامتر C برقراری تعادل بین مرزهای تصمیم‌گیری هموار و طبقه‌بندی نقاط داده تمرینی را کنترل می‌کند. برای داشتن ترکیبی مؤثر از این پارامترها و پیشگیری از بیش‌برازش همواره باید مقدار اعتبارسنجی متقابل را در نظر داشت (کرنزادی و همکاران، ۲۰۱۷).

ارزیابی کارایی مدل‌های حساسیت سیل

منحنی تشخیص عملکرد نسبی^{۱۳} نموداری است که در آن

براساس نقاط عطف منحنی فراوانی تجمعی پیکسل‌ها در پنج کلاس (خیلی کم، کم، متوسط، زیاد و خیلی زیاد) طبقه‌بندی می‌گردد. این مدل در نرم‌افزار ModEco نیز اجرا می‌شود (گوا و لیو، ۲۰۱۰^۱).

مدل جنگل تصادفی^۲

این الگوریتم، جنگلی را به طور تصادفی می‌سازد. جنگل ساخته شده، در واقع گروهی از درخت‌های تصمیم^۳ است. کار ساخت جنگل با استفاده از درخت‌ها اغلب اوقات به روش کیسه‌گذاری^۴ انجام می‌شود. ایده اصلی روش کیسه‌گذاری آن است که ترکیبی از مدل‌های یادگیری، نتایج کلی مدل را افزایش می‌دهد. به بیان ساده، جنگل تصادفی چندین درخت تصمیم ساخته و آن‌ها را با یکدیگر ادغام می‌کند تا پیش‌بینی‌های صحیح‌تر و پایدارتری حاصل شوند. این مدل براساس میانگین‌گیری از نتایج حاصل از تمامی درخت‌های تصمیم مربوط استوار است و برای بسیاری از مجموعه داده‌ها، طبقه‌بندی با صحت بالایی را انجام می‌دهد. برای تعیین تعداد درختان بهینه و تهیه نمودار میانگین مربعات خطای داده‌های آموزش و آزمون، جایی که با افزایش تعداد درختان تغییری در کاهش میزان میانگین مربعات خطای آموزش و آزمون مشاهده نمی‌شود، به عنوان تعداد درخت بهینه انتخاب می‌شود. یکی دیگر از پارامترهای مهم در اجرای الگوریتم جنگل تصادفی تعداد برآوردکننده‌ها k (متغیرهای مستقل) در هر گره برای برآورد متغیر وابسته است. یکی از ساده‌ترین راه‌های ممکن، استفاده از جذر کل تعداد متغیرهای مستقل مورد استفاده در مدل است که این مقدار به عنوان تعداد برآوردکننده‌ها در هر گره مورد استفاده قرار می‌گیرد. این مدل در نرم‌افزار R اجرا می‌شود (لی و همکاران، ۲۰۱۷).

مدل ماشین بردار پشتیبان^۶

این مدل نخست در ترسیم خط ساحل استفاده شد که در آن مدل با موفقیت قادر به تفکیک دو الگوی قایق‌های مسافرتی

7. Jakkula

8. Cristiani & Scholkopf

9. Kernel trick

10. Hyperplane

11. Cross validation

12. Overfitting

13. Relative operating characteristic

1. Guo & Lio

2. Random Forest Model

3. Decision Trees

4. Bagging

5. Lee

6. Support Vector Machine Model

آبراهه‌های اصلی بیشترین میزان احتمال وقوع سیل را دارند. سپس میزان تأثیرگذاری عوامل مؤثر بر تولید سیل در حوزه آبخیز سد کرج نیز تعیین شد (شکل ۸). با توجه به این شکل عوامل ارتفاع، تراکم زهکشی، شیب، میانگین بارندگی و شاخص توان آبراهه بیشترین میزان تأثیرگذاری را بر احتمال وقوع تولید سیل دارند. مساحت کلاس‌های تولید سیل منتج از هر سه مدل مورد استفاده در این تحقیق نیز به دست آمد (جدول ۲). براساس این جدول و نتایج مدل ماشین بردار پشتیبان به‌عنوان مدل برتر، بیشترین و کمترین مساحت مربوط به کلاس‌های تولید سیل خیلی کم و خیلی زیاد به‌ترتیب حدود ۲۱۴۵۰ و ۱۴۰۳۰ هکتار است.

در پایان، نتایج مدل‌ها با دسته داده‌های اعتبارسنجی در نرم‌افزار PMT با استفاده از شاخص مساحت زیرمنحنی عملکرد نسبی سیستم^۴ (AUROC) سنجیده شده است (شکل ۹، ۱۰ و ۱۱). همان‌طور که مشخص است، مدل ماشین بردار پشتیبان با مقدار مساحت زیر منحنی ۰/۸۸۵ به‌عنوان مدل برتر پیش‌بینی مکانی سیل انتخاب شده و نقشه احتمال مکانی تولید سیل حوزه آبخیز سد کرج براساس خروجی این مدل به‌عنوان نقشه نهایی مورد نیاز معرفی می‌گردد (شکل ۷). شاخص کاپا^۵ ۰/۶۸ نیز مؤید دقت بالا و برتری مدل مزبور نسبت به دو مدل دیگر است (جدول ۳). به عبارتی، هر دو شاخص، کارایی مطلوب مدل مورد استفاده را تأیید می‌کنند.

نسبت پیکسل‌هایی که به‌اشتباه توسط مدل به‌عنوان پهنه سیل شناسایی شده‌اند و در واقعیت فاقد واقعه سیل بوده‌اند، روی محور افقی (مثبت کاذب^۱ یا ۱-Specificity) در برابر مقدار مکمل آن یعنی نسبت پیکسل‌هایی که به‌درستی توسط مدل به‌عنوان پهنه سیل شناسایی شده‌اند (مثبت صحیح^۲ یا Susceptibility) روی محور عمودی ترسیم می‌گردد. این منحنی در نرم‌افزار ModEco محاسبه و ترسیم می‌شود. مدلی که بیشترین مساحت زیر این منحنی (AUC) را دارا باشد، دارای عملکرد نسبی بالاتری است. AUC برابر با ۰/۵ معادل مدل خشی (نتایج ناشی از شانس) بوده و هرچه این مقدار به عدد ۱ نزدیک‌تر شود، کارایی مدل افزایش می‌یابد (پونیتوس و اشنايدر، ۲۰۰۱). شاخص کاپا نیز برای ارزیابی مدل‌ها استفاده شد. ضریب کاپا و تحلیل آماری مبتنی بر آن اندازه عددی بین ۱- تا ۱+ است، که هرچه به ۱+ نزدیک‌تر باشد بیانگر وجود توافق متناسب و مستقیم بین پیش‌بینی مدل و واقعیت منطقه است. اندازه‌های نزدیک به ۱- نشان‌دهنده وجود توافق معکوس و اندازه‌های نزدیک به صفر عدم توافق را نشان می‌دهد.

نتایج

براساس نتایج بررسی هم‌خطی عوامل مؤثر بر تولید سیل ناگهانی، کلیه مقادیر فاکتور تورم واریانس^۳ کوچک‌تر از ۱۰ است؛ بنابراین هم‌خطی بحرانی بین داده‌ها منتفی و استفاده از کلیه عوامل در روند مدل‌سازی بلامانع است (جدول ۱). پس از تعلیم سه مدل جنگل تصادفی، بیشینه آنتروپی و ماشین بردار پشتیبان با دسته داده‌های تعلیمی، نقشه‌های احتمال وقوع تولید سیل ایجاد شد (اشکال ۵، ۶ و ۷). براساس مدل بیشینه آنتروپی (شکل ۵) از شرق به‌سمت غرب حوزه آبخیز احتمال وقوع تولید سیل افزایش می‌یابد و در آبراهه‌های اصلی احتمال تولید سیل بیشتر از سایر مناطق است. براساس مدل جنگل تصادفی (شکل ۶) در جنوب غربی حوزه آبخیز بیشترین میزان احتمال سیل خیزی مشاهده می‌شود. نقشه مدل ماشین بردار پشتیبان (شکل ۷) نشان داد که بیشترین میزان سیل‌خیزی از شرق و شمال به‌سمت جنوب غربی حوزه آبخیز افزایش می‌یابد و

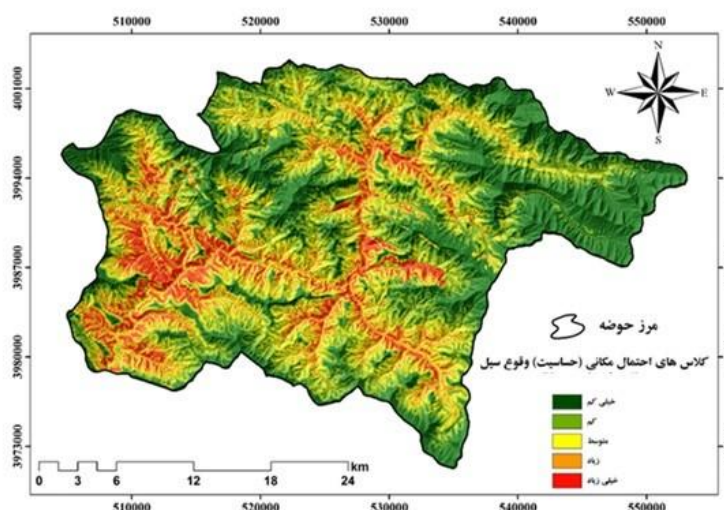
4. Area under relative operating characteristic
5. Kappa index

1. False positive
2. True positive
3. Variance inflation factor

جدول (۱): مقادیر هم خطی عوامل مؤثر بر تولید سیل

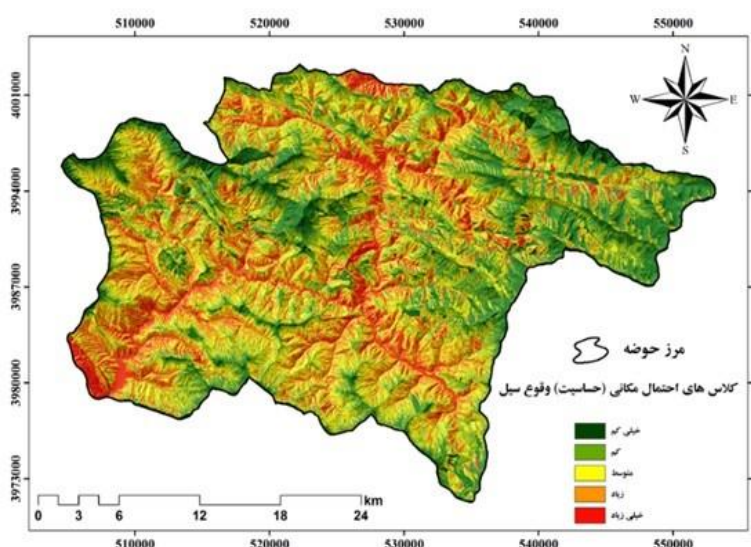
Table (1): The Coordinate values of factors influencing flood generation

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
(Constant)	1.869	0.648		2.885	0.004		
Aspect	-0.026	0.013	-0.136	-2.077	0.040	0.835	1.198
DD	0.007	0.083	0.007	0.088	0.930	0.501	1.995
Stream_D	9.579E-5	0.000	0.027	0.366	0.715	0.638	1.567
Plan_C	0.042	0.047	0.080	0.892	0.374	0.446	2.241
Profile_C	0.022	0.041	0.051	0.541	0.589	0.406	2.465
Rainfall	0.002	0.001	0.359	3.175	0.002	0.279	3.588
Slope	-0.006	0.009	-0.292	-0.639	0.524	0.116	8.580
SPI	0.000	0.000	0.335	0.708	0.480	0.336	2.976
TPI	0.002	0.006	0.033	0.279	0.780	0.252	3.962
TWI	0.031	0.019	0.152	1.634	0.104	0.414	2.418
DEM	-0.001	0.000	-1.049	-4.179	0.000	0.114	8.729
Hydro_G	-0.013	0.047	-0.019	-0.276	0.783	0.772	1.295
Litho	0.004	0.005	0.050	0.731	0.466	0.751	1.332
LULC	-0.015	0.020	-0.052	-0.752	0.453	0.757	1.321



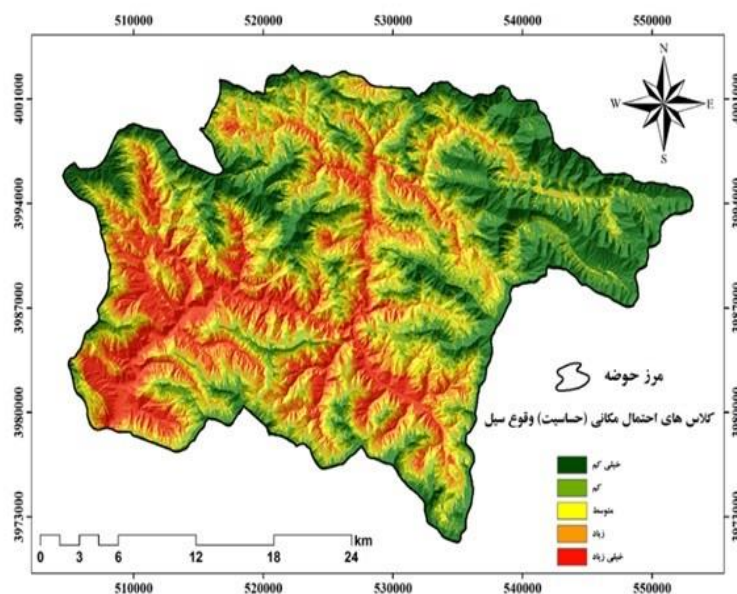
شکل (۵): کلاس‌های احتمال تولید سیل با استفاده از مدل بیشینه آنتروپی

Figure (5): Probability classes of flood generation using maximum entropy model

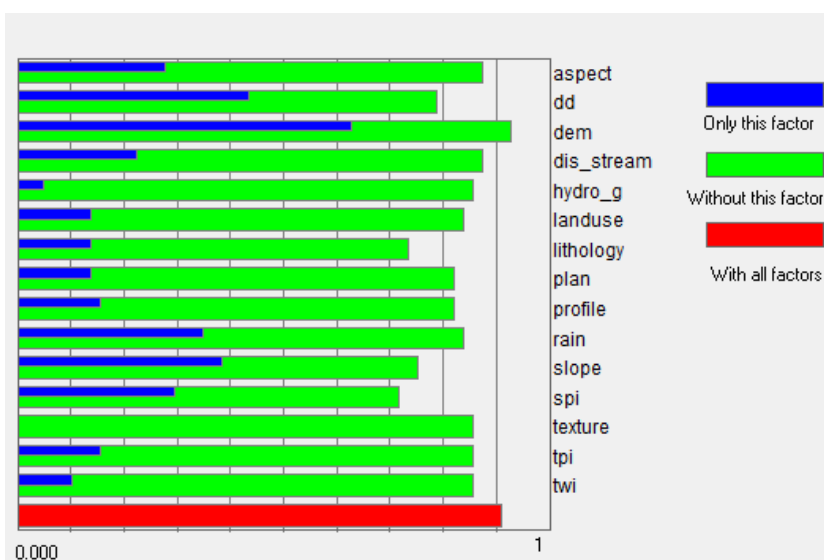


شکل (۶): کلاس‌های احتمال تولید سیل با استفاده از مدل جنگل تصادفی

Figure (6): Probability classes of flood generation using random forest model



شکل (۷): کلاس‌های احتمال تولید سیل با استفاده از مدل ماشین بردار پشتیبان
Figure (7): Probability classes of flood generation support vector machine model

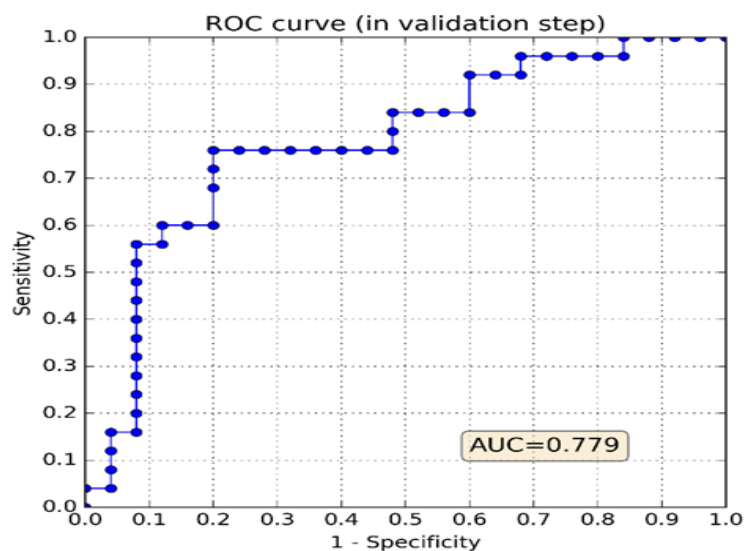


شکل (۸): میزان تأثیرگذاری عوامل مؤثر بر تولید سیل
Figure (8): Influence degree of the effective factors on flood generation

جدول (۲): مساحت کلاس‌های تولید سیل

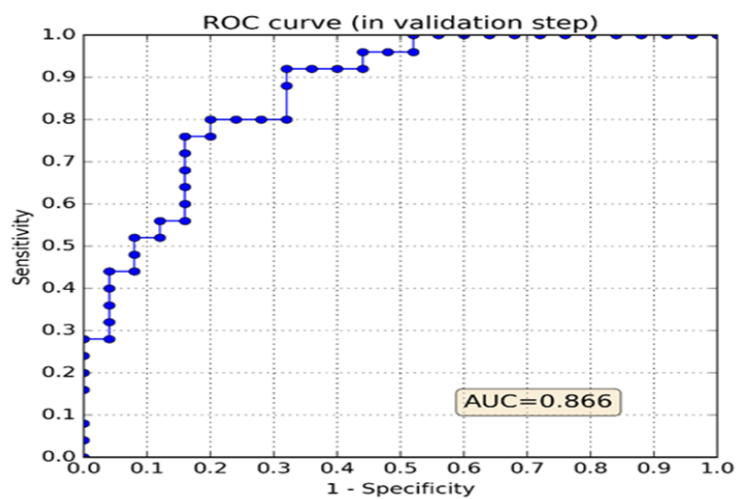
Table (2): Area of flood generation classes

جنگل تصادفی		بیشینه آنتروپی		ماشین بردار پشتیبان		کلاس حساسیت
درصد	هکتار	درصد	هکتار	درصد	هکتار	
۱۹/۴۷	۱۶۵۶۲/۱۹	۳۳/۳۶	۲۸۳۷۶/۶۲	۲۵/۲۲	۲۱۴۵۳/۹۷	خیلی کم
۲۱/۴۶	۱۸۲۵۹/۱۰	۲۵/۹۸	۲۲۰۹۷/۵۶	۲۰/۴۶	۱۷۴۰۳/۱۳	کم
۲۴/۴۷	۲۰۸۱۵/۸۶	۱۹/۱۵	۱۶۲۸۶/۹۹	۱۹/۱۱	۱۶۲۵۲/۵۰	متوسط
۲۴/۱۵	۲۰۵۴۷/۰۷	۱۴/۰۱	۱۱۹۱۹/۷۷	۱۸/۷۲	۱۵۹۲۶/۹۰	زیاد
۱۰/۴۵	۸۸۸۴/۲۲	۷/۵۰	۶۳۸۷/۵۰	۱۶/۴۹	۱۴۰۳۱/۵۰	خیلی زیاد
۱۰۰	۸۵۰۶۸	۱۰۰	۸۵۰۶۸	۱۰۰	۸۵۰۶۸	جمع مساحت



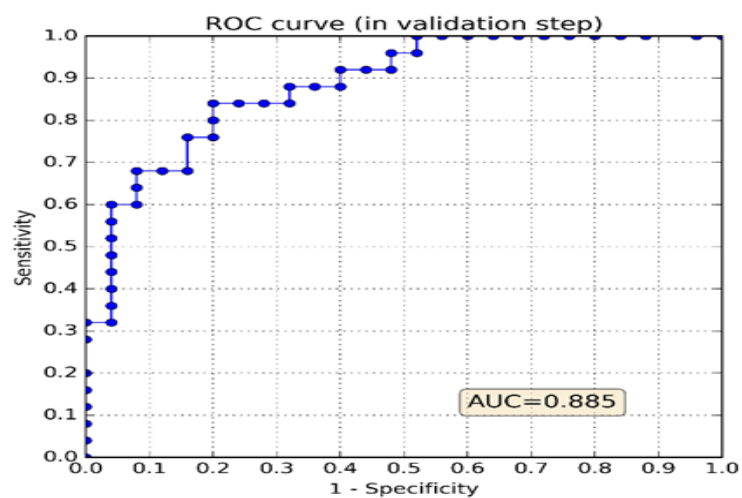
شکل (۹): نمودار ROC منتج از مدل بیشینه آنتروپی

Figure (9): The ROC graph obtained from the maximum entropy model



شکل (۱۰): نمودار ROC منتج از مدل جنگل تصادفی

Figure (10): The ROC graph obtained from the random forest model



شکل (۱۱): نمودار ROC منتج از مدل ماشین بردار پشتیبان

Figure (11): The ROC graph obtained from the support vector machine model

جدول (۳): مقادیر شاخص کاپا به‌ازای مدل‌های مختلف یادگیری ماشین

Table (3): The Kappa index values corresponding to various machine learning models

مدل	شاخص کاپا
بیشینه آنتروپی	۰/۳۰
جنگل تصادفی	۰/۴۸
ماشین بردار پشتیبان	۰/۶۸

زمینه‌ساز تصمیم‌گیری‌های بهینه در جهت مدیریت صحیح منابع طبیعی و کاهش خطرات سیلاب است. بررسی میزان کارایی مدل ماشین بردار پشتیبان، نشان داد که دقت این مدل در شناسایی مناطق حساس و تعیین احتمال وقوع مکانی تولید سیل در مراحل آموزش و اعتبارسنجی بهتر از سایر مدل‌ها بوده است. از این‌رو و براساس نتایج به‌دست‌آمده می‌توان گفت مدل ماشین بردار پشتیبان توانایی زیادی در تشخیص مناطق دارای پتانسیل تولید سیل در این حوزه آبخیز و احتمالاً در آبخیزهای مشابه دارد. به‌دلیل سرعت و دقت زیاد مدل‌های داده‌کاوی پیشنهاد می‌شود در پژوهش‌های مشابه به‌ویژه در کشورهای در حال توسعه که با کمبود منابع مالی و اطلاعاتی مواجه‌اند، نیز این مدل‌ها مورد استفاده قرار گیرند. ترکیب سیستم اطلاعات جغرافیایی با مدل‌های داده‌کاوی به‌منظور تعیین مناطق دارای حساسیت سیل، به‌ویژه در مناطقی که برای تهیه اطلاعات و داده‌های سیل با محدودیت زمان و هزینه مواجه‌اند نیز توصیه می‌گردد. طبق نقشه خروجی مدل ماشین بردار پشتیبان، بیشترین مساحت عرصه مورد مطالعه به کلاس حساسیت خیلی کم (۲۵ درصد) و کمترین سطح (۱۶ درصد) به کلاس خیلی زیاد اختصاص دارد. کلاس‌های متوسط و زیاد نیز ۳۸ درصد از مساحت آبخیز (۳۲۱۷۸ هکتار) را شامل می‌شوند. همچنین براساس نتایج این مطالعه، عوامل ارتفاع، تراکم زهکشی، شیب و میانگین سالانه بارندگی بیشترین نقش را در تولید سیل‌های ناگهانی به‌وقوع پیوسته گذشته داشته‌اند که با نتایج تحقیق اسدی نلیوان و همکاران (۲۰۱۸) تطابق کامل دارد. عامل تراکم زهکشی و کاهش زمان تمرکز که باعث افزایش تجمع جریان‌های سطحی و درنهایت افزایش میزان تولید سیل می‌شوند، با یافته‌های مطالعات حق‌زاده و همکاران (۲۰۱۷) منطبق است. شیب‌های بسیار تند که فرصت جذب و نفوذ



شکل (۱۲): نمونه‌ای از تخریب سیل در حوزه آبخیز سد کرج

Figure (12): A case of flood destruction in Karaj Dam Watershed

بحث و نتیجه‌گیری

شایان توجه است که همان‌طور که در تحقیق حاضر بدان پرداخته شد، قبل از استفاده از رخداد های سیل‌های پیشین در آبخیز برای مراحل یادگیری و اعتبارسنجی، ضرورت دارد با نگاه فرایند تشکیل سیل، رخداد های سیل در حوضه مورد بررسی با توجه به خصوصیات زمانی و مکانی بارش‌ها و نیز خصوصیات سیل‌ها به انواع سیل‌های ناگهانی، رودخانه‌ای و ناشی از ذوب برف تفکیک شوند (برای توضیح بیشتر رجوع شود به لالوزایی و همکاران، ۲۰۲۲). بدین ترتیب فرایند تشکیل سیل در مدل یادگیری ماشین لحاظ و منجر به حصول نتایج منطقی‌تری خواهد شد. این مورد از آنجایی اهمیت دارد که اشکال نپرداختن اکثر مدل‌های یادگیری ماشین به فرایندها را مورد توجه قرار می‌دهد و از بی‌توجهی به فرایندها پرهیز می‌کند.

براساس نتایج حاصل از این پژوهش، استفاده از مدل‌های یادگیری ماشین در تحلیل و بررسی احتمال تولید سیل در حوزه‌های آبخیز، موجب افزایش دقت در پیش‌بینی شده و

گیرد. نقشه کلاس‌های احتمال مکانی تولید سیل‌های ناگهانی می‌تواند به سیاست‌گذاران، مدیران و برنامه‌ریزان، برای مدیریت مناطق سیل‌خیز و اتخاذ تصمیمات کارشناسی، کمک شایانی نماید. از طرفی با به‌اشتراک‌گذاری این اطلاعات با سایر بهره‌برداران، می‌توان ضمن آگاه‌سازی به منظور دوری جستن از مکان‌های پرخطر و افزایش ظرفیت سازگاری و ظرفیت‌مدارا جوامع محلی، گام‌های مفیدی در جهت کاهش ریسک سیل در منطقه برداشت.

قدردانی

مفتخر است از مدیرکل محترم منابع طبیعی و آبخیزداری استان البرز، معاون محترم آبخیزداری و رئیس محترم اداره آموزش و ترویج آن اداره‌کل که در انجام بازدیدهای میدانی و تهیه داده‌ها، آمار و اطلاعات لازم همکاری نمودند، صمیمانه سپاس و قدردانی به عمل آید. این مقاله بخشی از نتایج رساله دکتری در رشته آبخیزداری است و دانشگاه علوم کشاورزی و منابع طبیعی گرگان پشتیبانی مالی رساله دکتری را متقبل شده است.

رواناب در خاک را به‌طور فزاینده‌ای کاهش می‌دهد، از دیگر عوامل افزایش وقوع سیل در این حوزه آبخیز به شمار می‌روند. توجه به این نکته ضروری است که میزان تأثیر متغیرهای محیطی از یک منطقه به منطقه‌ای دیگر متفاوت خواهد بود. بنابراین، اولویت‌بندی اهمیت عوامل محیطی یک منطقه را نمی‌توان به منطقه‌ای دیگر تعمیم داد. از این‌رو، برنامه‌ریزی مدیریت سیل مبتنی بر اثرگذاری عوامل محیطی هر منطقه باید براساس شرایط محیطی همان منطقه انجام گیرد. براساس نقشه احتمال وقوع تولید سیل تهیه‌شده، باید از هرگونه طرح‌های توسعه‌ای (ایجاد مناطق مسکونی، احداث ابنیه و تأسیسات مهم و زیربنایی) در بخش‌هایی که دارای پتانسیل سیل بالایی هستند، جلوگیری گردد. اجرای عملیات مدیریتی و کنترلی سیل از قبیل احداث سازه‌های آبخیزداری، بندهای کنترل جریان، دیواره‌های سیل‌گیر در رودخانه‌ها، احیای پوشش گیاهی در بالادست، رعایت الگوی صحیح کاربری اراضی، جلوگیری از تجاوز به حریم قانونی رودخانه‌ها، قانونمند کردن برداشت مصالح می‌تواند در محدوده‌های حساس مورد توجه و تأکید جدی قرار

منابع

1. Akbari, M., Ownegh, M., Asgari, H. R., Sadoddin A., Khosravi H. (2016). Soil erosion risk assessment using the CORINE model (Case Study: Semi-arid region in Golestan Province). *Desert Ecosystem Engineering*, 5 (12), 63-78.
2. Asadi Nalivan, O., Teimori, A., Rahmati, A., & Kornejadi, A. (2108). River management and flood control using unstructured methods. Noroozi Publications. 204p. Gorgan Iran.
3. Bai, M., Tahmasbipour, N., Zinivand, H., & Sadoddin, A., Kh, M. (2022). Predicting the effect of intervention scenarios in the vulnerability of local communities in a flood Zone (The Qarachai River, Ramyan-Golestan province, Iran). *Journal of Water and Soil Conservation*, 30 (2), 51-72.
4. Bronstert, A. (2003). Floods and climate change: interactions and impacts. *Risk Analysis: An International Journal*, 23 (3), 545-557.
5. Choubin, B., Moradi, E., Golshan, M., Adamowski, J., Sajedi-Hosseini, F., & Mosavi, A. (2019). An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. *Science of the Total Environment*, 651, 2087-2096.
6. Cristianini, N., & Scholkopf, B. (2002). Support vector machines and kernel methods: the new generation of learning machines. *American Association for Artificial Intelligence. A. I. Magazine*, 23 (3), 31-31.
7. Guo, Q., & Liu, Y. (2010). ModEco: an integrated software package for ecological niche modeling. *Ecography*, 33 (4), 637-642.
8. Haghighzadeh, A., Siahkamari, S., Haghiabi, A. H., & Rahmati, O. (2017). Forecasting flood-prone areas using Shannon's entropy model. *Journal of Earth System Science*, 126, 1-11.
9. Jakkula, V. (2006). Tutorial on support vector machine (SVM). *School of EECS, Washington State University*, 37 (2.5), 3.
10. Kjeldsen, T. R. (2010). Modelling the impact of urbanization on flood frequency relationships in the UK. *Hydrology Research*, 41 (5), 391-405.
11. Kornejadi, A., Ownegh, M., & Bahremand, A. (2017). Landslide susceptibility assessment using maximum entropy model with two different data sampling methods. *Catena*, 152, 144-162.
12. Kourgialas, N.N., & Karatzas, G. P. (2011). Flood management and a GIS modelling method to assess flood-hazard areas- a case study. *Hydrological Sciences Journal-Journal des Sciences Hydrologiques*, 56 (2), 212-225.
13. Lalezari, A., Sadoddin, A., Zare Garizi, A., &

- Sheikh, V.B. (2022). Process-based flood type separation for partial duration series (Sample study: the Gorganrood River Sub-Basins). *Journal of Iran Water Resources Research*, 3 (62), 18-30.
14. Lee, S., Kim, J., Jung, H., Lee, M., & Lee, S. (2017). Spatial prediction of flood susceptibility using random-forest and boosted-tree models in Seoul metropolitan city, Korea. *Geomatics, Natural Hazards and Risk*, 8 (2), 1185-1203.
 15. Miller, A. J. (1990). Flood hydrology and geomorphic effectiveness in the central Appalachians. *Earth Surface Processes and Landforms*, 15 (2), 119-134.
 16. Nachappa, T.G., Tavakkoli, S., Gholamnia, Kh. Ghorbanzadeh, O., Rahmati, O., & Blaschke, T. (2020). Flood susceptibility mapping with machine learning, multi-criteria decision analysis and ensemble using Dempster Shafer Theory. *Journal of Hydrology*, 590, 1252.
 17. Rahmati, O., Zeinivand, H., & Besharat, M. (2016). Flood hazard zoning in Yasooj region, Iran, using GIS and multi-criteria decision analysis. *Geomatics, Natural Hazards and Risk*, 7 (3), 1000-1017.
 18. Sadler J.M., Goodall, J.L., Morsy, M.M., & Spencer, K. (2018). Modeling Urban Coastal Flood Severity from Crowd-Sourced Flood Reports Using Poisson Regression and Random Forest. *Journal of Hydrology*.
 19. Singh, A. K. (2010). Landslide management: concept and philosophy. *Disaster Prevention and Management: An International Journal*, 19 (1), 119-134.
 20. Shao, G., Pang, B., Xu, Z., Peng, D., & Xu, L. (2019). Assessment of urban flood susceptibility using semi-supervised machine learning model. *Science of the Total Environment*, 629, 940-949.
 21. Special Reporting Committee on Iran Floods in 2019. (2020). Final report: narration, analysis, lesson learnt, and recommendations. University of Tehran.
 22. Tehrany, M. S., Pradhan, B., & Jebur, M. N. (2014a). Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. *Journal of Hydrology*, 512, 332-343.
 23. Tehrany, M. S., Lee, M. J., Pradhan, B., Jebur, M. N., & Lee, S. (2014b). Flood susceptibility mapping using integrated bivariate and multivariate statistical models. *Environmental Earth Sciences*, 72 (10), 4001-4015.
 24. Tehrany, M. S., Pradhan, B., Mansor, S., & Ahmad, N. (2015). Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. *Catena*, 125, 91-101.
 25. Wang, Z., Lai, C., Chen, X., Yang, B., Zhao, S., & Bai, X. (2015). Flood hazard risk assessment model based on random forest. *Journal of Hydrology*, 527, 1130-1141.

Zoning of Flash Flood Generation Probability Using Machine Learning Techniques (A Case Study: Karaj Dam Watershed, Iran)

Abdolvahed Kordi,¹ Amir Sadoddin,^{2*} Gahangir Mohammadi,³ VahedBerdi Sheikh,⁴ Omid Asadi Nalivan⁵

Received: 17/04/2025

Accepted: 13/07/2025

Extended Abstract

Introduction: Flooding stands as a major hydro-climatic extreme event and one of the most severe natural disasters, posing significant threats to socio-environmental systems. Given its multifaceted and dynamic nature, floods are influenced by numerous factors, with both natural and anthropogenic agents collectively impacting their magnitude, intensity, extent, and duration. A fundamental step in managing flood-prone areas is to identify the spatial extent of potential floods. By spatially analyzing and investigating floodplains, a risk-based management framework can be established, leading to effective preventive measures. Therefore, flood prevention and management necessitate the initial delineation of high flood-potential areas using modeling methods, followed by the identification of the underlying drivers of these events. In recent years, Iran has experienced an intensifying frequency and impact of flash floods, resulting in considerable damage. The Karaj Dam Watershed in Alborz Province, in particular, has been severely affected, creating substantial challenges for travelers, tourists, and local residents. This ongoing issue underscores the urgent need for enhanced flood management strategies. Consequently, this research focuses on the spatial analysis and susceptibility assessment of flash flood-prone areas. Our aim is to provide essential and practical insights to inform planning, decision-making, and the implementation of various flood mitigation measures by relevant authorities and stakeholders.

Materials and Methods: To investigate the probability of flash flood generation across the Karaj Dam Watershed in Alborz Province, we first compiled a comprehensive map of past flash flood events. This involved field visits, consultation with local resources, and integrating data from various sources within the ArcGIS 10.8 software environment. Based on an extensive review of relevant literature, we identified fourteen variables hypothesized to influence flash flood occurrence. To ensure an optimal combination of predictors, we applied the Variant Inflation Factor (VIF) test for feature selection, screening out redundant variables. For model development and validation, the flash flood event data was divided using a 10-fold cross-validation method, with 70% allocated for training and 30% for validation. We then processed and analyzed the spatial data, including environmental variables and hydrologic information, using three distinct machine learning algorithms: Maximum Entropy, Random Forest, and Support Vector Machine. The accuracy of each susceptibility model was rigorously evaluated using comprehensive criteria, including the Relative Operating Characteristic (ROC) curve and the Kappa index. The model demonstrating the most acceptable accuracy was subsequently selected to generate a spatial probability map of flash flood generation for the study area. The insights derived from this research offer valuable contributions to flood hazard assessment and management efforts.

1. PhD Student, Department of Watershed Management, University of Agricultural Sciences and Natural Resources, Gorgan, Iran. E-mail: Vahed.kordi@yahoo.com

2. Corresponding Author, Professor, Department of Watershed Management, University of Agricultural Sciences and Natural Resources, Gorgan, Iran. E-mail: amir.sadoddin@gau.ac.ir

3. Associate Prof., Department of Forest, University of Agricultural Sciences and Natural Resources, Gorgan, Iran. E-mail: Mohamadi.jahangir@gau.ac.ir

4. Professor, Department of Watershed Management, University of Agricultural Sciences and Natural Resources, Gorgan, Iran. E-mail: V.sheikh@yahoo.com

5. Assistant Prof., Department of Water Engineering, Faculty of Agriculture, University of Maragheh. Iran. E-mail: O.asadi@maragheh.ac.ir

Results: Our analysis revealed that the Support Vector Machine (SVM) model proved to be the most suitable algorithm for determining flash flood susceptibility within the study area. This model achieved an impressive Area Under the Curve (AUC) value of 0.88 and a Kappa coefficient of 0.68, demonstrating strong predictive capability and outperforming both the Maximum Entropy and Random Forest models. The resulting flood susceptibility map indicates that 53% of the study area, approximately 46,200 hectares, exhibits moderate, high, or very high flash flood potential. Conversely, the remaining 47% of the watershed, about 38,860 hectares, falls into the low and very low flood probability classes. Furthermore, our research identified the key contributing factors to flash flooding in the region. In order of significance, elevation, drainage density, slope, and average annual rainfall emerged as the most important drivers influencing flood occurrence.

Discussion and Conclusions: Historically, the Karaj Dam Watershed has experienced a unique interplay of climatic and geomorphological factors that amplify flash flood risk. Extreme cold conditions at higher altitudes lead to substantial winter snow accumulation. When this is followed by heavy rainfall in spring and summer, it causes a rapid increase in water flow, resulting in significant flash flooding. This aligns perfectly with the watershed's natural characteristics: its steep slopes limit the soil's capacity to absorb rainwater, and when combined with high drainage density and intense, short-duration rainfall, these factors significantly elevate the risk of flash flooding. The findings from this research are crucial for both improving our understanding of flood phenomena and for practically identifying flood-prone areas. By pinpointing the key contributing factors to flash flood occurrences, we can prioritize management measures tailored to the region's specific conditions, ultimately aiming to reduce flood-related damages. The generated flood susceptibility map serves as a vital tool, providing a roadmap for implementing effective preventive measures. Ultimately, this research offers an effective framework for developing robust flood hazard management plans. It can significantly contribute to reducing human losses and economic damages by guiding informed decision-making and planning processes. We strongly recommend that responsible organizations and agencies utilize these insights, including the detailed zoning of flood susceptibility and the identified effective factors, to optimize financial resource allocation and ensure efficient use of public resources in flood mitigation efforts.

Keywords: Data mining, Flood occurrence probability, Flood generation, Flood zonation.