نشریه علمی

علوم و مهندسی آبخیزداری ایران
Iran-Watershed Management
Science & Engineering

سال هفدهم- شماره ٦١- تابستان ١٤٠٢

Vol. 17, No. 61, Summer 2023

مقایسه مدل‌های یادگیری ماشینی در پهنه‌بندی حساسیت سیل‌خیزی حوزه آبخیز سد کرج

مهدی تیموری[1]، فرزانه وکیلی تجره[2]، ملیحه مزین[3] و مرضیه رمضانی[4]

تاریخ دریافت:۱۴۰۱/۰۲/۱۲ تاریخ پذیرش: ۱۴۰۱/۰۵/۱۱

چکیده

هدف از این تحقیق تعیین مناطق دارای حساسیت سیل‌خیزی با استفاده از روش‌های CART, RF و BRT می‌باشد. ۱۲ عامل موثر بر پتانسیل سیل شامل ارتفاع، شیب، جهت شیب، فاصله از آبراهه، سنگ‌شناسی، بارندگی، کاربری اراضی، SPI، TPI، TWI، انحناء توپوگرافی و RSP انتخاب شدند. از مجموع ۸۲ نقطه سیل، به صورت تصادفی و نسبت ۷۰ به ۳۰ درصد به عنوان داده‌های آزمون و اعتبارسنجی طبقه‌بندی شدند. همچنین جهت تعیین مهم‌ترین پارامترها از روش جنگل تصادفی استفاده شد. از منحنی ROC نیز جهت قدرت پیش‌بینی مدل استفاده شد. بر اساس روش جنگل تصادفی لایه‌های ارتفاع، فاصله از آبراهه، بارندگی، کاربری اراضی و RSP به ترتیب مهم‌ترین عوامل موثر بر حساسیت و احتمال وقوع سیل بودند. بر اساس نمودار ROC میزان دقت مدل RF به عنوان مدل برتر، هم در بخش آموزش (۰/۸۸٤) و هم اعتبارسنجی (۰/۸٥٦) خیلی خوب بوده است. بر اساس نقشه نهایی حساسیت سیل ۳۲/۷ درصد از پهنه حوزه مورد مطالعه دارای حساسیت سیل متوسط رو به بالا می‌باشد. نتایج نشان داد که با توجه به دقت عالی نقشه توزیع مکانی حساسیت سیل، نتایج این پژوهش می‌تواند مورد استفاده تصمیم‌گیران، مدیران محلی و مسئولان به منظور کاهش خسارات قرار گیرد.

کلیدواژه‌ها: *جنگل تصادفی، سد کرج، حساسیت سیل، یادگیری ماشینی.*

۱- نویسنده مسئول و استادیار گروه مهندسی طبیعت، دانشکده کشاورزی شیروان، دانشگاه بجنورد، پست الکترونیک: m.teimouri@yahoo.com
۲- دانشجوی دکتری علوم و مهندسی آبخیز، گروه احیای مناطق خشک و کوهستانی، دانشکده منابع طبیعی، دانشگاه تهران.
۳-استادیار گروه مرتع و آبخیزداری، دانشکده منابع طبیعی، دانشگاه صنعتی خاتم‌الانبیاء بهبهان.
٤- دانش آموخته کارشناسی مرتع و آبخیزداری دانشگاه یزد.

مقدمه

امروزه پدیده سیل یکی از پیچیده‌ترین رخدادهای طبیعی است که نیاز به مدیریت سیلاب را پیش از وقوع واقعه نمایان می‌کند و بیش از سایر بلایای دیگر، هر ساله در نقاط مختلف دنیا و به خصوص در کشور ایران منجر به ایجاد خسارت‌های مالی و جانی و هم‌چنین تخریب اراضی مرتعی و کشاورزی و منازل مسکونی می‌شود. بنابراین تهیه نقشه حساسیت سیلاب اولین گام در برنامه مدیریت جامع سیلاب است. بر اساس تحقیقات کوهورست و همکاران سیل شایع‌ترین رخداد طبیعی در جهان محسوب می‌شود [۱۰]. سیل‌ها در مقیاس جهانی ویرانگرترین پدیده طبیعی و موثرترین عامل در تعداد مرگ و میر و خسارات مالی هستند [۸]. در سال‌های اخیر وقوع سیل بیش از آن که یک پدیده طبیعی باشد، از عوارض به هم خوردن تعادل اکوسیستم و شرایط جغرافیایی نشات گرفته است. افزایش جمعیت و در پی آن کمبود سکونت‌گاه برای ساخت و ساز خصوصاً در کلان‌شهرها، موجب تغییر مرفولوژی حوزه‌های‌آبخیز، تشدید ریسک سیلاب‌های احتمالی و تهدید امنیت و آرامش روانی ساکنان این مناطق گردیده است. تجاوز به حریم رودخانه‌ها، مسیل‌ها و آبراهه‌ها، الگوی زهکشی طبیعی را تغییر می‌دهد و باعث جاری شدن جریان‌های مازاد بر ظرفیت آبراهه‌ها و مسیل‌های شهری می‌شود [۲]. با توجه به این‌که پدیده سیل ماهیت چندبعدی و پویا دارد، قطعاً تمامی عوامل طبیعی، انسانی و سازمانی - مدیریتی در کنار یکدیگر بر وقوع، شدت، گستره و تداوم سیل تاثیر خواهند داشت. از جمله می‌توان به فاکتورهای تأثیرگذار نظیر تغییر الگوی بارش، عدم رعایت اصول آمایش سرزمین و تغییر کاربری اراضی، جنگل‌زدایی، آسیب‌پذیری جوامع انسانی و شرایط اقتصادی، مکان‌یابی نادرست زیرساخت‌ها نظیر جاده‌ها و سازه‌های تقاطعی به ویژه ظرفیت آبگذری پل‌ها، عدم مدیریت صحیح دشت‌های سیلابی، فقدان سند جامع راهبردی مخاطرات محیطی و برنامه اقدام جامع، توجه نکردن به رویکرد مدیریت جامع حوزه‌های‌آبخیز، عدم کنترل و بررسی به موقع مسیرهای خروج سیل قبل از وقوع و در نهایت کمبود برنامه‌های آموزشی لازم قبل از وقوع برای کاهش آسیب‌پذیری اشاره نمود. مسلم است ذکر تنها یک عامل به عنوان علت وقوع سیل، موجب فاصله گرفتن از ریشه‌یابی جامع و اصولی موضوع خواهد شد. در سال‌های اخیر تکنیک‌های پیشرفته مبتنی بر سیستم اطلاعات جغرافیایی و سنجش از دور در ارزیابی نقشه‌های حساسیت سیل و سایر بلایای طبیعی توسط روش‌های داده‌کاوی

[Downloaded from jwmsei.ir on 2026-01-29]
[DOR: 20.1001.1.20089554.1402.17.61.4.8]

و یادگیری ماشینی اجرا شده است و دلیل استفاده از روش‌های چندبعدی مذکور ماهیت پیچیده حوزه‌های آبخیز و بلایای طبیعی است. یکی از روش‌های کاهش خسارت سیل، پهنه‌بندی سیل می‌باشد که اطلاعات ارزشمندی را در رابطه با سیلاب‌ها و اثرات آن بر اراضی دشت سیلابی و حریم رودخانه‌ها ارائه می‌دهد [۱۷]. به دلیل هزینه‌های بالای کنترل سیلاب با استفاده از روش‌های سازه‌ای تکنیک‌های یادگیری ماشینی در سال‌های اخیر به صورت کاملا کاربردی در بحث‌های پهنه‌بندی و تهیه نقشه‌های حساسیت سیل مورد استفاده قرار گرفته است [۷، ۱۱، ۱۹ و۱۲]. در این زمینه مطالعات زیادی با روش‌های مختلف انجام شده است. آوند و همکاران (۲۰۲۰) با استفاده از دو مدل جنگل تصادفی و خطی تعمیم‌یافته بیزین به بررسی حساسیت سیل در حوزه آبخیز تجن در شهرستان ساری پرداختند. آن‌ها از ۱۳ عامل موثر بر سیل استفاده کردند و بیان کردند که سه پارامتر ارتفاع، فاصله از آبراهه و شیب به ترتیب مهم‌ترین عوامل موثر بر سیل در این حوزه هستند. همچنین بر اساس مقدار AUC مدل جنگل تصادفی را برتر دانستند [۳]. عرب‌عامری و همکاران [۲]. به مقایسه کارایی روش‌های آماری (EBF, FR) و تصمیم‌گیری چند معیاره (TOPSIS, VIKOR) برای تعیین حساسیت سیل در مناطق شمال ایران پرداختند. نتایج نشان داد که عوامل شیب، فاصله از آبراهه و کاربری اراضی از اهمیت کلیدی در مدل‌سازی حساسیت سیل در منطقه مورد مطالعه برخوردارند و مدل EBF با نرخ پیش‌بینی بالاتر نسبت به سایر مدل‌ها به عنوان مدل برتر معرفی گردید [۲]. چوبین و همکاران (۲۰۱۹) حساسیت سیل را با استفاده از مدل‌های تحلیل چند متغیره، طبقه‌بندی و درخت‌های رگرسیون (CART) و ماشین بردار پشتیبان مورد بررسی قرار دادند. نتایج حاصل از تحقیق آن‌ها نشان داد که مدل تحلیل چند متغیره بالاترین دقت پیش‌بینی (۸۹ درصد) را داشته و مدل‌های SVM (۸۸ درصد) و CART (۸۳ درصد) با اختلاف جزئی در مراتب بعدی اهمیت قرار دارند. همچنین، تحلیل حساسیت عوامل نشان داد که عوامل شیب، تراکم‌زهکشی و فاصله از رودخانه جزء مهم‌ترین عوامل در تهیه نقشه حساسیت سیل بوده‌اند [٦]. ناچپا و همکاران (۲۰۲۰) جهت تهیه نقشه پهنه‌بندی سیل در استان سالزبورگ کشور اتریش از روش‌های تصمیم‌گیری چند معیاره (AHP،ANP) و یادگیری ماشینی (RF،SVM) و مقایسه آن‌ها استفاده کرده‌اند و جهت بهینه‌کردن نقشه‌های مذکور از روش دمپستر- شفر بهره بردند. آن‌ها جهت تهیه نقشه‌ها از ۱۱ عامل موثر بر سیل استفاده کردند. نتایج نشان داد که عملکرد بهتر مدل‌ها به ترتیب RF، SVM، ANP و AHP بوده است. در واقع روش‌های یادگیری ماشینی عملکرد بهتری نسبت به روش‌های تصمیم‌گیری چند معیاره دارند. لازم به ذکر است که مدل دمپستر- شفر به افزایش دقت مدل‌ها کمک کرد. در نهایت آن‌ها بیان کردند که ترکیب ٤ تا مدل با همدیگر بهترین نتیجه را ارائه می‌دهد [۱۵]. جمع‌بندی مرور منابع نشان‌دهنده اهمیت تعیین مناطق دارای حساسیت سیل جهت مدیریت آن با توجه به خسارات گسترده سیل و همچنین توانایی بالای روش‌های یادگیری ماشینی در این امر مهم و حیاتی می‌باشد. هدف از این تحقیق پهنه‌بندی حساسیت سیل در حوزه آبخیز سد کرج با استفاده از روش‌های CART, RF و BRT و همچنین تعیین و اولویت‌بندی عوامل موثر بر سیل با استفاده از روش RF است.

مواد و روش‌ها

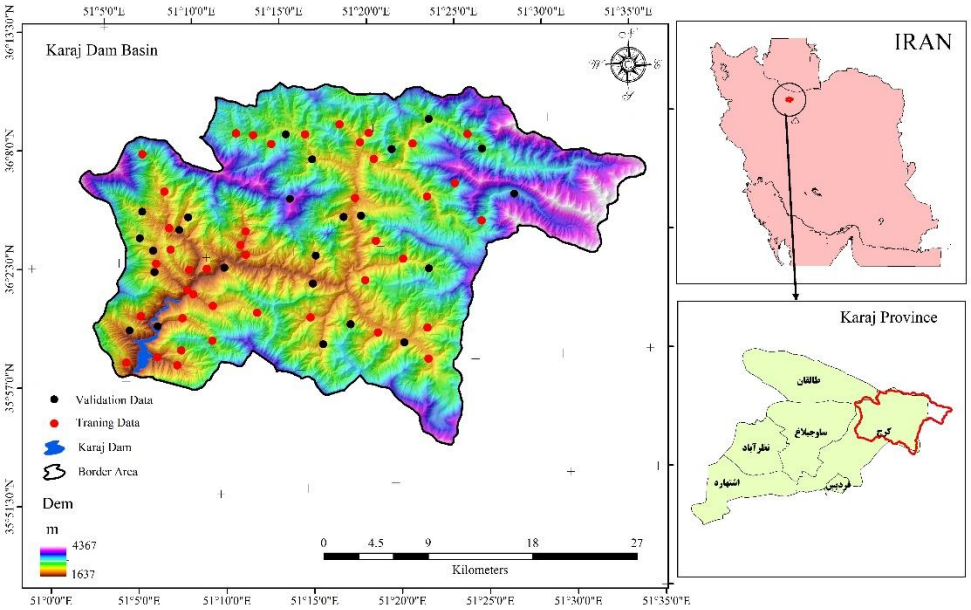
حوزه آبخیز سد کرج در استان البرز و بین مختصات جغرافیایی (بین ۵۱ درجه و ۵ دقیقه تا ۵۱ درجه و ۳۵ دقیقه طول شرقی و ۳۵ درجه و ۵۱ دقیقه تا ۳٦ درجه و ۱۳ دقیقه عرض شمالی) واقع شده است (شکل ۱). دامنه ارتفاع از سطح دریا از ٤۳٦۷ متر تا خروجی حوزه با ارتفاع ۱٦۳٦ متر متغیر است. کم‌ترین و بیش‌ترین مقدار بارندگی متوسط سالانه به‌ترتیب ٦۰۸ و ۱۰۹۹ میلی‌متر محاسبه شده است. در شکل ۱ نقاط سیل و تقسیم‌بندی آن‌ها به دو بخش آموزشی و اعتبارسنجی جهت مدل‌سازی و پهنه‌بندی حساسیت سیل‌خیزی نشان داده شده است.

در تحقیق حاضر جهت تعیین مناطق دارای پتانسیل سیل‌خیزی از ۱۲ عامل شامل ارتفاع، شیب، جهت شیب، انحنای توپوگرافی، فاصله از آبراهه، لیتولوژی، بارندگی، کاربری اراضی، شاخص قدرت آبراهه (SPI)، شاخص موقعیت توپوگرافیک (TPI)، موقعیت شیب نسبی (RSP) و شاخص رطوبت توپوگرافیک (TWI) استفاده شده است. در این تحقیق از مجموع ۸۲ نقطه سیل، به صورت تصادفی ۲۵ نقطه سیل (۳۰ درصد) به عنوان داده‌های اعتبارسنجی و ۵۷ نقطه سیل (۷۰ درصد) به عنوان داده‌های آزمون (آموزش) طبقه‌بندی شدند (شکل ۱) [۵، ۱۱، ۱۵ و ۱٦]. شکل ۲ نمودار جریانی را نشان می‌دهد.

شکل ۳ عوامل مورد استفاده در تحقیق حاضر را نشان می‌دهد. به منظور تهیه نقشه عوامل توپوگرافیک از نقشه مدل رقومی ارتفاع حوزه آبخیز سد کرج با قدرت تفکیک مکانی ۱۲/۵ متر استفاده شد. نقشه DEM از سایت https://vertex.daacc.asf.alaska.edu دانلود شد. ارتفاع بر روی میزان بارندگی، نوع و تراکم پوشش گیاهی تاثیرگذار است [۵]. و از این طریق به صورت غیرمستقیم بر وقوع سیلاب تاثیرگذار است. عوامل شیب، جهت شیب و انحنای توپوگرافی با استفاده از DEM و به ترتیب از توابع Slope، Aspect و Curvature تهیه شدند. نقشه شیب بیانگر تغییرات ارتفاع در راستای سطح افق می‌باشد و افزایش شیب زمین باعث کاهش نفوذپذیری و افزایش رواناب می‌شود. جهت شیب بر روی میزان دریافت بارندگی، میزان رطوبت و میزان دریافت انرژی خورشید اثرگذار است و بر همین اساس نوع و تراکم پوشش گیاهی را تحت تاثیر قرار می‌دهد [١٤و ۱۷].انحنای توپوگرافی هم‌گرایی و واگرایی جریان را در سطح شیبدار نشان می‌دهد و بر سرعت جریان سطحی تاثیرگذار است. مقادیر مثبت به مناطق محدب (تسریع رواناب)، مقادیر منفی مناطق مقعر (نگه‌داشت آب) و سلول‌های با مقدار صفر

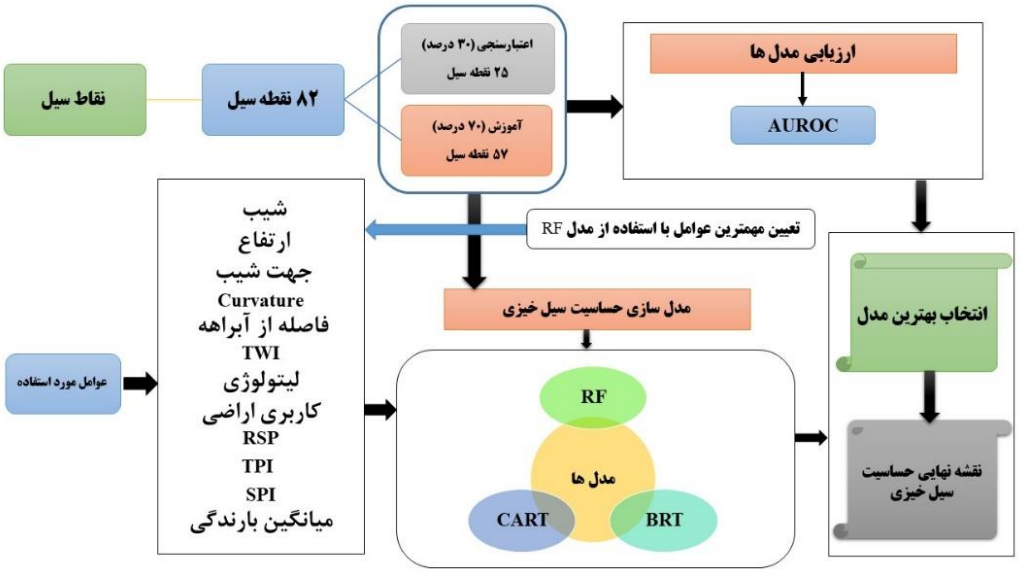
[Downloaded from jwmsei.ir on 2026-01-29]

[DOR: 20.1001.1.20089554.1402.17.61.4.8]



شکل ۱- موقعیت حوزه آبخیز سد کرج در ایران و استان البرز و نقاط سیل آموزشی و اعتبارسنجی

Fig 1. Location of Karaj dam basin in Iran and Alborz province and traning and validation flood points



شکل ۲- نمودار جریانی مراحل تحقیق

Fig 2. Methodological fowchart in this study

به مناطق مسطح تعلق می‌گیرد. شبکه زهکشی به لیتولوژی، شیب و شرایط اقلیمی مربوط می‌شود و نقش اساسی در زمان تمرکز و اوج سیلاب دارد. نقشه آبراهه‌ها از نرم‌افزار SAGA GIS استخراج و لایه فاصله از آبراهه بر اساس تابع Euclidean distance در نرم‌افزار ArcGIS تهیه شد. بارندگی‌های با شدت زیاد و در مدت زمان کم یکی از اصلی‌ترین دلایل سیل در بیشتر مناطق ایران است. نقشه بارندگی با استفاده از اطلاعات ۱۳۸ ایستگاه باران‌سنجی تهیه شده است. ابتدا میانگین بارندگی ۲۲ ساله (۱۳۷٦-۱۳۹۸) برای هر ایستگاه محاسبه شد و سپس با استفاده از روش کریجینگ جهانی در نرم‌افزار ArcGIS پهنه‌بندی بارندگی انجام گرفت. لیتولوژی تعیین‌کننده شدت و ظرفیت نفوذپذیری خاک بوده و رواناب سطحی را نیز تحت تاثیر قرار می‌دهد [٤]. سنگ‌شناسی به عنوان یک عامل مهم در نوع شبکه زهکشی، هیدرولوژی و عمق و عرض آبراهه‌ها در نظر گرفته می‌شود [۳]. لایه کاربری اراضی از اداره کل منابع طبیعی و آبخیزداری استان البرز دریافت شده است و به کمک کارشناسان مربوطه در برخی مناطق به دلیل تغییر کاربری اراضی اصلاحاتی بر روی آن انجام شده است. کاربری اراضی به صورت مستقیم و غیرمستقیم بر روی پارامترهای سیکل هیدرولوژی و پتانسیل

[Downloaded from jwmsei.ir on 2026-01-29]

[DOR: 20.1001.1.20089554.1402.17.61.4.8]

سیل تأثیرگذار است [۱۵]. لایه‌های شاخص موقعیت توپوگرافیک، شاخص قدرت آبراهه، شاخص رطوبت توپوگرافیک و موقعیت شیب نسبی با استفاده از نرم‌افزار SAGA GIS تهیه شدند. شاخص TPI تفاوت بین پیکسل مرکزی و میانگین پیکسل‌های اطراف را نشان می‌دهد. (رابطه ۱).

$$TPI = Z_0 - \frac{\sum_{1-n} Z_n}{n} \quad (۱)$$

که در آن Z_0: ارتفاع پیکسل مرکزی؛ Z_n: ارتفاع شبکه اطراف پیکسل مرکزی؛ n: تعداد کل نقاط مورد استفاده در ارزیابی.

مقادیر مثبت نشان‌دهنده خطالرس‌ها و مقادیر منفی نشان‌دهنده دره‌ها و چاله‌ها و مقادیر صفر نشان‌دهنده مناطق مسطح می‌باشد. شاخص SPI شاخص معرف توان فرسایش جریان آب می‌باشد (رابطه ۲) و افزایش آن باعث افزایش سرعت جریان و افزایش گل‌آلودگی می‌شود.

$$SPI = A_s * \tan\beta \quad (۲)$$

A_s: مساحت حوزه آبخیز و β: گرادیان شیب بر حسب درجه است. شاخص رطوبت توپوگرافی به صورت زیر تعریف می‌گردد (رابطه ۳):

$$TWI = IN\left(\frac{A_s}{\tan\beta}\right) \quad (۳)$$

که در آن As سطح حوزه آبخیز و β شیب به درجه می‌باشد. این شاخص همبستگی زیادی با عمق خاک، درصد سیلت، مواد عالی و میزان فسفر دارد. شاخص TWI تأثیر توپوگرافی بر تولید روان‌آب را به‌صورت کمی نشان می‌دهد و توزیع مکانی رطوبت خاک را برآورد می‌کند. اگر مقدار آن بالا باشد باعث کاهش نفوذ آب می‌شود [۱۳] و در نتیجه آن مناطق دارای رطوبت زیاد در هنگام بارندگی می‌توانند باعث تشدید سیلاب شوند [۱۷]. شاخص RSP نشان‌دهنده موقعیت شیب هر سلول بین کف دره تا خطالراس است (رابطه ٤).

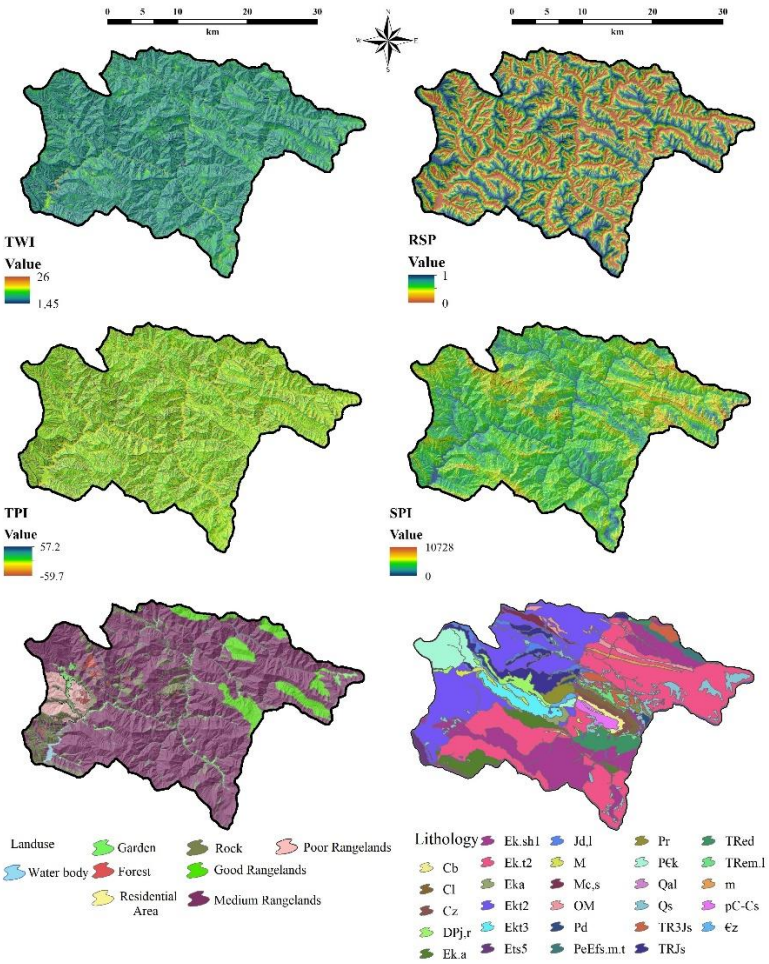
$$RSP = \left(\frac{z(s)-z(s)_v}{z(s)_r-z(s)_v}\right) \times 100 + 0.5 \quad (٤)$$

که در آن $Z(s)$: ارتفاع؛ $z(s)_v$: ارتفاع دره؛ $z(s)_r$: ارتفاع خطالراس.

مدل‌های مورد استفاده و ارزیابی کارایی مدل‌ها

مدل جنگل تصادفی (Random Forest)

جنگل تصادفی یک روش جدید و قدرتمند است که پیشرفت‌های قابل ملاحظه‌ای را در تکنولوژی داده‌کاوی ارائه داده است، با این



شکل ۳- عوامل مورد استفاده جهت تعیین مناطق دارای حساسیت سیل

Fig 3. Factors used to determine flood susceptibility areas

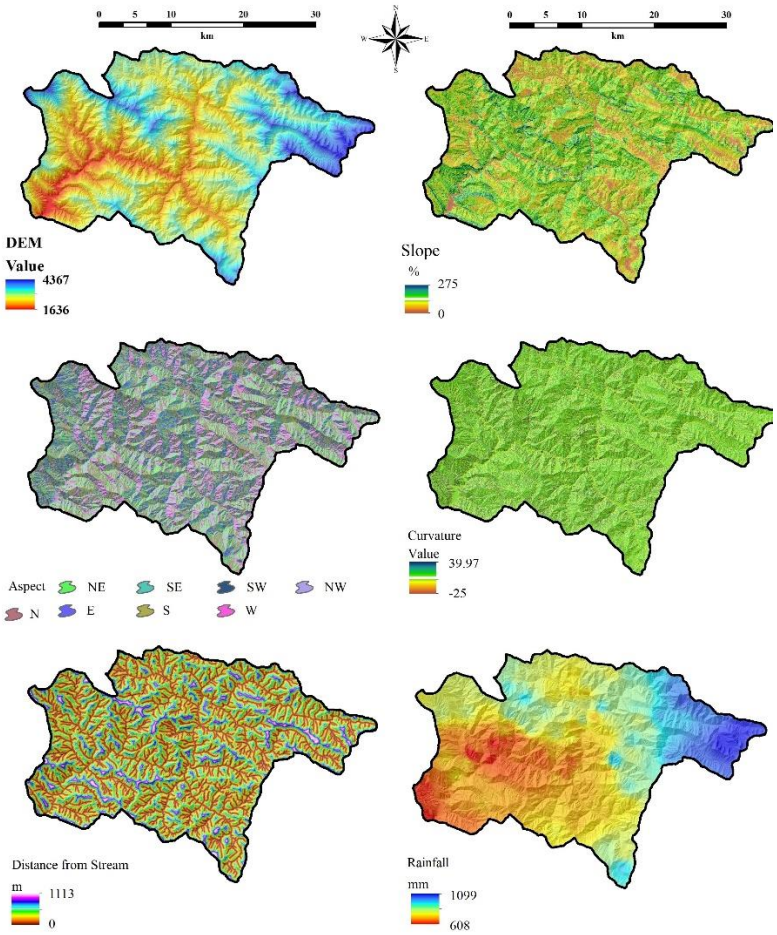
[Downloaded from jwmsei.ir on 2026-01-29]

[DOR: 20.1001.1.20089554.1402.17.61.4.8]

وجود در مطالعات اکولوژیک نسبتاً ناشناخته است. رویکرد جنگل تصادفی مبتنی بر روش‌های جدید ترکیب اطلاعات است که در آن تعداد زیادی درخت تصمیم ایجاد گردیده، سپس تمام درختان با هم برای پیش‌بینی ترکیب می‌گردند [۱]. زمانی که متغیرهای پیشگو و هدف مشخص گردیدند، جنگل‌های تصادفی با رویاندن یک درخت تصمیم شبیه CART شروع می‌شوند، این درخت از چندین لحاظ با درخت استاندارد CART متفاوت است، پارامترهای کلیدی برای مدل جنگل تصادفی، تعداد درختان و تعداد متغیرهای پیشگو هستند. براساس این دو پارامتر، درخت تصمیم به بزرگترین اندازه ممکن رویانده و بدون هرس رها می‌شود [۳] . مزایای جنگل‌های تصادفی را می‌توان به صورت زیر خلاصه نمود: ۱) صحت؛ مدل‌های جنگل تصادفی اغلب نسبت به یک تک درخت دارای صحت قابل توجه بیشتری می‌باشند. صحت آن، اغلب قابل رقابت با بهترین روش‌های جایگزین می‌باشد. ۲) مقاومت به بیش برازش؛ رویاندن تعداد زیاد درخت خطای بیش برازش را ایجاد نمی‌کند. ۳) انتخاب پیشگوها از بین تعداد زیادی متغیرهای کاندید تحلیگر، نیازی به انتخاب متغیر یا تقلیل داده ندارد؛ جنگل تصادفی به‌طور خودکار، در نهایت بهترین متغیرهای پیشگو را شناسایی می‌کند. ٤) سرعت؛ درختان با سرعت زیاد رویانده می‌شوند، چرا که تعداد کمی متغیر در هر زمان مورد استفاده قرار می‌گیرد. ۵) توانایی مدل‌سازی اثرات متقابل پیچیده میان متغیرهای پیش‌بینی کننده ٦- انعطاف‌پذیری برای انجام انواع مختلفی از تجزیه و تحلیل داده‌های آماری، از جمله رگرسیون و طبقه‌بندی [۱۵].

درخت طبقه‌بندی و رگرسیون (Classification and Regression Tree)

تجزیه و تحلیل داده‌های پیچیده اکولوژیکی نیازمند روش‌های تحلیلی قوی و انعطاف‌پذیری است که بتواند روابط غیر خطی، اثرات متقابل و داده‌های از دست رفته را کنترل کند. علاوه بر این درک و ارائه نتایج توسط این روش‌ها باید ساده و براحتی قابل تفسیر باشد. درختان طبقه‌بندی و رگرسیون برای تجزیه و تحلیل داده‌های پیچیده اکولوژیکی ایده‌آل و مناسب هستند. CART یکی از روش‌های رگرسیون ناپارامتری است که درخت تصمیم‌گیری را بر اساس الگوریتم تقسیم‌بندی دوتایی بازگشتی رشد می‌دهد. درخت به



ادامه شکل ۳

Fig 3 Continue

[Downloaded from jwmsei.ir on 2026-01-29]

[DOR: 20.1001.1.20089554.1402.17.61.4.8]

صورت گرافیکی با گره ریشه (حاوی داده‌های تقسیم نشده) در بالا وگره‌های شاخه (داده‌های تقسیم شده) و گره‌های برگ (تقسیم‌های نهایی) در پایین، نشان داده می‌شود که این موضوع به درک تفسیر روابط کمک می‌کند [٦]. ساختار سلسله مراتبی درخت رگرسیون و طبقه‌بندی بدان معنی است که پاسخ به یک متغیر به مقادیر در گره‌های بالاتر نیز بستگی دارد، بنابراین اثر متقابل بین متغیرها به طور خودکار مدلسازی می‌شود [٢٠]. اولین اقدام تصمیم‌گیری در مورد انتخاب بهترین متغیر تقسیم کننده است که بتواند متغیر هدف را به بهترین نحو تقسیم کند. هنگامی که بهترین متغیر تقسیم کننده انتخاب شد، داده‌ها به دو گروه تقسیم می‌شوند (تقسیم دوگانه). جنبه تقسیم بازگشتی اشاره به این واقعیت دارد که هر گره به نوبه خود به دو گره فرزند تقسیم می‌شود و آن گره‌ها نیز پس از آن تقسیم می‌شوند و به همین ترتیب این روند ادامه پیدا می‌کند. این تقسیم‌بندی تا زمانی ادامه می‌یابد که گروه‌ها همگن شده یا تعداد مشاهدات در درون گره‌ها کمتر از مقدار حداقل تعریف شده توسط کاربر باشد. هر گره توسط تعداد مشاهدات، مقدار میانگین متغیر پاسخ آن مشاهدات و میزان متغیر پیشگو تقسیم کننده آن، مشخص می‌شود. یک موضوع مهم در درخت رگرسیون و طبقه‌بندی، تصمیم‌گیری در مورد زمان توقف تقسیم‌ها است، چنان‌چه تقسیم شدن‌ها متوقف نشود، الگوریتم درخت در نهایت همه اطلاعات را تقسیم کرده و منجر به بیش برازش یا فرایادگیری می‌شود. چنین رویکردی یک درخت بزرگ و پیچیده ایجاد می‌کند که این درخت را نمی‌توان برای پیش بینی موارد دیگر استفاده کرد. از سوی دیگر، درخت کوچک، بعید است که ویژگی‌های مهم را در بر گیرد [٢٠]. مزایای استفاده از درختان عبارتند از: ١- استفاده از طیف گسترده‌ای از انواع متغیرها از جمله طبقه‌ای، رتبه‌ای و داده‌های بقا ٢- سهولت و مقاوم بودن ٣- سهولت تفسیر ٤- توانایی کنترل داده‌های از دست رفته. بنابراین درختان می‌توانند جایگزنی مناسب برای بسیاری از تکنیک‌های آماری سنتی از جمله رگرسیون چندگانه، رگرسیون لجستیک و مدل لگاریتمی- خطی باشند. هر چند ساخت درختان رگرسیون و طبقه‌بندی آسان بوده و تفسیر آن‌ها به راحتی نیز امکان‌پذیر است. اما یک ضعف اساسی آن این است که مدل برازش یافته واریانس بالایی دارد. یک تغییر اندک در داده‌ها اغلب منجر به تغییرات زیاد در شکل درخت برازش یافته می‌گردد. این امر تفسیر درخت رگرسیون و طبقه‌بندی را با مشکل مواجه می‌نماید. برای غلبه بر این مشکل راه حل‌های متعددی وجود دارد که تمامی این روش‌ها مستلزم برازش درختان متفاوت زیاد به داده‌ها و میانگین‌گیری پیش بینی‌ها از درختان است. این رویکردها به روش ترکیبی معروف می‌باشند که جنگل تصادفی و درخت رگرسیون تقویت شده دو روش بسیار متداول این رویکرد می‌باشند [٦].

مدل درخت رگرسیون تقویت شده (Boosted Regression Trees)

BRT ترکیبی از تکنیک‌های آماری و یادگیری ماشین است. این تکنیک یکی از تکنیک‌های مختلفی است که با استفاده از ترکیب مدل‌های متعدد به بهبود عملکرد تک مدل کمک می‌کند. BRT از ترکیب دو الگوریتم استفاده می‌کند: مدل رگرسیونی CART و ساخت و ترکیب مجموعه‌ای از مدل‌ها (Boosting). Boosting الگوریتم کارآمد و ساده‌ای برای افزایش دقت مدل است و بر این اساس کار می‌کند که ساخت، ترکیب و میانگین‌گیری تعداد زیادی مدل، بهتر و دقیق‌تر از ایجاد یک مدل است [١]. Boosting بر بزرگ‌ترین ضعف تک درخت تصمیم که دارای برازش نسبتاً ضعیف است غلبه می‌نماید. در BRT تنها اولین درخت از کل داده‌های آموزشی ساخته می‌شود، درخت‌های بعدی بر روی داده‌های باقیمانده درخت قبل از خود، رشد داده می‌شوند. درخت‌ها بر روی همه داده‌ها ساخته نمی‌شوند و فقط از تعدادی از داده‌ها استفاده می‌کنند. ایده اساسی در این روش ترکیب مجموعه‌ای از مدل‌های پیش‌بینی کننده ضعیف (خطای پیش‌بینی بالا) برای رسیدن به پیش‌بینی قوی (خطای پیش‌بینی کم) است. در هر مرحله، هر مجموعه داده طبقه‌بندی می‌شود و این طبقه‌بندی به عنوان وزن برای برازش درخت بعدی استفاده می‌گردد. موارد نادرست طبقه‌بندی‌شده، وزن بیشتری از مواردی که به درستی طبقه‌بندی شده‌اند دریافت می‌کنند و در نتیجه مواردی که به سختی طبقه‌بندی می‌شوند وزن بسیار زیادی دریافت می‌کنند و در نتیجه شانس بیشتری برای طبقه‌بندی درست بدست می‌آورند. مزایای BRT به شرح زیر است: ١) پرداختن به انواع مختلف متغیرهای پیشگو ٢) اصلاح داده‌های از دست رفته ٣) بدون نیاز به تبدیل داده‌ها یا حذف نقاط دورافتاده ٤) برازش روابط پیچیده غیرخطی ٥) کنترل خودکار اثر متقابل بین متغیرها [١]. مدل رگرسیون تقویت شده در نرم‌افزار R و با استفاده از بسته gbm اجرا شده است است.

منحنی ROC نموداری است که در آن نسبت پیکسل‌هایی که به اشتباه توسط مدل به عنوان پهنه سیل شناسایی شده‌اند و در واقعیت فاقد واقعه سیل بوده‌اند روی محور افقی (مثبت کاذب یا Specificity) و در برابر مقدار مکمل آن یعنی نسبت پیکسل‌هایی که به درستی توسط مدل به عنوان پهنه سیل شناسایی شده‌اند (مثبت صحیح یا Sensitivity) روی محور عمودی ترسیم می‌گردد. مساحت زیر این منحنی AUC نام داشته و مدلی که بیش‌ترین مقدار AUC را دارا باشد دارای عملکرد نسبی بالاتری است. AUC برابر با ٠/٥ معادل مدل خنثی (نتایج ناشی از شانس) بوده و هر چه این مقدار به عدد ١ نزدیک‌تر شود کارایی مدل افزایش می‌یابد [١٧].

نتایج

نتایج در بخش ارتفاع نشان داد که با کاهش ارتفاع احتمال وقوع سیل بیشتر شده و از ارتفاع ٢٢٥٠ متری به بالا کاهش احتمال وقوع سیل اتفاق افتاده است. تغییرات انحنای توپوگرافی نشان داد که میزان حساسیت سیل در مناطق با انحنای محدب افزایش پیدا کرده است. تغییرات شیب نشان داد که تا شیب ٤٢ درصد در سطح حوزه به دلیل وجود مراتع خوب و باغات و در مواردی جنگل‌های دست کاشت

[Downloaded from jwmsei.ir on 2026-01-29]

[DOR: 20.1001.1.20089554.1402.17.61.4.8]

در این شیب‌ها حساسیت سیل کاهش پیدا کرده است. لازم به ذکر است که این مناطق بیشتر محل عبور سیل هستند (شیب‌های پایین). در شیب‌های بیشتر با افزایش شیب افزایش حساسیت سیل اتفاق افتاده است که دلیل آن افزایش میزان رواناب و کاهش نفوذپذیری در شیب‌های بالاتر است. همچنین شیب‌های بالاتر، بیشتر حالت صخره‌ای دارند که باعث افزایش میزان رواناب و سرعت آن و در نتیجه افزایش میزان حساسیت سیل می‌شوند. نقشه جهت شیب نشان داد که در جهات شمال غربی، شمال و مناطق مسطح بیشترین حساسیت سیل اتفاق افتاده است. نقشه فاصله از آبراهه نشان داد که با افزایش فاصله از آبراهه به دلیل اینکه بیشترین میزان تجمع جریان در آبراهه اتفاق می‌افتد، میزان حساسیت سیل کاهش پیدا کرده است. اصولا افزایش میزان دبی در آبراهه و خروج آب از آبراهه‌ها باعث ایجاد سیل می‌شود. در بحث لیتولوژی کاهش نفوذپذیری آب سطحی و وجود رخنمون سنگی در واحدهایی نظیر سنگ آهک دولومیتی؛ توف‌های سبز ضخیم لایه، کنگلومرا و آهک؛ سیلت سنگ و شیل خاکستری تیره و ماسه سنگ حساسیت سیل را افزایش داده است. بر اساس نقشه کاربری اراضی بیشترین میزان حساسیت سیل در کاربری‌های مراتع فقیر و در مناطق دارای رخنمون سنگی بوده است. با افزایش مقدار موقعیت شیب نسبی در مناطق نزدیک به یال حساسیت سیل افزایش پیدا می‌کند. نقشه شاخص موقعیت توپوگرافیک نشان داد که مناطق دارای مقادیر زیر صفر شامل مناطق دره‌ای و چاله‌ها (افزایش رطوبت و کاهش نفوذپذیری) به سمت مناطق مسطع میزان حساسیت سیل افزایش پیدا کرده است. نتایج نشان داد که افزایش میزان حساسیت سیل در اثر افزایش شاخص رطوبت توپوگرافی رخ داده است. نقشه میزان بارندگی و حساسیت سیل نشان داد که با افزایش میزان بارندگی میزان حساسیت سیل کاهش پیدا کرده است.

جدول ۱ نتایج وزن متغیرها حاصل از اجرای مدل RF را جهت اولویت‌بندی عوامل موثر بر حساسیت سیلخیزی نشان می‌دهد. بر اساس مدل RF عوامل ارتفاع و فاصله از آبراهه مهم‌ترین عوامل در تعیین مناطق دارای حساسیت سیل بوده‌اند و عوامل انحنای توپوگرافی و جهت شیب کمترین اثر را داشته‌اند. عوامل ارتفاع، فاصله از آبراهه، بارندگی و کاربری اراضی در مطالعات زیادی به عنوان عوامل موثر بر سیل معرفی شده‌اند [۱۷].

شکل ٤ نقشه نهایی مناطق دارای حساسیت سیل را در حوزه آبخیز سد کرج نشان می‌دهد. بر اساس مدل RF به عنوان مدل برتر ۳۲/۷ درصد از پهنه حوزه مورد مطالعه دارای پتانسیل حساسیت سیل متوسط، زیاد و خیلی زیاد می‌باشد. همچنین جدول ۲ مساحت و درصد مساحت هر یک از طبقات حساسیت سیل را نشان می‌دهد. طبقه حساسیت سیل خیلی کم بیشترین مساحت معادل ٤١ درصد سطح حوزه و طبقه خیلی زیاد کمترین میزان مساحت معادل ٥ درصد سطح حوزه را بر اساس مدل برتر شامل می‌شوند. بر اساس نقشه نهایی، مناطق اطراف آبراهه‌های اصلی و مناطق پایین دست حوزه در محدوده سد کرج دارای بیشترین میزان حساسیت به سیل هستند.

جدول ۱- وزن متغیرهای مورد استفاده برای تعیین مهم‌ترین عوامل بر اساس مدل RF

Table 1. Weight of variables used to determine the important factors based on the RF model

Weight	Factors	عوامل
19.95	DEM	ارتفاع
18.12	Distance from stream	فاصله از آبراهه
17.72	Rainfall	بارندگی
16.25	Landuse	کاربری اراضی
15.23	RSP	RSP
14.88	Lithology	لیتولوژی
14.09	Slope	شیب
11.12	TWI	TWI
9.82	SPI	SPI
7.55	TPI	TPI
6.43	Curvature	انحنای توپوگرافی
4.91	Slope Aspect	جهت

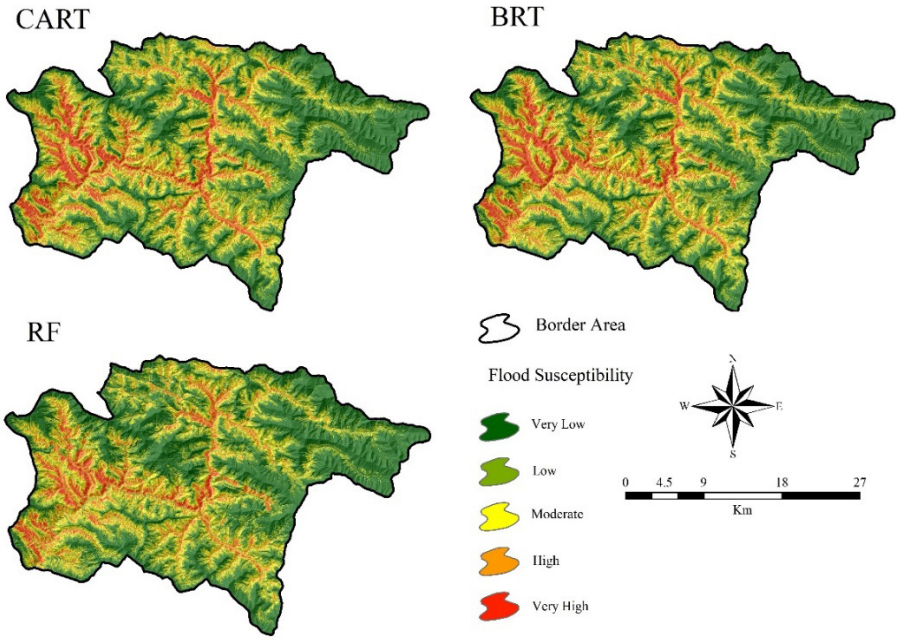
[Downloaded from jwmsei.ir on 2026-01-29]

[DOR: 20.1001.1.20089554.1402.17.61.4.8]

جدول ۲- مساحت و درصد مساحت طبقات حساسیت سیل در حوزه آبخیز سد کرج

Table 2. Area and percentage of flood susceptibility classes in Karaj Dam Basin

BRT		CART		RF		
درصد مساحت	مساحت (هکتار)	درصد مساحت	مساحت (هکتار)	درصد مساحت	مساحت (هکتار)	طبقه حساسیت
Area (%)	Area (h)	Area (%)	Area (h)	Area (%)	Area (h)	Susceptibility
35.8	30419	36.2	30794	41	34867	خیلی کم Very low
25.9	21989	25.1	21322	26.3	22370	کم Low
19	16200	18.6	15829	16.9	14375	متوسط Moderate
12.7	10801	13.3	11334	10.8	9167	زیاد High
6.6	5656	6.8	5786	5	4285	خیلی زیاد Very high



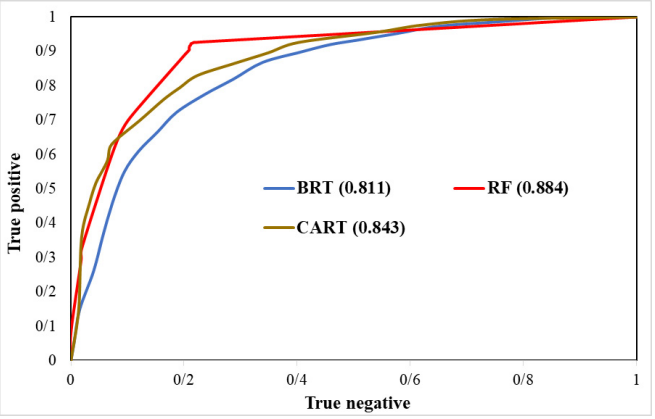
شکل ٤- نقشه حساسیت سیل در حوزه آبخیز سد کرج

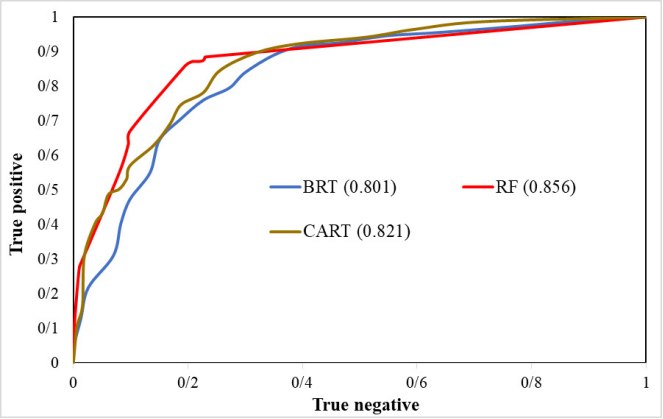
Fig 4. Flood susceptibility map in Karaj Dam Basin

شکل ۵ نتایج حاصل از قدرت پیش‌بینی مدل بر اساس مقادیر ROC را در دو بخش آموزش و اعتبارسنجی نشان می‌دهد. بر اساس این شکل میزان دقت پیش‌بینی مدل هم در بخش آموزش (۰/۸۸٤) و هم اعتبارسنجی (۰/۸۵٦) در مدل RF به عنوان مدل برتر خیلی خوب بوده است. لذا با توجه به مقادیر ROC می‌توان به نتایج پیش‌بینی مدل اعتماد کرد. مدل‌های CART و BRT در رتبه‌های بعدی با اختلاف اندکی قرار دارند. نتایج نشان داد که مدل RF دارای دقت بالاتری (اختلاف اندک) در تعیین مناطق حساس به وقوع سیل می‌باشد.

[Downloaded from jwmsei.ir on 2026-01-29]

[DOR: 20.1001.1.20089554.1402.17.61.4.8]





شکل ۵- نتایج مقادیر ROC در بخش بالا: آموزش و پایین: اعتبارسنجی مدل

Fig 5. The ROC value for three flood susceptibility models using the validation and traning dataset

بحث و نتیجه‌گیری

سیل از جمله حوادث طبیعی است که فراوانی وقوع آن در دو دهه اخیر روند افزایشی داشته و خسارات ناشی از آن همه ساله بخش وسیعی از کشور را تحت تاثیر قرار داده است. عوامل زیادی در یک حوزه آبخیز بر تولید رواناب در سطح حوزه اثرگذار هستند که هر کدام از این عوامل پتانسیل متفاوتی در تغییر و نوع جریان و همچنین شدت و بزرگی رخداد سیلاب دارند. لذا تعیین مهم‌ترین و اثرگذارترین فاکتورهای موثر بر سیلاب جهت کنترل و کاهش خسارات ضروری به نظر می‌رسد. در تحقیق حاضر عوامل ارتفاع، فاصله از آبراهه، بارندگی، کاربری اراضی و RSP مهم‌ترین عوامل موثر بر حساسیت سیل در حوزه سد کرج بودند که از بین این عوامل دو مورد کاربری اراضی و فاصله از آبراهه قابل مدیریت هستند. به عبارت دیگر با جلوگیری از تغییر کاربری اراضی و همچنین رعایت حد بستر و حریم رودخانه (پهنه سیلاب با دوره بازگشت‌های مختلف) می‌توان تا حدود زیادی حساسیت سیل را کاهش داد و خسارات را به حداقل رساند. عوامل ارتفاع، فاصله از آبراهه، بارندگی و کاربری اراضی در مطالعات زیادی به عنوان عوامل موثر بر سیل معرفی شده‌اند [۱۷]. نتایج نشان داد که در مناطق نزدیک به سد کرج و محدوده اطراف آبراهه‌های اصلی و رودخانه کرج افزایش احتمال وقوع سیل و خسارات افزایش پیدا کرده است. از دلایل این امر تغییر کاربری‌های صورت گرفته در سال‌های اخیر و افزایش اراضی کشاورزی و باغی و مسکونی است که در معرض سیل بودن را افزایش داده و لذا سبب افزایش حساسیت سیل و خسارات شده‌اند که باید این موارد به صورت ویژه مورد بررسی و بازبینی قرار بگیرند. به عبارت دیگر کاهش فاصله با حریم اصلی رودخانه و در معرض قرار گرفتن املاک، اراضی و افراد باعث افزایش خسارات در مواقع سیلابی شده‌اند. لذا می‌توان از طریق اعمال مدیریت و فرهنگ‌سازی در این زمینه خسارات سیل را کاهش داد. از دلایل کاهش حساسیت سیل در ارتفاعات بالاتر حوزه آبخیز سد کرج می‌توان به افزایش بارش به صورت برف (ذوب تدریجی و نفوذپذیری و عدم تولید رواناب) اشاره کرد. در ارتفاعات بالاتر پوشش گیاهی بیشتر حالت بالشتکی به خود می‌گیرد و از تشکیل رواناب و وقوع سیل جلوگیری می‌کند [۹]. نتایج این بخش با تحقیقات [۲ و ۳]. همخوانی دارد. مناطق محدب در سطح حوزه باعث افزایش و تسریع در تمرکز جریان می‌شوند که تشدید سیل را به همراه دارد و واگرایی جریان در این مناطق زیاد می‌باشد. به عبارت دیگر مناطق محدب محل شروع رواناب و تشکیل سیل هستند. یکی از دلایل افزایش حساسیت سیل در مناطق شیبدار، طول شیب زیاد است که تجمع جریان در طول شیب‌های بیشتر افزایش پیدا کرده و همچنین سرعت جریان افزایش پیدا می‌کند. مجددی و همکاران [۱۲] و آوند و همکاران [۳] نقش شیب را نسبت به حساسیت سیل کاهشی می‌دانند که دلیل آن را وجود پوشش گیاهی جنگلی در مناطق شیبدار شمال کشور معرفی کردند [۳ و ۱۲]. از دلایل افزایش حساسیت سیل در جهات شمال و شمال غرب حوزه سد کرج، افزایش نگهداشت رطوبت سطحی و افزایش رواناب و کاهش نفوذپذیری در این جهات می‌باشد. لازم به ذکر است در جهات مذکور در بخش‌هایی که میزان پوشش گیاهی افزایش پیدا کرده است، حساسیت سیل کاهش پیدا کرده است. از دلایل افزایش میزان حساسیت سیل در مراتع فقیر می‌توان به فرسایش سطحی و چرای بی‌رویه دام‌ها و در نتیجه کاهش میزان پوشش گیاهی و افزایش میزان رواناب اشاره کرد. در مناطق دارای رخنمون سنگی نیز به دلیل افزایش میزان رواناب در این سطوح میزان حساسیت سیل افزایش پیدا کرده است. با افزایش مقدار موقعیت شیب نسبی در مناطق نزدیک به یال که این مناطق بیشتر حالت صخره سنگ دارند، به دلیل عدم نفوذپذیری و هدایت جریان به سمت مناطق دره به مرور میزان حساسیت سیل افزایش پیدا می‌کند. افزایش میزان حساسیت سیل در اثر افزایش شاخص رطوبت توپوگرافی به دلیل کاهش میزان نفوذپذیری و افزایش رواناب و در نتیجه تشدید میزان حساسیت سیل رخ داده است [۱۸]. از دلایل کاهش حساسیت در اثر بارندگی می‌توان به افزایش میزان پوشش گیاهی در اثر افزایش

[Downloaded from jwmsei.ir on 2026-01-29]

[DOR: 20.1001.1.20089554.1402.17.61.4.8]

میزان بارندگی اشاره کرد. عامل پوشش گیاهی یکی از عوامل اصلی نفوذ آب و کاهش میزان رواناب و در نتیجه کاهش میزان حساسیت سیل می‌باشد. تهیه نقشه حساسیت به وقوع سیل، پیش زمینه‌ای در شناخت عوامل موثر بر وقوع سیل، رخداد آن، مدیریت خطر و روش‌های پیشگیری از خطر و ریسک سیل است. در این پژوهش نقشه حساسیت سیل حوزه آبخیز سد کرج با درنظر گرفتن عوامل مختلف و تأثیرگذار و با مدل‌های CART, BRT و RF تهیه شد. نتایج ROC نشان داد که دقت مدل در برآوردکردن مناطق دارای حساسیت سیل در مرحله آموزش و در مرحله اعتبارسنجی در هر سه مدل خیلی خوب بوده است و مدل RF نسبت به سایر مدل‌ها با اختلاف اندکی کارایی بالاتری دارد که با نتایج [۳، ۵ و ۲۱]. مطابقت دارد. بر اساس نقشه پهنه‌بندی حساسیت سیل با استفاده از مدل RF به عنوان مدل برتر ۱٤۳۷٦ هکتار در طبقه متوسط و ۱۳٤۵۱ هکتار در طبقه زیاد و خیلی زیاد حساسیت سیل قرار دارند که تمهیدات ویژه مدیریتی در این مناطق می‌تواند حساسیت و خسارات سیل را کاهش دهد. در مناطق نزدیک به آبراهه‌ها به دلیل تعداد زیادی ویلا و باغ ویلا، رستوران و ... حساسیت سیل افزایش پیدا کرده است و در محدوده سد کرج نیز رخنمون‌های سنگی بزرگ با شیب بسیار زیاد و با کم‌ترین میزان نفوذپذیری قابل مشاهده است که باعث تشدید سیلاب در این نواحی شده است. با توجه به نتایج مدل‌ها ترکیب سیستم اطلاعات جغرافیایی با سنجش از دور و مدل‌های یادگیری ماشینی به منظور تعیین مناطق دارای حساسیت سیل مخصوصا در کشورهای در حال توسعه که دسترسی به اطلاعات و داده‌های سیل با محدودیت زمان و هزینه مواجه هستند پیشنهاد می‌گردد.

منابع

1. Abedi, R., Costache, R., Moghadam, H., Pham, Q. 2021. Flash-flood susceptibility mapping based on XGBoost, random forest and boosted regression trees. Geocarto International. doi.org/10.1080/10106049.2021.1920636.

2. Arabameri, A., Rezaei, K., Cerdà, A., Conoscenti, C., & Kalantari, Z. 2019. A comparison of statistical methods and multi-criteria decision making to map flood hazard susceptibility in Northern Iran. Science of the Total Environment, 660(10), 443-458.

3. Avand, M., Moradi, H.R., Ramazanzadeh, M. 2020. Flood Susceptibility Mapping Using Random Forest Machine Learning and Generalized Bayesian Linear Model. Environment and water Enginerring. 6(1), 83-95. [In Persian].

4. Bomers, A., van der Meulen, B., Schielen, R. M. J., & Hulscher, S. J. M. H. 2019. Historic flood reconstruction with the use of an Artificial Neural Network. Water Resources Research, 55(11), 9673-9688.

5. Chen, W., Hong, H., Li, S. 2019. Flood susceptibility modelling using novel hybrid approach of reduced-error pruning trees with bagging and random subspace ensembles. Journal of Hydrology. 575, pp. 864–873.

6. 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.

7. Costache, R. 2019. Flash-Flood Potential assessment in the upper and middle sector of Prahova river catchment Romania. A comparative approach between four hybrid models," Science of the Total Environment. 659, pp. 1115–1134.

8. Darabi, H., Shahedi, K., Mardian, M. 2016. Flood susceptibility and probability mapping using frequency ratio method in Pol-Doab Shazand watershed. J. Eng. Watershed Manage. 8(1), 68-79. (In Persian)

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

10. Koehorst, B.A.N., Kjekstad, O., Patel, D., Lubkowski, Z., Knoeff, J.G., Akkerman, G.J. 2005. Work package 6 Determination of Socio-Economic Impact of Natural Disasters. Assessing socioeconomic Impact in Europe, 173p.

11. Mohammadi, A., Valizaeh, K.h., Karimzadeh, S., Shahabi, H., Al. Ansari, N. 2020. Flood Detection and Susceptibility Mapping Using Sentinel-1 Time Series, Alternating Decision Trees, and Bag-ADTree Models. Hindawi, Complexity, V. 4271376, https://doi.org/10.1155/2020/4271376.

12. Mojaddadi, H., Habibnezhad, M., Shahedi, K., Pradhan, B. 2020. The Efficiency of an Ensemble Frequency Ratio-Support Vector Machine Model in the Detection of Flood-Prone Areas of the Kalat Basin. Iranian Journal of Eco Hydrology. 7(1): 77-95. (In Persian).

13. Mokarram, M., Roshan, G., Negahban, S. 2015. Landform classification using topography position index case study: salt dome of Korsia-Darab plain, Iran. Modeling Earth Systems and Environment. 1: 40.

14. Muniraj, K., Jesudhas, C.J., Chinnasamy, A. 2019. Delineating the Groundwater Potential Zone in Tirunelveli Taluk, South Tamil Nadu, India, Using Remote Sensing, Geographical Information System GIS and Analytic Hierarchy Process AHP Techniques. Proceedings of the National Academy of Sciences, India Section A:

[Downloaded from jwmsei.ir on 2026-01-29]

[DOR: 20.1001.1.20089554.1402.17.61.4.8]

Physical Sciences. Doi.org/10.1007/s40010-019-00608-5.

15. 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, 125275.

16. Rahmati, O., Darabi, H., Haghighi, A., Stefandis, S., Kornejady, A., Asadi Nalivan, O., Tien Bui, D. 2019. Urban Flood Hazard Modeling Using Self-Organizing Map Neural Network. Water, 11, 2370, doi: 10.3390/w11112370.

17. Shahabi, H. 2021. Flood susceptibility mapping in northern regions of Iran using advanced data mining algorithms (Case study: Haraz watershed). Journal of Regional Planning. 11(41): 167-184. (In Persian).

18. Tehrany, M., Shabani, F., Neamah Jebur, M., Hong, H., Chen, W., Xie, X. 2017. GIS-based spatial prediction of flood prone areas using standalone frequency ratio, logistic regression, weight of evidence and their ensemble techniques. Geomatics, Natural Hazards and Risk. 82, 1538-1561.

19. Toda, L.L., Yokingco, J.C.E., Paringit, E.C., Lasco, R.D. 2017. A LiDAR-based flood modelling approach for mapping rice cultivation areas in Apalit, Pampanga. Applied Geography. 80: 34-47.

20. Wang, Y., Fang, Z., Hong, H., Costache, R., Tang, X. 2021. Flood susceptibility mapping by integrating frequency ratio and index of entropy with multilayer perceptron and classification and regression tree. Journal of Environmental Management. 289, 112499. doi.org/10.1016/j.jenvman.2021.112449.

21. Zhao, 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.

[Downloaded from jwmsei.ir on 2026-01-29]

[DOR: 20.1001.1.20089554.1402.17.61.4.8]

نشریه علمی
علوم و مهندسی آبخیزداری ایران
Iran-Watershed Management
Science & Engineering

Vol. 17, No. 61, Summer 2023 سال هفدهم- شماره ٦١- تابستان ١٤٠٢

Comparison of Machine Learning Models in Flood Susceptibility Zoning in Karaj Dam Basin

M. Teimouri[1], F. Vakili Tajareh[2], M. Mozayyan[3] and M. Ramezani[4]

Received: 02-05-2022 Accepted: 02-08-2022

Abstract

The present research aims to determine areas with flood susceptibility using CART, RF and BRT models. Twelve factors affecting flood susceptibility including altitude (DEM), slope, aspect, distance from stream, lithology, rainfall, land use, SPI, TPI, TWI, curvature and RSP were selected. Out of 82 flood points, 70 percent to 30 percent were randomly classified as training and validation data. Also, random forest method was used to determine the most important parameters. The ROC curve was also used to validation the model. According to the random forest model, DEM, distance from stream, rainfall, land use and RSP were the most important factors affecting the susceptibility and probability of floods, respectively. According to the ROC chart, the accuracy of the RF model as a superior model has been very good in both training (0.884) and validation (0.856). According to the final flood susceptibility map, 32.7 percent of the study area has a medium to high flood susceptibility. The results showed due to the high accuracy of the spatial distribution map of flood susceptibility can be promising for decision makers, local managers and policy-makers to reduce flood damages.

Keywords: Random forest, Karaj dam basin, Flood susceptibility, Machine learning.

[Downloaded from jwmsei.ir on 2026-01-29]

[DOR: 20.1001.1.20089554.1402.17.61.4.8]

1. Corresponding Author and Assistant Professor Department of Nature Engineering Shirvan Faculty of Agriculture, University of Bojnord, Iran
2. Ph.D Student in Watershed Science and Engineering, Faculty of Natural Resources, University of Tehran.
3. Assistant Professor, Department of Rangeland and Watershed Management, Faculty of Natural Resources, Behbahan Khatam Alanbia University of Technology
4. Bachelor's degree in rangeland and watershed management of Yazd University