



تعیین پتانسیل سیل با استفاده از مدل‌های یادگیری ماشین CART، GLM و GAM (مطالعه موردی: حوضه کشکان)

حسین یوسفی^۱، حجت الله یونسی^{۲*}، داود داودی مقدم^۳، آزاده ارشیا^۴، زهرا شمسی^۵

تاریخ ارسال: ۱۴۰۰/۰۴/۱۲

تاریخ پذیرش: ۱۴۰۰/۰۷/۰۷

مقاله پژوهشی

چکیده

سیل پدیده‌ای است که موجب آسیب‌های زیست‌محیطی و اقتصادی-اجتماعی بسیاری می‌شود. هدف از این پژوهش، ارزیابی کارایی مدل‌های یادگیری ماشین CART، GLM و GAM در شناسایی مناطق حساس به خطر سیلاب در حوضه کشکان است. استان لرستان و به‌ویژه حوضه کشکان شامل: سلسله، دلفان، دوره، خرم‌آباد، پلدختر و کوه‌دشت، سیل‌خیز است و دفعات بسیاری دچار خسارات ناشی از سیل شده است و در فروردین ۱۳۹۸، بزرگ‌ترین سیل ۲۰۰ سال اخیر را تجربه کرده است؛ در همین راستا از عوامل مختلف شامل: ارتفاع، جهت شیب، انحنای زمین، درصد شیب، فاصله از رودخانه، تراکم زهکشی، خاک، سنگ‌شناسی، کاربری اراضی و شاخص رطوبت توپوگرافی استفاده شد. نقشه رقومی تمام عوامل نام‌برده در نرم‌افزار ArcGIS 10.5 و در قالب پایگاه داده تهیه شد. موقعیت ۱۲۳ واقعه سیل ثبت‌شده در سال‌های اخیر در این حوضه، جمع‌آوری و به‌صورت تصادفی در دو دسته آموزش مدل (۸۶ واقعه) و اعتبارسنجی مدل (۳۷ واقعه) در مدل‌سازی‌ها استفاده شد. با استفاده از مدل‌های یادگیری ماشین و عوامل مؤثر محیطی، نقشه‌های پیش‌بینی پتانسیل سیل تهیه شدند و سپس با استفاده از روش‌های منحنی مشخصه AUC و شاخص TSS اعتبارسنجی شدند. نتایج حاصل از اعتبارسنجی مدل‌ها نشان داد که مدل یادگیری ماشین CART با $AUC=0/91$ و شاخص $TSS=0/88$ دقیق‌ترین مدل در پیش‌بینی پتانسیل خطر سیل بوده و پس از آن مدل GAM با $AUC=0/87$ و شاخص $TSS=0/84$ و مدل GLM با $AUC=0/83$ و شاخص $TSS=0/88$ قرار دارند. دقت ۰/۹۱ مدل CART نشان‌دهنده دقت عالی این مدل برای حوضه کشکان است. این مدل، مساحت بیشتری از حوضه را تحت شرایط پتانسیل بالا و متوسط خطر سیل‌گیری نشان می‌دهد که اغلب مناطق غربی و همچنین مناطق مرکزی حوضه (کوه‌دشت، خرم‌آباد و پلدختر) را شامل می‌شوند که دقیقاً بخش‌هایی از همین مناطق در سیل بزرگ سال ۹۸ هم زیر آب رفتند و لازم است در اولویت اول برنامه‌ریزی و مدیریت ریسک سیل قرار گیرند.

واژه‌های کلیدی: شاخص رطوبت توپوگرافی، نقشه سیل، AUC، ROC، TSS

^۱ دانشیار دانشکده علوم و فنون نوین، دانشگاه تهران Email: hosseinyousefi@ut.ac.ir

^۲ * استادیار گروه مهندسی آب، دانشکده کشاورزی و منابع طبیعی، دانشگاه لرستان Email: yonesi.h@lu.ac.ir

^۳ دکتری آبخیزداری، دانشکده کشاورزی و منابع طبیعی، دانشگاه لرستان Email: d.davoudi.m@gmail.com

^۴ دانشجوی دکتری سازه‌های آبی، دانشکده کشاورزی و منابع طبیعی، دانشگاه لرستان Email: azadeharshia69@gmail.com

^۵ دانشجوی دکتری سازه‌های آبی، دانشکده کشاورزی و منابع طبیعی، دانشگاه لرستان Email: azm.shams@gmail.com



مقدمه

را با استفاده از مدل ترکیبی فرآیند تحلیل شبکه و منطق فازی بررسی کردند. بر اساس مدل تحلیل شبکه معیار شیب و سنگ‌شناسی بیشترین ضریب تأثیر را داشتند. کنعانی سادات و همکاران به ارزیابی حساسیت سیل با روش‌های تصمیم‌گیری چندمعیاره و GIS برای تعدادی حوضه در کردستان پرداختند. طبق ارزیابی نقاط سیل، مدل ANP ارزش زیادی در مقایسه با AHP نشان داد. در این مطالعه معیار TWI و تراکم آبراهه بیشترین وزن را داشتند (Kanani-Sadat et al, 2019). بوی و همکاران پهنه‌بندی سیل را با استفاده از الگوریتم تکاملی BuiNovel و روش -های رگرسیون لجستیک و نروفازی در مازندران انجام دادند (Bui et al, 2018). آزاد طلب و همکاران (۱۳۹۹) در مطالعه‌ای سیل خیزی شهر سنندج را با استفاده از روش‌های WOE و EBF مورد بررسی قرار دادند و نتایج نشان داد که در منطقه مورد مطالعه، مدل EBF دارای عملکرد بهتری است. تهرانی و همکاران، خطر سیل را با در نظر گرفتن متغیرهایی از جمله شاخص رطوبت توپوگرافی، شاخص توان آبراهه، شیب، کاربری اراضی، زمین‌شناسی، خاک و انحنا حوضه با استفاده از روش‌های یادگیری درخت تصمیم ماشین بردار پشتیبان ارزیابی کردند. نتیجه ارزیابی صحت مدل‌ها با استفاده از AUC به ترتیب با مقدار ۰/۸۸ و ۰/۸۵ دقت مناسب آن‌ها را نشان داد (Tehrany et al, 2019). فرامرزی و همکاران (۱۳۹۸) به پهنه‌بندی وقوع مخاطره سیلاب در پارک ملی گلستان با مدل‌های جنگل تصادفی، آنتروپی بیشینه و درخت رگرسیون تقویت‌شده پرداختند و دقت این مدل‌ها با استفاده از روش منحنی ROC بررسی شد. نتایج مدل‌ها برای این منطقه، اهمیت زیاد ارتفاع از سطح دریا، فاصله از آبراهه، بارش و فاصله از جاده را در وقوع مخاطره سیل نشان دادند. مختاری و همکاران (۱۳۹۹) به تهیه نقشه خطر وقوع سیلاب در حوضه با استفاده از ANP و GIS پرداختند. پارامترهای مختلف مؤثر در سیل از جمله شیب، لیتولوژی، کاربری اراضی، شاخص قدرت آبراهه، شاخص رطوبت توپوگرافی، ضریب گراویلیوس و تراکم زهکشی در نظر گرفته شدند و در نهایت معیارهای بارش، لیتولوژی، کاربری اراضی و شیب، بیشترین وزن را داشتند. نقشه پهنه‌بندی در ArcGIS تهیه شد و ۵۰ درصد

ذخیره و بهره‌برداری از رواناب‌ها به‌وسیله‌ی عملیات پخش سیلاب هنگام بارش‌های شدید و وسیع، می‌تواند باعث افزایش ذخایر آب زیرزمینی و نیز مهار سیل گردد (برخورداری و همکاران، ۱۳۹۳). پهنه‌بندی مناطق مستعد سیل، اولین و مهم‌ترین گام در جلوگیری از زمین‌لغزش است (آزادی و همکاران، ۱۳۹۹). روش‌های مختلفی از جمله روش‌های فازی، سلسله مراتبی (اسماعیلی و همکاران، ۱۳۹۹)، ارزیابی چندمعیاره فضایی Arianpour and (Jamali, 2015)، فاکتور وزنی (Khosravi et al, 2016)، مدل ترکیبی هوش مصنوعی (Chapi et al, 2017)، مدل -های EBF و WOE (آزادی و همکاران، ۱۳۹۹) و بسیاری دیگر از روش‌ها به دلیل اهمیت تعیین نقشه حساسیت سیل، همواره در سال‌های اخیر مورد توجه قرار گرفته‌اند. عابدینی و فتحی (۱۳۹۴) به پهنه‌بندی خطر وقوع سیلاب با استفاده از فرآیند تحلیل شبکه در حوضه خیاو چای پرداختند و مدل را با استفاده از پارامترهای مختلف مؤثر در سیل اجرا کردند و نتایج نشان داد که بیش از ۱۵ درصد مساحت حوضه در معرض خطر سیلاب با پتانسیل بسیار بالاست. این سطوح اغلب با شیب کمتر از ۳۵ درصد، درصد پوشش گیاهی کم، نواحی پست و نیز حاشیه رودخانه‌ها را تشکیل می‌دهند. حسینی و همکاران (۱۳۹۴) به تعیین پهنه‌های سیل‌گیر رود کشکان با استفاده از مدل هیدرولیکی HEC-RAS پرداختند. در این پژوهش، بازه‌ای به طول ۵ کیلومتر از رودخانه کشکان مورد بررسی قرار گرفت و نتایج نشان داد که از کل مساحت سیل‌گیر مربوط به دوره بازگشت ۱۰۰ سال، ۷۷ درصد آن مستعد سیل -گیری در دوره بازگشت‌های کوتاه‌مدت است و حتی برای دوره بازگشت دو سال هم زمین‌های زراعی اطراف رودخانه در معرض خطر سیل قرار دارند. سیاه‌کمری و زینی وند (۱۳۹۵) در مطالعه‌ای به تهیه نقشه پتانسیل سیل حوضه مادرسو در استان گلستان با روش‌های آماری WOE و SI پرداختند و دقت این روش‌ها را با استفاده از منحنی ROC خیلی خوب ارزیابی کردند. عابدینی و بهشتی جاوید (۱۳۹۵) در پژوهشی وضعیت سیل‌گیری حوضه ليقوان چای

مناطق از کشور در سال‌های اخیر همواره درگیر سیلاب با خسارات زیاد بودند و لازم است این مناطق در اولویت‌های مطالعات و مدیریت قرار گیرند و این نکته اهمیت پژوهش‌هایی همچون پژوهش حاضر را نشان می‌دهد؛ بنابراین مسئله مهم، تهیه نقشه اولویت‌بندی منطقه از نظر پتانسیل سیل‌خیزی است. ضرورت مطالعه‌ای با موضوع پژوهش حاضر در منطقه کشکان این است که حوضه کشکان جزو مناطقی است که همواره درگیر خسارات بسیار و پرتکرار ناشی از سیل شده و نیز در سال ۱۳۹۸ شدیدترین سیلاب را تجربه کرده است؛ بنابراین بررسی‌های مختلف جهت مدیریت سیلاب در این منطقه، همواره مهم و ضروری است.

هدف کلی از مطالعه‌ی حاضر، ارائه‌ی نقشه‌ی پهنه‌بندی سیلاب به منظور کاهش مخاطرات سیل این حوضه‌ی سیل-خیز با استفاده از مدل‌های یادگیری ماشین CART، GLM و GAM و بهره‌گیری از نرم‌افزار ArcGIS 10.5 برای بهبود تصمیم‌گیری و مدیریت سیل در منطقه است. هدف فرعی نیز مقایسه دقت مدل‌های یادگیری ماشین با دیگر مدل‌های به‌کاررفته در این منطقه است.

در مطالعه حاضر، فرض بر این است که دقت مدل‌های یادگیری ماشین بیشتر از مدل‌های دیگر مطالعه‌شده در این منطقه است. آنچه که به‌عنوان نوآوری مطالعه حاضر اهمیت دارد، این است که تا به حال از مدل‌های یادگیری ماشین CART، GLM و GAM برای تهیه نقشه پتانسیل سیل حوضه کشکان استفاده نشده، بنابراین می‌توان نتایج این پژوهش را با مطالعات و روش‌های دیگر در حوضه سیل‌خیز کشکان مقایسه کرد و همچنین دقت نقشه‌هایی حاصل از سه مدل یادشده را بررسی و تحلیل کرد.

سیلی که در فروردین سال ۹۸ اتفاق افتاد، یکی از بی‌سابقه‌ترین سیل‌های ثبت‌شده در ۲۰۰ سال اخیر بود؛ بنابراین، تهیه نقشه‌های پتانسیل خطر سیل برای شناسایی مناطق سیلاب در حوضه کشکان برای بهبود مدیریت و تصمیم‌گیری ضروری است.

منطقه در معرض خطر وقوع سیلاب قرار داشت. آزادی و همکاران (۱۳۹۹) به پهنه‌بندی حساسیت خطر سیل در حوضه رودخانه کشکان با استفاده از دو مدل WOE و EBF پرداختند و برای هر دو مدل فاکتورهای مختلف مؤثر بر سیلاب را در نظر گرفتند و برای اعتبارسنجی از منحنی ROC استفاده کردند. بر اساس مدل EBF، ۴۷ درصد از سطح منطقه و بر اساس مدل WOE، ۵۲ درصد از منطقه در کلاس خطر متوسط تا خیلی زیاد قرار داشت. بیشترین صحت به مدل EBF اختصاص پیدا کرد و در نتیجه این مدل برای این منطقه مناسب‌تر اعلام شد. حسین زاده و همکاران (۱۳۹۹) در مطالعه‌ای به پهنه‌بندی حساسیت سیل با استفاده از نوعی تحلیل ترکیبی از معیارهای هیدروژئومورفیک و فرایند تحلیل سلسله مراتبی در حوضه سنقر استان کرمانشاه پرداختند. بدین منظور از معیارهای ارتفاع، شیب، فاصله از آبراهه، ژئومورفولوژی، بارش، کاربری اراضی، زمین‌شناسی، قدرت جریان و شاخص رطوبت توپوگرافی استفاده شده است. در این منطقه ۲۶ درصد در معرض حساسیت سیل زیاد و ۵۹ درصد در معرض حساسیت سیل کم قرار دارد. سه عامل ارتفاع، شیب و تراکم شبکه زهکشی و بارش تأثیر عمده‌ای بر حساسیت سیلاب این حوضه داشتند. یوسفی و همکاران (۱۴۰۰) در مطالعه‌ای با توجه به ضرورت مطالعات سیلاب در حوضه سیل‌خیز کشکان به تعیین مناطق مستعد سیل با مدل‌های SI، FR و Shannon و ارائه نقشه پهنه‌بندی به منظور کاهش مخاطرات سیل در این حوضه پرداختند. بیشترین صحت برای این منطقه به مدل آنتروپی شانون (۰/۸۲)، خیلی خوب) اختصاص داشت و بعد از آن، مدل نسبت فراوانی و شاخص آماری (۰/۷۸، خوب)، مناسب این منطقه معرفی شدند. نتایج نشان داد که مدل آنتروپی شانون، مساحت بیشتری از حوضه را تحت شرایط پتانسیل بالای خطر سیل-گیری نشان می‌دهد (حدود ۴۰ درصد از مساحت منطقه در طبقه خطر سیل زیاد و خیلی زیاد) که اغلب مناطق غربی و همچنین مناطق مرکزی حوضه را شامل می‌شوند که در کوه‌دشت، خرم‌آباد و پلدختر قرار دارند که لازم است در اولویت اول برنامه‌ریزی و مدیریت ریسک سیل در این حوضه قرار گیرند.

مواد و روش‌ها

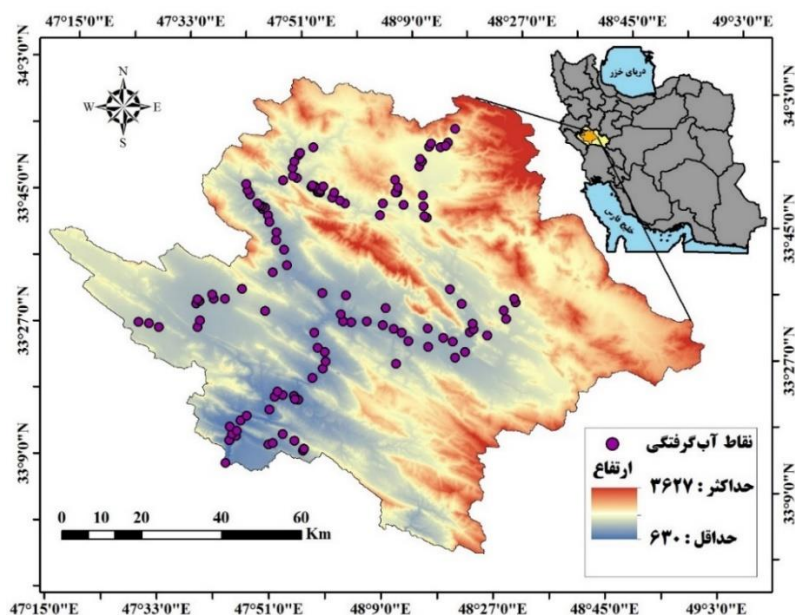
منطقه مورد مطالعه

استان لرستان در غرب ایران واقع شده است و وسعت آن حدود ۲۸۵۵ کیلومترمربع است. رودخانه کشکان با طول ۲۹۰ کیلومتر، سیل خیزترین زیر حوضه کرخه است. از نظر توزیع سازندهای زمین‌ساختی، حدود ۵۰ درصد سطح این حوضه از سازندهای آبرفتی جدید، امیران و مشترک آسماری- شهبازان که دارای پتانسیل بسیار بالای تولید رواناب هستند، تشکیل شده است. وضعیت فیزیوگرافی این حوضه بیانگر این واقعیت است که فاکتورهای مساحت، زمان تمرکز، ضریب فشردگی حوضه، پتانسیل بسیار بالایی در تولید رواناب در مدت‌زمان کوتاه و افزایش سیل‌خیزی

آن دارند. بر اساس پژوهش‌هایی که به برآورد و پیش‌بینی احتمال وقوع سیل و دوره‌های بازگشت آن در این حوضه صورت گرفته، بیانگر احتمال وقوع ۹۹/۹۹ درصد هر سال سیلی به میزان دبی ۶۰۶/۳۲ مترمکعب و هر ۲۵ سال یک‌بار سیلی با دبی ۱۰۹۲/۹ مترمکعب (بدون محاسبه دبی پایه) در نقطه خروجی این حوضه، یعنی شهر پل‌دختر است. سیل فروردین ۱۳۹۸ رودخانه کشکان با آبدهی پیک ۶۵۰۰ مترمکعب در ثانیه و از خسارت‌بارترین سیل‌ها بوده است (آزادی و همکاران، ۱۳۹۹). نمایی از سیل فروردین ۹۸ رودخانه کشکان در شکل (۱) و موقعیت مکانی ۱۲۳ نقطه سیل‌خیز با توجه به اطلاعات شرکت آب منطقه‌ای استان لرستان در شکل (۲) آورده شده است



شکل (۱): نمایی از سیل فروردین ۹۸ رودخانه کشکان



شکل (۲): موقعیت جغرافیایی حوزه آبخیز کشکان و موقعیت نقاط سیل

۲-ج). برای تهیه نقشه شیب از مدل رقومی ارتفاع^۲ (با قدرت تفکیک ۳۰ متر) و بسته الحاقی Slope نرم‌افزار ArcGIS 10.5 استفاده شد. نقشه به‌دست‌آمده به پنج کلاس ۰-۴، ۴-۸، ۸-۱۳، ۱۳-۲۰، و بیشتر از ۲۰ درصد گروه‌بندی شد (شکل ۲-د). نقشه فاصله از رودخانه‌های مجاور بر اساس لایه رقومی شبکه جریان حوزه آبخیز کشکان در نرم‌افزار ArcGIS 10.5 تهیه شد. نقشه حاصل به پنج کلاس (Tehrany et al, 2015)، ۰-۲۷۰، ۲۷۰-۵۸۲، ۵۸۲-۹۶۷، ۹۶۷-۱۵۳۰، و ۱۵۳۰-۱۴۸۴ متر گروه‌بندی شد (شکل ۲-ث). با توجه به نقش تراکم آبراه‌ها در انتقال آب و میزان نفوذ آن در زمین (Mutzner et al, 2016) نقشه تراکم زهکشی نیز بر اساس لایه رقومی شبکه آبراهه در نرم‌افزار ArcGIS 10.5 تهیه شد و سپس در چهار کلاس ۰-۲، ۲-۳، ۳-۴، ۴-۳، و ۳-۱/۳-۱/۳ گروه‌بندی شد (شکل ۲-ج). به دلیل تأثیر خصوصیات خاک روی میزان نفوذپذیری زمین و تولید رواناب، نقشه خاک منطقه نیز در این پژوهش مورد استفاده قرار گرفت (شکل ۲-چ). نقشه سنگ‌شناسی نیز با استفاده از نقشه‌های زمین‌شناسی ۱:۱۰۰۰۰۰ سازمان زمین‌شناسی تهیه و رقومی شد. (شکل ۲-ح). کاربری اراضی یکی از مهم‌ترین عوامل مؤثر در تشکیل رواناب سطحی و پتانسیل سیل حوزه‌های آبخیز

انتخاب عوامل مؤثر در وقوع سیل

عوامل زیادی از جمله کاربری اراضی، شیب زمین، خاک-شناسی و غیره بر وقوع سیل و آب‌گرفتنی در حوزه‌های آبخیز تأثیرگذار هستند (Bates, 2012) با بررسی منابع علمی (Tehrany et al, 2014) پارامترهای ارتفاع، جهت شیب، انحنا، زمین، درصد شیب، فاصله از رودخانه، تراکم زهکشی، خاک، سنگ‌شناسی، کاربری اراضی و شاخص رطوبت توپوگرافی برای پیش‌بینی پتانسیل سیل حوضه کشکان انتخاب شدند. به‌منظور تهیه نقشه‌های ارتفاع زمین و جهت شیب از مدل رقومی ارتفاع و نرم‌افزار ArcGIS 10.5 استفاده شد. در نهایت نقشه ارتفاع حوضه کشکان بر اساس منابع علمی (Tehrany et al, 2013) به پنج کلاس کمتر از ۱۲۵۳ متر، ۱۲۵۳-۱۴۷۸، ۱۴۷۸-۱۶۹۹، ۱۶۹۹-۱۹۵۹، ۱۶۹۹ و بیشتر از ۱۹۵۹ متر گروه‌بندی شد (Tehrany et al, 2014) (شکل ۲-الف). نقشه انحنا^۱ زمین^۱ نیز بر اساس مدل رقومی ارتفاع و با استفاده از بسته Curvature نرم‌افزار ArcGIS 10.5 تهیه شد. این نقشه بر اساس منابع علمی (Tehrany et al, 2014) به سه کلاس انحنا مقعر، انحنا محدب و تخت (بدون انحنا) گروه‌بندی شد (شکل

² - Digital Elevation Model: DEM

¹ - Curvature



(1984) و ایده اصلی این مدل، تقسیم داده‌ها به بخش‌های کوچک‌تر است به طوری که این بخش‌ها حاوی اطلاعات تا حد امکان تفکیک شده باشند (Gordon, 2013). مدل CART روشی برای ایجاد مدل‌های پیش‌بینی از داده‌ها است. این مدل داده‌های ورودی به خود را به صورت بازگشتی پخش می‌کند و قادر به پردازش متغیرهای پیشگو و هدف دسته‌ای است (Loh, 2011).

به منظور استفاده از مدل CART، بایستی ابتدا یک ویژگی هدف را در مجموعه داده‌ها تعیین کنیم. این مدل کار خود را از گره ریشه آغاز می‌کند. داده‌های آموزشی به دو گره فرزند و سپس هر گره فرزند به ترتیب به دو گره نوه تقسیم می‌شوند. در فرآیند رشد درخت، مدل CART در جستجوی سؤالاتی است که هر گره والد را به دو گره فرزند همگن تقسیم نماید. منظور از همگن بودن گره این است که همه رکوردهای موجود در آن متعلق به یک دسته خاص باشند، چون در این صورت آن گره به برگ تبدیل می‌شود. به عبارت دیگر این مدل به دنبال ویژگی‌هایی از مجموعه داده‌ها است که خاصیت جداکنندگی بیشتری دارند. هر چه درخت بیشتر رشد کند، گره‌ها همگن‌تر می‌شوند و اطلاعات بیشتر نمایان می‌شود (Sharma and Kumar, 2016). فرآیند رشد درخت تا هنگام رسیدن به درختی با اندازه حداکثر و تا زمانی که عملیات تقسیم به دلیل کمبود داده‌ها متوقف نشود، ادامه خواهد داشت. پس از ساخت درخت، عملیات هرس کردن درخت با اندازه حداکثر توسط یکی از روش‌های هرس کردن با شروع از برگ‌ها به سمت ریشه انجام می‌شود. مکانیسم CART قصد تولید تنها یک درخت را ندارد، بلکه تلاش می‌کند تا یک توالی از درخت‌های هرس شده‌ی تودرتو را ایجاد کند به طوری که هر یک از آن‌ها گزینه‌هایی هستند که می‌توانند در انتها به عنوان درخت بهینه انتخاب شوند. یک درخت خوب به وسیله ارزیابی کارایی آن بر روی داده‌های آزمایشی مستقل شناسایی می‌شود. در نهایت می‌توان درخت تصمیم ساخته‌شده را بر روی مجموعه داده‌های آزمایشی اعمال

است. نقشه کاربری اراضی بر اساس رقومی سازی و اصلاح نقشه کاربری اراضی اداره کل منابع طبیعی استان لرستان تهیه شد. نقشه کاربری اراضی به کلاس‌های زراعت، مرتع، باغ، جنگل، شهر و فرودگاه تقسیم شد (شکل ۲-خ).

شاخص رطوبت توپوگرافی^۱ (TWI) یکی از شاخص‌های مؤثر در برر سی پتانسیل سیل حوزه‌های آبخیز به شمار می‌رود (Tehrany et al, 2013) شاخص رطوبت توپوگرافی بر اساس رابطه (۱) تعریف می‌شود.

$$TWI = Ln\left(\frac{As}{S}\right) \quad (1)$$

که در آن As سطح ویژه حوزه آبخیز و β شیب زمین (برحسب درجه) می‌باشند (Regmi et al, 2010). نقشه شاخص رطوبت توپوگرافی حوزه آبخیز کشکان بر اساس مدل رقومی ارتفاع در نرم‌افزار ArcGIS 10.5 تهیه شد. در نهایت نقشه حاصل به چهار کلاس ۵-۲، ۶-۵، ۷-۶ و ۷-۱۸ تقسیم شد (Tehrany et al, 2013) (شکل ۲-د).

بررسی آمار موقعیت سیل در دهه‌های اخیر

به منظور بررسی حساسیت خطر سیلاب، تجزیه و تحلیل سیل‌های ثبت شده ضروری است (Bate, 2012). گزارش‌های شرکت آب منطقه‌ای استان لرستان حاکی از این است که حوزه آبخیز کشکان در دهه‌های اخیر به شدت تحت تأثیر سیل‌های خطرناک بوده است. ۱۲۳ نقطه که از جمله نقاط سیل این حوزه هستند در این پژوهش مورد استفاده قرار گرفت که در هر مدل سازی به صورت تصادفی به یک گروه متشکل از ۸۶ واقعه سیل (۷۰ درصد) برای آموزش^۲ و یک گروه متشکل از ۳۷ واقعه سیل (۳۰ درصد) برای اعتبار-سنجی^۳ تقسیم شدند. موقعیت واقعه‌های سیل ثبت شده در شکل ۲ مشخص می‌باشد.

مدل درخت‌های طبقه‌بندی و رگرسیون (CART)

مدل درخت‌های طبقه‌بندی و رگرسیون توسط بریمن و همکاران، سال ۱۹۸۴ معرفی شد (Breiman et al,

³ - Validation

¹ - Topographic Wetness Index: TWI

² - Training

خود نیز از مدل‌های خطی توسعه یافته است (Brenning, 2009). در این مدل برخلاف مدل رگرسیون خطی اجازه داده می‌شود که داده‌ها شکل منحنی پاسخ را تعیین کنند. این تکنیک مطالعه، روابط غیرخطی بین متغیرهای توضیحی و پاسخ را ممکن می‌سازد در حالی که نسبت به بیشتر مدل‌ها، کمتر در معرض بیش‌برازش^۱ در مدل‌سازی مخاطرات طبیعی است. هم‌چنین، مدل GAM برخلاف اکثر تکنیک‌های یادگیری ماشین به راحتی قابل تفسیر است (Goetz et al, 2011). پایه اصلی این مدل بر مبنای جایگزینی عملکرد خطی هر یک از متغیرها است، همان‌طور که در مدل GLM با یک تابع تجربی به کار برده می‌شود تا داده‌ها، بهترین عملکرد را از خود نشان دهند.

درواقع مدل GAM ترکیب توابع صاف کننده خطی و غیرخطی را به صورت یک روش افزایشی امکان‌پذیر می‌کند (Davoudi Moghaddam et al, 2019).

در پژوهش حاضر، این مدل با استفاده از بسته GAM در نرم‌افزار R روی داده‌های ورودی اجرا شد و در نهایت نقشه پتانسیل وقوع سیل حوضه کشکان با استفاده از این مدل نیز تهیه شد.

اعتبارسنجی نقشه‌های پیش‌بینی پتانسیل سیل

به منظور اعتبارسنجی نقشه‌های پیش‌بینی پتانسیل سیل حوضه کشکان، مجموعه‌ای متمایز از داده‌های موقعیت سیل که در مدل‌سازی‌ها استفاده نشده بودند به کار گرفته شدند. در این مرحله ۳۷ واقعه سیل (۳۷ نقطه: ۳۰ درصد) مرحله اعتبارسنجی در نرم‌افزار R با نقشه‌های پیش‌بینی پتانسیل سیل روی هم‌گذاری و مورد تجزیه و تحلیل قرار گرفتند. سپس بر اساس منابع علمی (Tehrany et al, 2014 and Tehrany et al, 2013) از روش منحنی ROC^۲ برای ارزیابی و اعتبارسنجی دقت نقشه‌های پیش‌بینی پتانسیل سیل حوضه کشکان استفاده شد. مساحت زیر این منحنی^۳ (AUC) میزان دقت نقشه‌های پیش‌بینی سیل را به صورت کمی نشان می‌دهد. چنانچه سطح زیر منحنی بین ۰/۵ تا ۰/۶ باشد، نشان‌دهنده تصادفی بودن نتایج است، اگر سطح زیر منحنی بین ۰/۷ تا ۰/۸ باشد،

نمود. منظور از اعمال کردن مدل، پیش‌بینی مقدار ویژگی دسته برای یک رکورد آزمایشی بر اساس مدل ساخته شده است (Sharma and Kumar, 2016). این روند ارزیابی شامل مقایسه هر یک از مقادیر موجود در ویژگی‌های هر رکورد آزمایشی با قواعد موجود در مدل ساخته شده است. در این پژوهش، با استفاده از مدل CART و نیز به کارگیری بسته rpart در نرم‌افزار R، پتانسیل وقوع سیل حوضه کشکان مورد بررسی قرار گرفت. نقاطی که به منظور آموزش مدل انتخاب شده بودند به عنوان متغیر وابسته و لایه‌های مربوط به فاکتورهای محیطی مؤثر بر وقوع سیل به عنوان متغیرهای مستقل به مدل معرفی شدند و سپس مدل، اجرا و نقشه حساسیت سیل منطقه تهیه شد.

مدل خطی تعمیم یافته (GLM)

مدل خطی تعمیم یافته یکی از رایج‌ترین روش‌های آماری است که برای مدل‌سازی مکانی متغیرهای محیطی به کار گرفته می‌شود. این مدل یک روش پارامتری بوده و بسط مدل‌های خطی است. هدف اصلی این روش، یافتن بهترین مدل برای نشان دادن رابطه بین متغیر وابسته و چندین متغیر مستقل است (Ozdemir and Altural, 2013) در این مدل، فرمول ارائه می‌شود و رابطه بین متغیرهای تبیینی و پاسخ به وسیله پارامتر برآورد شده رگرسیون به اضافه فواصل اطمینان سنجش می‌شود. مدل‌های خطی تعمیم یافته برای مواقعی که مشاهدات به طور نرمال توزیع نیافته‌اند و زمانی که سایر روش‌های مدل رگرسیون مناسب نیستند، ابداع شدند. این مدل در بین روش‌های مدل‌سازی دارای عملکرد خوبی است (Davoudi Moghaddam et al, 2019). در پژوهش حاضر این مدل با استفاده از بسته GLM در نرم‌افزار R روی پارامترهای ورودی اجرا شد و در نهایت نقشه پتانسیل وقوع سیل حوضه کشکان با استفاده از این مدل تهیه شد.

مدل جمعی تعمیم یافته (GAM)

مدل جمعی تعمیم یافته (GAM) مدلی نا پارامتریک است که بسط مدل‌های خطی تعمیم یافته (GLM) است و

³ - Area Under Curve: AUC

¹ - Overfitting

² - Receiver Operating Characteristic: ROC



مثبت ۱ متغیر است که مقدار منفی ۱ حاکی از عملکرد تصادفی مدل، مقدار ۰ حاکی از یک مدل ب و مقدار مثبت ۱ حاکی از مدل عالی است (Davoudi et al, 2020)

بررسی درصد مشارکت فاکتورهای محیطی در سیل
در این مطالعه با استفاده از روش آنالیز اهمیت و دقیق ترین مدل مشخص شده به بررسی مشارکت فاکتورهای محیطی مختلف روی وقوع سیلاب پرداخته شده و درصد مشارکت آن‌ها را اولویت بندی می‌کند. در این روش، به صورت چرخشی هر پارامتر به صورت آگاهانه از گردونه مدل سازی حذف شده و با استفاده از بقیه پارامترهای محیطی مدل ساخته می‌شود. سپس، عملکرد پیش‌بینی مدل ایجاد شده با مدل ساخته شده از تمام فاکتورها مقایسه می‌شود. در نتیجه، توزیع عامل حذف شده قابل آزمون است (داودی مقدم و حقی زاده، ۱۳۹۹). فلوچارت مراحل انجام تحقیق حاضر در شکل (۳) آورده شده است.

نتایج دارای اعتبار خوب، سطح زیر منحنی بین ۰/۸ تا ۰/۹ باشد، نتایج دارای اعتبار بسیار خوب و در صورتی که سطح زیر منحنی بیشتر از ۰/۹ باشد، اعتبار نتایج به دست آمده در سطح عالی قرار دارند (Tehrany et al, 2015). از شاخص TSS^1 نیز برای ارزیابی دقت نقشه‌های خروجی استفاده شد که با استفاده از فرمول (۲) محاسبه می‌شود (Davoudi Moghaddam et al, 2020):

$$TSS = TPR - FPR$$

که در آن TPR^2 نرخ مثبت واقعی و FPR^3 نرخ مثبت کاذب می‌باشند و با استفاده از فرمول‌های ذیل محاسبه می‌شوند:

$$TPR = \frac{TP}{TP + FN}$$
$$FPR = \frac{FP}{FP + TN}$$

که در آن TP^4 و TN^5 تعداد پیکسل‌هایی هستند که به درستی گروه بندی شده‌اند، FP^6 (خطای تیپ یک) و FN^7 (خطای تیپ دو) تعداد پیکسل‌هایی هستند که به اشتباه گروه بندی شده‌اند. مقدار شاخص TSS بین منفی ۱ تا

⁵ - True Negatives: TN

⁶ - False Positives: FP

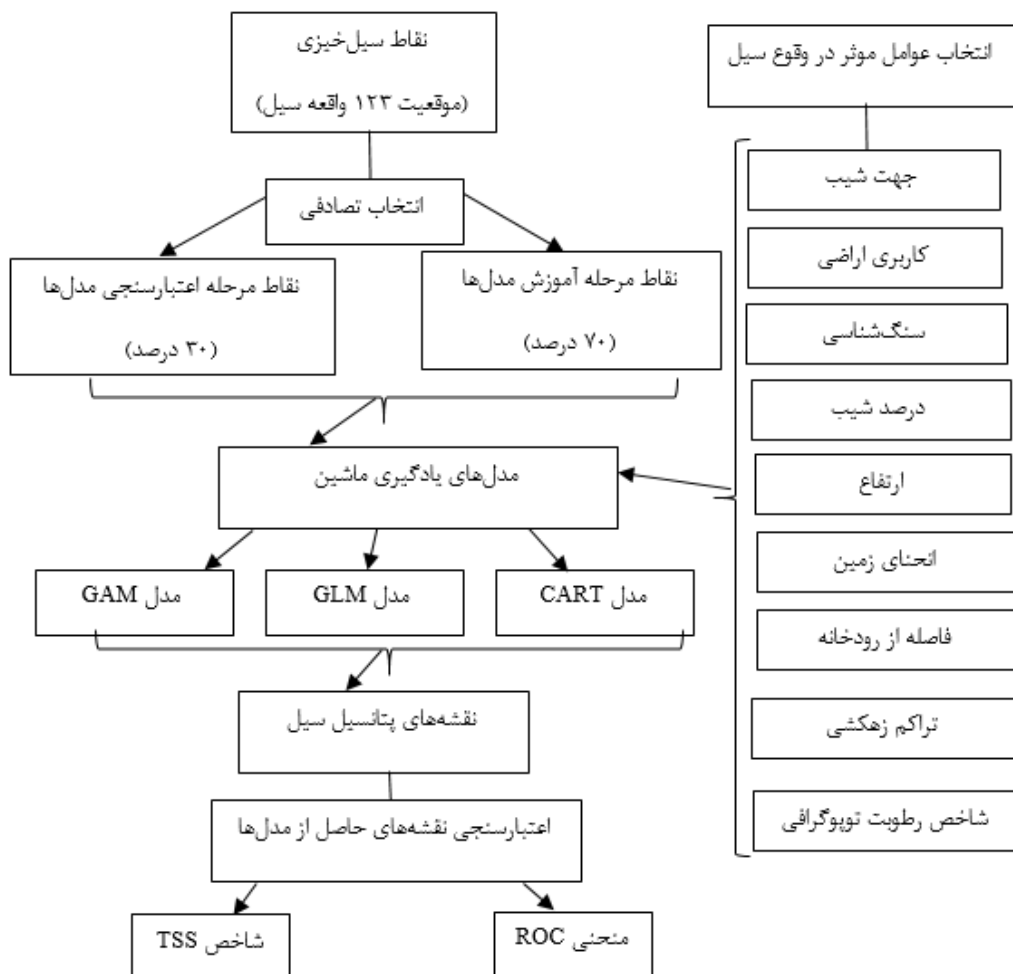
⁷ - False Negatives: FN

¹ - True Skill Statistic: TSS

² - True Positive Rate: TPR

³ - False Positive Rate: FPR

⁴ - True Positives: TP

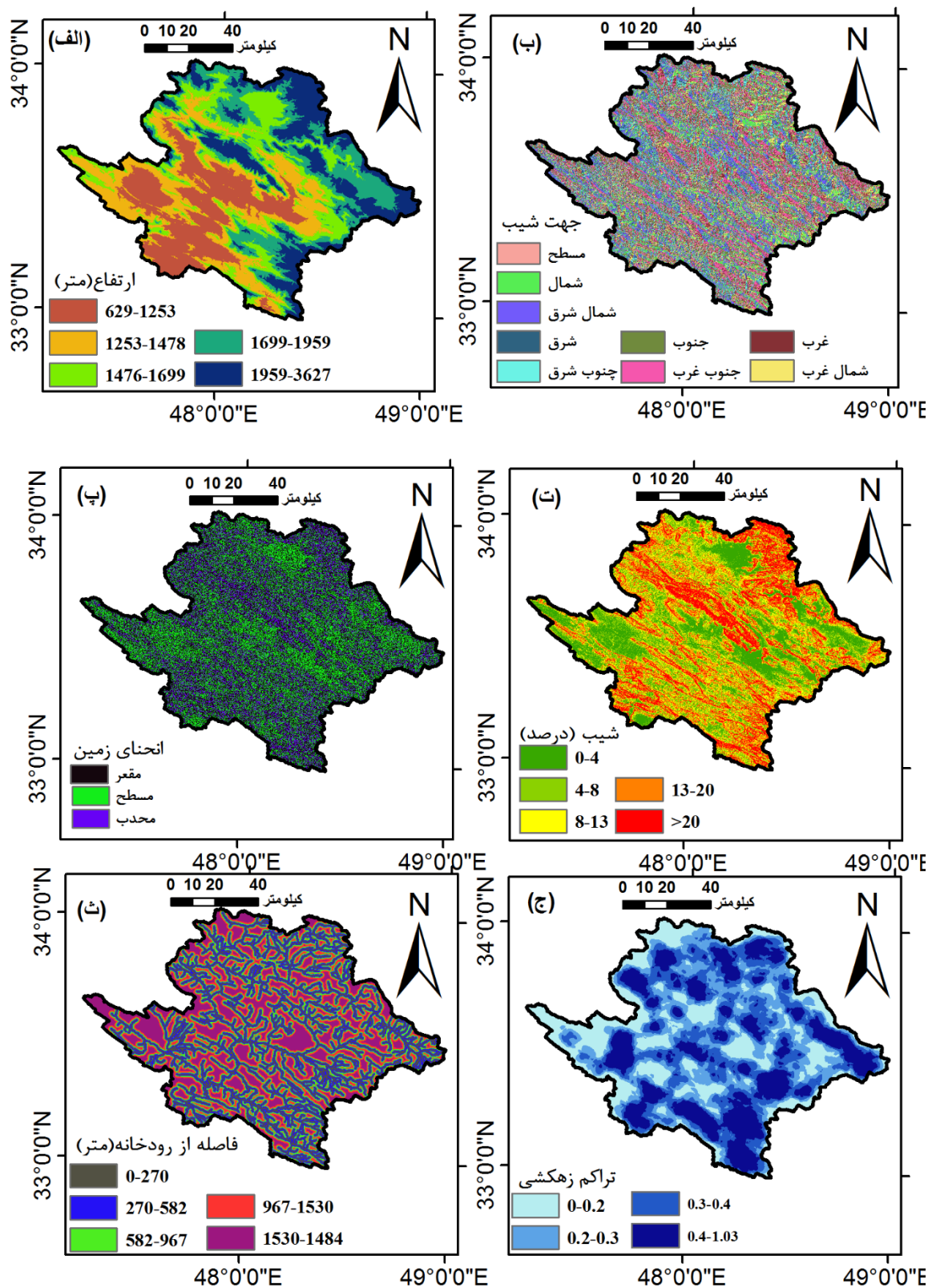


شکل (۳): فلوچارت مراحل مختلف پژوهش

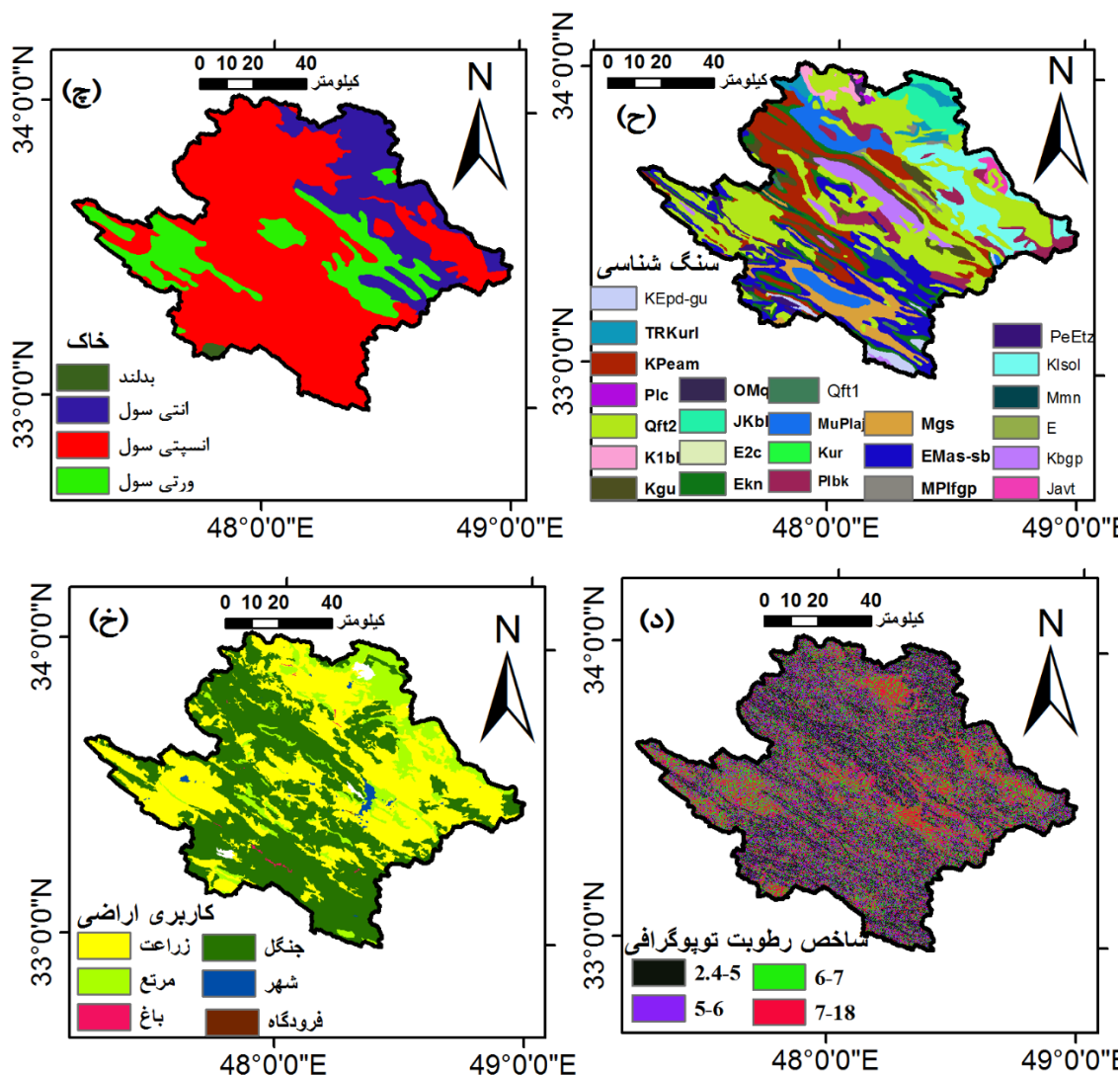
نتایج و بحث

تهیه نقشه‌های عوامل مؤثر محیطی

نقشه عوامل ورودی به مدل‌های CART، GLM و GAM در شکل (۴) آورده شده است.



شکل (۴): نقشه ارتفاع (الف)، جهت شیب (ب)، انحناى زمین (پ)، شیب (ت)، فاصله از رودخانه (ث)، تراکم زهکشی (ج)



ادامه شکل (۴): نقشه خاک (چ) سنگ‌شناسی (ح) کاربری اراضی (خ) و شاخص رطوبت توپوگرافی (د)، حوضه کشکان

از حوضه کشکان را در مدل‌های مختلف به خود اختصاص داده‌اند که در جدول (۱) قابل مشاهده می‌باشند.

ارزیابی دقت نقشه‌های پیش‌بینی پتانسیل سیل

به منظور ارزیابی دقت نقشه‌های پیش‌بینی پتانسیل سیل و آب‌گرفتگی از دو روش منحنی ROC و شاخص TSS و با استفاده از ۳۷ واقعه سیل مربوط به دوره‌های گذشته که در مدل‌سازی‌ها استفاده نشده بودند استفاده شد. دقت پیش‌بینی پتانسیل سیل در مدل‌های مختلف بر اساس روش منحنی ROC و شاخص TSS محاسبه شده که نتایج در جدول ذیل قابل مشاهده می‌باشند. نتایج اعتبارسنجی نشان داد که مدل یادگیری ماشین CART با سطح زیر منحنی ۰/۹۱ و شاخص TSS ۰/۸۸ دقیق‌ترین مدل در

تهیه نقشه‌های پتانسیل خطر سیل و آب‌گرفتگی

تجزیه و تحلیل‌های مقایسه‌ای بین موقعیت وقوع سیل - های گذشته و پارامترهای محیطی مؤثر بر وقوع سیل انجام گرفت و نقشه‌های پیش‌بینی پتانسیل خطر سیل و آب‌گرفتگی حوضه آبخیز کشکان با استفاده از مدل‌های GAM و GLM و CART تهیه شدند و در نهایت بر اساس منابع مختلف (Tehrany et al, 2014) با استفاده از تکنیک Quantile در نرم‌افزار ArcGIS 10.5 به پنج کلاس پتانسیل خطر آب‌گرفتگی خیلی کم، کم، متوسط، زیاد و خیلی زیاد طبقه‌بندی شدند که در شکل (۵) ارائه شده‌اند. نتایج نشان داد که کلاس‌های با پتانسیل خطر خیلی کم، کم، متوسط، زیاد و خیلی زیاد درصد مساحت‌های مختلفی



می‌باشند، گرچه مدل‌های یادگیری ماشین در صد کمتری از مساحت حوضه را تحت خطر سیل متوسط تا خیلی زیاد پیش‌بینی می‌کنند. مدل CART حدود ۲۰ درصد از مساحت حوضه را دارای پتانسیل بالای سیل نشان می‌دهد، اما در مطالعه یوسفی و همکاران (۱۴۰۰)، نتایج نشان داد که مدل آنترابی شانون، مساحت بیشتری از حوضه را تحت شرایط پتانسیل بالای خطر سیل‌گیری نشان می‌دهد (حدود ۴۰ درصد از مساحت منطقه در طبقه خطر سیل زیاد و خیلی زیاد) که البته همان‌طور که اشاره شد این مناطق همانند مطالعه حاضر در قسمت‌های غربی و مرکزی حوضه کشکان قرار دارند.

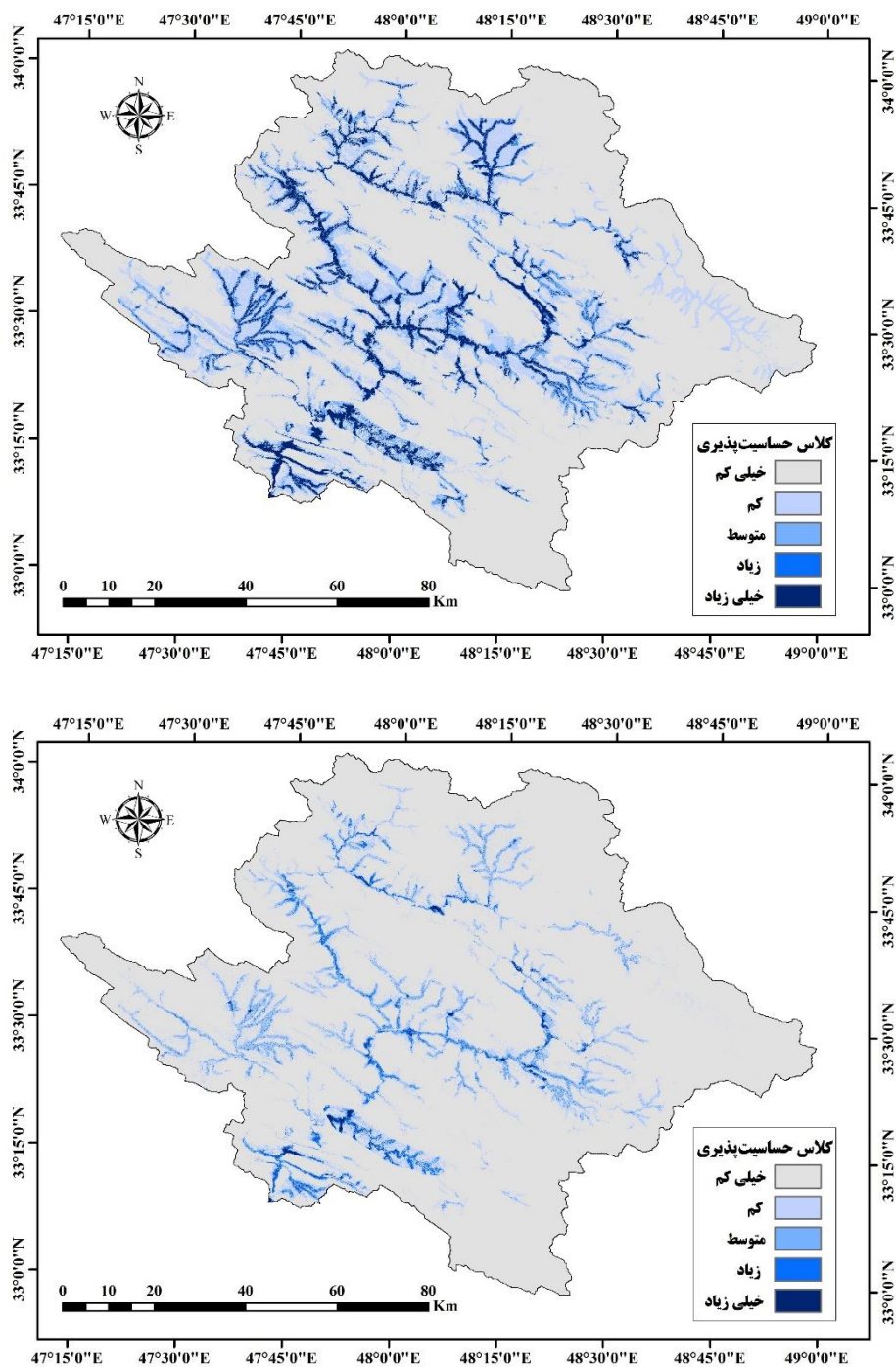
درصد مشارکت فاکتورهای مؤثر در وقوع سیل

نتایج بررسی درصد مشارکت فاکتورهای محیطی در وقوع سیلاب با استفاده از روش آنالیز اهمیت و مدل CART در شکل (۶) ارائه شده است. بر اساس نتایج حاصل، مهم‌ترین پارامترهای محیطی مؤثر بر وقوع سیلاب در حوضه کشکان، فاصله از رودخانه، ارتفاع و شیب هستند که با نتایج یوسفی و همکاران (۱۴۰۰) مطابقت دارد. عوامل سنگ‌شناسی، کاربری اراضی، تراکم زهکشی، شاخص رطوبت توپوگرافی، خاک، جهت شیب و انحنای زمین به ترتیب در رتبه‌های بعد قرار گرفته‌اند. با افزایش فاصله از رودخانه، تعداد وقوع سیلاب‌ها و آب‌گرفتگی‌ها کاهش می‌یابند که نشان می‌دهد مناطق نزدیک به رودخانه‌ها از حساس‌ترین مناطق هنگام وقوع سیلاب هستند که با نتایج (Fernandez and Lutz, 2010) مطابقت دارد. بررسی‌های مربوط به پارامتر ارتفاع نشان می‌دهد که با افزایش ارتفاع، حساسیت نسبت به وقوع سیلاب کاهش یافته است که با نتایج داودی‌مقدم و حقی زاده (۱۳۹۹) نیز هم‌سو است. در ارتباط با پارامتر شیب نیز بررسی‌ها نشان می‌دهد که شیب‌های کمتر پتانسیل وقوع سیلاب بیشتری دارند که علت آن را می‌توان به امکان تمرکز بیش‌تر جریان آب در شیب‌های کم دانست.

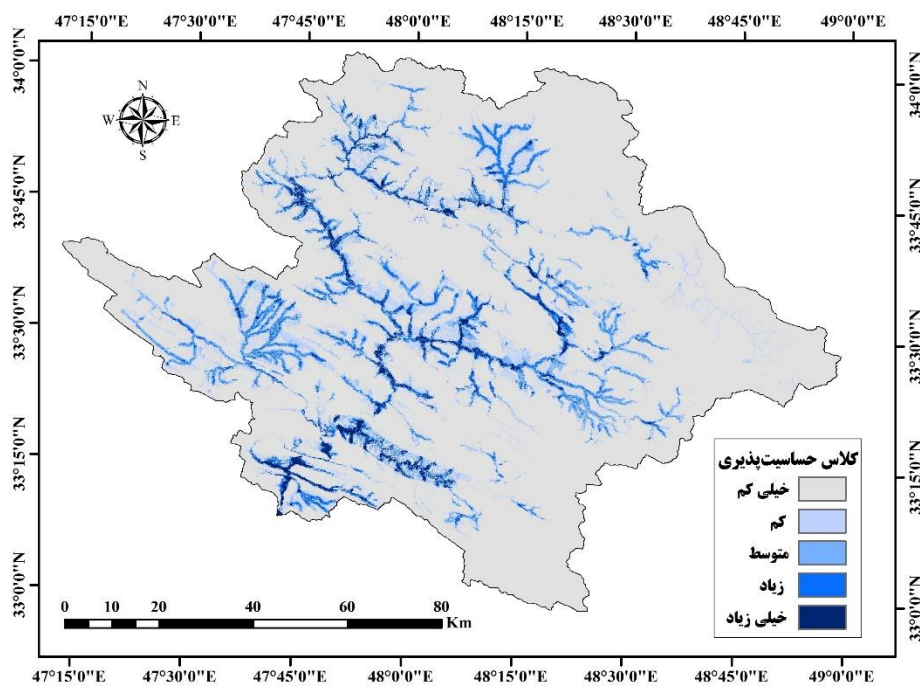
پیش‌بینی پتانسیل خطر سیل بوده و پس‌از آن مدل GAM با سطح زیر منحنی ۰/۸۷ و شاخص TSS ۰/۸۴ و مدل GLM با سطح زیر منحنی ۰/۸۳ و شاخص TSS ۰/۸۰ قرار دارند (جدول ۲). بنابراین، در شناسایی مناطق حساس به سیل و آب‌گرفتگی در حوزه آبخیز کشکان مدل CART دارای دقت عالی ۰/۹۱ و مدل‌های GAM و GLM دارای دقت بسیار خوب ۰/۸۷ و ۰/۸۳ می‌باشند. در بسیاری از مطالعات همچون آزادی و همکاران (۱۳۹۹) و یوسفی و همکاران (۱۴۰۰) فقط از روش منحنی ROC برای بررسی دقت مدل‌ها استفاده شده است اما در پژوهش حاضر از شاخص TSS نیز بهره گرفته شد. دقت مدل CART که ۰/۹۱ است از دقت مدل‌های دیگر که برای همین حوضه استفاده شده است، بیشتر است. در مطالعه یوسفی و همکاران (۱۴۰۰) در همین حوضه، دقت مدل‌های SI، Shannon، FR اعدادی زیر ۸۵ درصد است اما در مطالعه حاضر، مدل GAM با سطح زیر منحنی ۰/۸۷ و مدل CART با سطح زیر منحنی ۰/۹۱ دقت بالاتری دارند.

نتایج پژوهش حاضر با یافته‌های مطالعه محققان (Davoudi Moghaddam et al (2019) مبنی بر دقت عالی و بالاتر مدل CART نسبت به دو مدل GAM و GLM در پیش‌بینی خطر سیلاب هم‌سو است. از مقایسه نتایج تحقیق حاضر با پژوهش‌های انجام شده توسط آزادی و همکاران (۱۳۹۹) و یوسفی و همکاران (۱۴۰۰) که به مدل سازی پتانسیل خطر سیلاب با استفاده از مدل‌های آماری در همین منطقه پرداخته‌اند، مشاهده شد اگرچه استفاده از مدل‌های آماری راحت‌تر است، اما به‌صورت کلی دقت مدل‌های یادگیری ماشین در مدل سازی مکانی مخاطرات طبیعی از مدل‌های آماری بالاتر است که این یافته نیز با نتایج مطالعات (Tehrany et al (2014) و داودی‌مقدم و حقی زاده (۱۳۹۹) هم‌سو است.

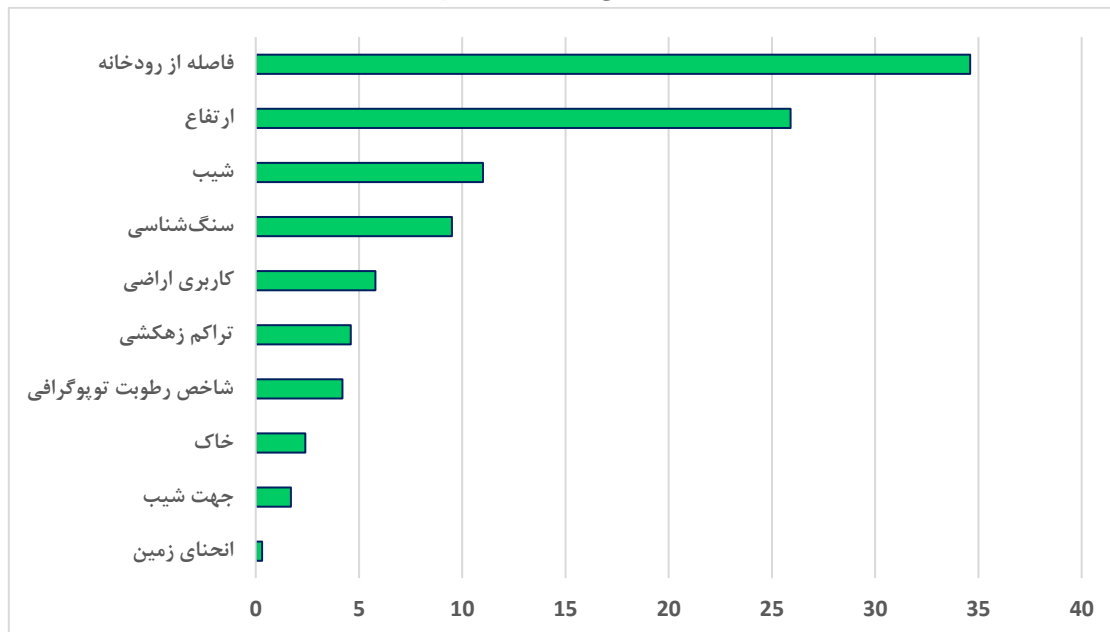
ذکر این نکته نیز ضروری است که در روش‌های به‌کاررفته در مطالعات آزادی و همکاران (۱۳۹۹) و یوسفی و همکاران (۱۴۰۰) نیز مناطق مرکزی و غربی حوضه، بیشتر در معرض خطر سیل هستند که با نتایج مطالعه حاضر هم‌سو



شکل (۵): پهنه‌بندی پتانسیل خطر سیل و آب‌گرفتگی در حوضه کشکان (به ترتیب از بالا، مدل‌های 1-CART, 2-GLM, 3-GAM)



ادامه شکل (۵): پهنه‌بندی پتانسیل خطر سیل و آب‌گرفتگی در حوضه کشکان (به ترتیب از بالا، مدل‌های 1-CART, 2-GLM, 3-GAM)



شکل (۶): درصد مشارکت فاکتورهای مؤثر در وقوع سیل حوضه کشکان

جدول (۱): درصد مساحت کلاس‌های حساسیت پذیری در مدل‌های مختلف

CART		GLM		GAM		حساسیت مناطق به سیل‌گیری
درصد	مساحت کیلومتر مربع	درصد	مساحت کیلومتر مربع	درصد	مساحت کیلومتر مربع	
۶۷	۶۸۴۷	۸۰	۸۲۰۲	۷۵	۷۶۷۸	خیلی کم
۱۵	۱۵۲۰	۸	۸۵۶	۸	۸۵۷	کم
۶	۵۹۴	۳	۳۳۲	۵	۵۰۶	متوسط
۲	۹۷	۲	۱۳۴	۵	۳۴۶	زیاد
۱۰	۵۲۲	۵	۵۸	۷	۱۹۳	خیلی زیاد

جدول (۲): دقت پیش‌بینی مدل‌ها بر مبنای روش‌های ارزیابی منحنی ROC و شاخص TSS

مدل‌ها			معیار ارزیابی
CART	GLM	GAM	
۰/۸۸	۰/۸۰	۰/۸۴	TSS
۰/۹۱	۰/۸۳	۰/۸۷	AUC

پس از آن مدل GAM با $AUC=0/87$ و شاخص $0/84$ و TSS= $0/88$ و مدل GLM با $AUC=0/83$ و شاخص $0/88$ و TSS= $0/91$ قرار دارند. دقت مدل CART نشان‌دهنده دقت عالی این مدل برای حوضه کشکان است. مقایسه نتایج این پژوهش با یافته‌های دیگر محققان که در قسمت نتایج و بحث آورده شد، نشان داد که مدل‌های یادگیری ماشین در مقایسه با مدل‌های آماری از جمله مدل‌های نسبت فراوانی، شاخص آماری، آنتروپی شانون، شواهد وزنی و شواهد قطعی که اجرای آن‌ها راحت‌تر است از دقت بالاتری برخوردار می‌باشند؛ همچنین بر اساس نتایج به دست آمده مشخص شد که در حوضه کشکان پارامترهای محیطی فاصله از رودخانه، ارتفاع و شیب، تأثیرگذارترین عوامل در وقوع سیلاب می‌باشند. بر اساس نقشه پتانسیل خطر سیل تهیه شده با استفاده از مدل CART، حدود ۲۰ درصد از مساحت حوضه کشکان در رده پتانسیل خطر متوسط تا خیلی زیاد قرار گرفته است. با توجه به نقشه پتانسیل خطر سیل، بسیاری از سازه‌های انسانی و زمین‌های کشاورزی در نواحی بحرانی و حساس به وقوع سیل قرار گرفته‌اند که نشان می‌دهد وقوع سیل در حوضه کشکان می‌تواند منجر به

نتیجه‌گیری

در سال‌های اخیر مدل‌های یادگیری ماشین محبوبیت زیادی در زمینه مدل‌سازی مکانی مخاطرات طبیعی به دست آورده‌اند. با توجه به اینکه سیلاب یکی از پرخطرترین حوادث طبیعی در ایران است و با توجه به هزینه‌های بالای نمونه‌برداری و مشکل کمبود داده در اکثر حوضه‌های کشور، استفاده از مدل‌سازی مکانی برای تهیه نقشه‌های پتانسیل خطر سیلاب با هدف مشخص کردن مناطق بحرانی و حساس به وقوع سیلاب ضروری است، زیرا یکی از گام‌های پایه در کنترل و کاهش خسارت سیل‌های آینده است. در این پژوهش به شناسایی مناطق حساس به وقوع سیل در حوضه کشکان با استفاده از مدل‌های یادگیری ماشین CART، GLM و GAM و با بهره‌گیری از نرم‌افزارهای ArcGIS و R پرداخته می‌شود. بر اساس نتایج، نقشه سیل تهیه شده با استفاده از مدل CART با $AUC=0/91$ و شاخص $TSS=0/88$ نسبت به دو مدل GLM و GAM دارای بیشترین دقت است؛ بنابراین دقیق‌ترین مدل در پیش‌بینی پتانسیل خطر سیل بوده و



این پژوهش مشخص شد که تجزیه و تحلیل عوامل محیطی، با استفاده از داده‌های سنجش از دور و مدل‌سازی مکانی، ابزاری قدرتمند برای پیش‌بینی مناطق حساس به شمار می‌رود. با توجه به دقت بالاتر مدل‌های یادگیری ماشین و به‌خصوص مدل CART و با توجه به کمبود داده‌های هیدرولوژیکی در اکثر حوزه‌های آبخیز کشور و کاربرد عالی مدل‌سازی مکانی برای مدیریت بلایای طبیعی در کشور پیشنهاد می‌شود پژوهشگران از این مدل به‌منظور تهیه نقشه‌های پتانسیل خطر سیلاب در مناطق دیگر کشور نیز استفاده و نتایج را با مدل‌های دیگر مقایسه کنند.

خسارات جانی و مالی فراوانی شود. یافته‌های این پژوهش با توجه به ایجاد و طرح نقشه پتانسیل خطر سیل و آب‌گرفتگی حوضه سیل‌خیز کشکان می‌تواند برای کارهای مدیریتی و کنترلی سیلاب، بسیار کاربردی باشد و می‌تواند به مدیران و برنامه‌ریزان برای انجام اقدامات مناسب به‌منظور پیش‌گیری و کاهش خطرات ناشی از سیل در آینده کمک نماید؛ هم‌چنین می‌توان از آن به‌منظور شناسایی مناطق مناسب و ایمن برای توسعه‌های عمرانی آینده در منطقه مورد مطالعه استفاده کرد و از ساخت‌وساز در مناطق با پتانسیل خطر سیل زیاد و خیلی زیاد خودداری نمود. در

منابع

- آزاد طلب، م.، شهابی، ه.، شیرزادی، ع.، چپی، ک. ۱۳۹۹. پهنه‌بندی خطر سیلاب در شهر سنجند با استفاده از مدل‌های ترکیبی شاخص آماری و تابع شواهد قطعی. ۹ (۳۶): ص ۲۷-۴۰.
- آزادی، ف. صدوق، س. ح.، قهرودی، م.، شهابی، ه. ۱۳۹۹. پهنه‌بندی حساسیت خطر سیل در حوضه آبخیز رودخانه کشکان با استفاده از دو مدل WOE و EBF. نشریه علمی جغرافیا و مخاطرات محیطی. ۹ (۱): ص ۴۵-۶۰.
- اسماعیلی علویجه، ا.، کریمی، س.، علوی پور، ف. س. ۱۳۹۹. ارزیابی آسیب‌پذیری مناطق شهری در برابر سیل با منطق فازی (مطالعه موردی: منطقه ۲۲ تهران). فصلنامه علوم و تکنولوژی محیط زیست. ۲۲ (۳): ص ۳۴۹-۳۶۱.
- برخورداری، ج.، زارع مهرجردی، م.، یوسفی، م. ۱۳۹۳. تاثیر پخش سیلاب بر برخی خصوصیات پوشش گیاهی و خاک در ایستگاه آبخوان سرچاهان-استان هرمزگان. نشریه پژوهش‌های آبخیزداری. ۲ (۲۷): ص ۳۳-۴۲.
- حسین زاده، م.، پناهی، ر.، تربند، ت. ۱۳۹۹. پهنه‌بندی حساسیت سیل در حوضه آبریز سنقر در استان کرمانشاه. نشریه اکوهیدرولوژی. ۷ (۴): ص ۸۷۳-۸۸۹.
- حسینی، س. م.، جعفربیگلو، م.، گراوند، ف. ۱۳۹۴. تعیین پهنه‌های سیل‌گیر رود کشکان با استفاده از مدل هیدرولیکی به‌منظور کاهش مخاطرات سیل. نشریه دانش مخاطرات. ۲ (۳): ص ۳۵۵-۳۶۹.
- داودی مقدم، د.، حقی زاده، ع. ۱۳۹۹. شناسایی مناطق مستعد سیل‌گیری و تعیین مهم‌ترین عامل‌های مؤثر بر وقوع آن با استفاده از مدل پیشینه‌ی بی‌نظمی در آبخیز تشان خوزستان. نشریه پژوهش‌های آبخیزداری. ۳۳ (۴): ص ۹۶-۱۱۲.
- سیاه کمری، ص.، زینی وند، ح. ۱۳۹۵. پتانسیل‌یابی مناطق مستعد سیل با استفاده از مدل شاخص آماری و وزن شواهد (حوزه آبخیز مادرسو، گلستان). نشریه سنجش از دور و سامانه اطلاعات جغرافیایی در منابع طبیعی. ۷ (۴): ص ۱۱۶-۱۳۳.
- عابدینی، م.، بهشتی جاوید، ا. ۱۳۹۵. پهنه‌بندی خطر وقوع سیلاب حوضه آبخیز ليقوان چای با استفاده از مدل فرآیند تحلیل شبکه و سیستم اطلاعات جغرافیایی. ۱۶ (۵۵): ص ۲۹۳-۳۱۲.
- عابدینی، م.، فتحی، م. ح. ۱۳۹۴. پهنه‌بندی خطر وقوع سیلاب با استفاده از فرآیند تحلیل شبکه (مطالعه موردی: حوضه آبخیز خیاو چای). نشریه هیدروژئومورفولوژی. ۳: ص ۹۹-۱۲۰.



- عرب عامری، ع.، پورقاسمی، ح. ر.، شیرانی، ک. ۱۳۹۶. پهنه بندی حساسیت سیل گیری با استفاده از روش ترکیبی نوین تئوری بیزین— فرایند تحلیل سلسله مراتبی (مطالعه موردی: حوضه آبخیز نکا — استان مازندران). نشریه اکوهیدرولوژی. ۴ (۲): ۴۴۷-۴۶۲.
- فرامرزی، ح.، حسینی، س. م.، پورقاسمی، ح.، فرنقی، م. ۱۳۹۸. ارزیابی و پهنه بندی وقوع مخاطره سیلاب در پارک ملی گلستان. نشریه اکوهیدرولوژی. ۶ (۴): ۱۰۵۵-۱۰۶۸.
- مختاری، د.، رضایی مقدم، م. ح.، رحیم پور، ت.، معزز، س. ۱۳۹۹. تهیه نقشه خطر وقوع سیلاب در حوضه آبریز گمناب چای با استفاده از مدل ANP و تکنیک GIS. نشریه اکوهیدرولوژی. ۷ (۲): ۵۰۹-۴۹۷.
- یوسفی، ح.، یونسسی، ح.، ارشیا، ا.، یاراحمدی، ی.، گودرزی، ا. ۱۴۰۰. تعیین مناطق مستعد سیل با مدل های SI, FR و Shannon به منظور کاهش مخاطرات سیل (مطالعه موردی: حوزه آبخیز کشکان). نشریه اکوهیدرولوژی. ۸ (۱): ۳۱۹-۳۰۷.
- Arianpour, M. and Jamali, A. A. 2015. Flood Hazard Zonation using Spatial Multi-Criteria Evaluation (SMCE) in GIS (Case Study: Omidieh-Khuzestan). *European Online Journal of Natural and Social Sciences*. 4(1): 39 – 49.
- Bates PD. 2012. Integrating remote sensing data with flood inundation models: how far have we got? *Hydrol. Process*. (26):2515–2521.
- Breiman L, Friedman JH, Olshen RA, Stone CJ. 1984. *Classification and regression trees*. Wadsworth, Belmont, CA.
- Brenning A. 2009. Benchmarking classifiers to optimally integrate terrain analysis and multispectral remote sensing in automatic rock glacier detection. *Remote Sens Environ* 113 (1):239–247.
- Bui, D. T., Panahi, M., Shahabi, H., Singh, V. P., Shirzadi, A., Chapi, K., Ahmad, B. B. 2018. Novel hybrid evolutionary algorithms for spatial prediction of floods. *Scientific reports*, 8(1), 15364.
- Chapi, K., Singh, V. P., Shirzadi, A., Shahabi, H., Bui, D. T., Pham, B. T., Khosravi, K. 2017. A novel hybrid artificial intelligence approach for flood susceptibility assessment. *Environmental Modelling & Software*. 95, 229-245.
- Davoudi Moghaddam, D., Pourghasemi, H.R., Rahmati, O. 2019. Assessment of the Contribution of Geo-environmental Factors to Flood Inundation in a Semi-arid Region of SW Iran: Comparison of Different Advanced Modeling Approaches. In *Natural Hazards GIS-Based Spatial Modeling Using Data Mining Techniques, Advances in Natural and Technological Hazards Research*. Springer: Cham, Switzerland. Volume 48, pp. 59–78.
- Davoudi Moghaddam, D., Rahmati, O., Haghizadeh, A., Kalantari, Z. 2020. A Modeling Comparison of Groundwater Potential Mapping in a Mountain Bedrock Aquifer: QUEST, GARP, and RF Models. *Water* 2020, 12, 679. <https://doi.org/10.3390/w12030679>.
- Fernandez DS, Lutz MA. 2010. Urban flood hazard zoning in Tucuman Province, Argentina, using GIS and multicriteria decision analysis. *Engineering Geology*. 111: 90-98.
- Goetz JN, Guthrie RH, Brenning A. 2011. Integrating physical and empirical landslide susceptibility models using generalized additive models. *Geomorphology* 129:376–386
- Gordon, L. 2013. Using classification and regression trees (CART) in SAS® enterprise miner TM for applications in public health. *Public Health*.
- Kanani-Sadat, Y., Arabsheibani, R., Karimipour, F., Nasserli, 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*. Volume 572, pp 17-31.
- Khosravi, K., Pourghasemi, H. R., Chapi, K., Bahri, M. 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.
- Loh, W.Y. 2011. *Classification and regression trees*. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 1(1):14-23.



Mutzner R, Tarolli P, Sofia G, Parlange MB, Rinaldo A. 2016. Field study on drainage densities and rescaled width functions in a high-altitude alpine catchment. *Hydrological Processes*. 30(13): 2138-2152.

Ozdemir A, Altural T. 2013. A comparative study of frequency ratio, weights of evidence and logistic regression methods for landslide susceptibility mapping: Sultan Mountains, SW Turkey. *J Asian Earth Sci* 64:180-197.

Regmi, N. R., Giardino, J.R., Vitek, J. D. 2010. Modeling susceptibility to landslides using the weight of evidence approach: Western Colorado, USA. *Geomorphology*. 115(1):172-187.

Sharma, H., Kumar, S. 2016. A survey on decision tree algorithms of classification in data mining. *International Journal of Science and Research*. 5(4):209497.

Tehrany, M. S., Pradhan, B., Jebur, M. N. 2013. Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS. *J. Hydrol*. 504(1):69-79.

Tehrany MS, Pradhan B, Jebur MN. 2014. Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. *J. Hydrol*. 512(1):332-343.

Tehrany, M. S., Jones, S., Shabani, F. 2019. Identifying the essential flood conditioning factors for flood prone area mapping using machine learning techniques. *Catena*. 1(175):174-92.

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(1):91-101.



Determination of Flood potential Using CART, GLM and GAM Machine learning Models (Case study: Kashkan Basin)

Hossein Yosefi^۱, Hojjatallah Yonesi^{۲*}, Davoud Davoudimoghadam^۳, Azadeh Arshia^۴, Zahra Shamsi^۵

Abstract

Flood is a phenomenon that causes a lot of environmental and socio-economic damage. The purpose of this study is to evaluate the efficiency of CART, GLM and GAM machine learning models in identifying flood risk areas in the Kashkan basin. Lorestan province and especially Kashkan basin, including: Selseleh, Delfan, Doreh, Khorramabad, Poldakhtar and Kuhdasht, is flooded and has suffered flood damage many times and in April 2019, experienced the largest flood of the last 200 years. In this regard, various factors including: height, slope direction, land curvature, slope percentage, distance from the river, drainage density, soil, lithology, land use and topographic moisture index were used. The digital map of all the mentioned factors was prepared in ArcGIS10.5 software and in the form of a database. The location of 123 flood events recorded in recent years in this basin was collected and randomly used in two categories of model training (86 cases) and model validation (37 cases) in modeling. Using machine learning models and environmental factors, flood potential prediction maps were prepared and then validated using AUC characteristic curve methods and TSS index. The results of model validation showed that CART machine learning model with AUC = 0.91 and TSS = 0.88 index was the most accurate model in predicting flood risk potential, followed by GAM model with AUC = 0.87 and TSS index = 0.84 and GLM model with AUC = 0.83 and TSS index = 0.88. Accuracy 0.91 CART model indicates the excellent accuracy of this model for the Kashkan basin. This model shows a larger area of the basin under high potential and moderate flood risk conditions, which include most of the western areas as well as the central areas of the basin (Kuhdasht, Khorramabad and Poldakhtar), which are exactly parts The same areas were flooded in the great flood of 2019 and it is necessary to be in the first priority of flood risk planning and management in this basin.

Keywords: AUC, flood map, ROC, Topographic moisture index, TSS

^۱ Associate Professor, Department of New Energy and Environment, University of Tehran, Iran. Email: hosseinyousefi@ut.ac.ir

^{۲*} Assistant Professor, Department of Water Engineering, Lorestan University, Khorramabad, Iran Email: yonesi.h@lu.ac.ir

^۳ PhD in Watershed Management, Lorestan University, Khorramabad, Iran Email: d.davoudi.m@gmail.com

^۴ PhD student in Hydraulic Structures, Lorestan University, Khorramabad, Iran Email: azadeharshia69@gmail.com

^۵ PhD student in Hydraulic Structures, Lorestan University, Khorramabad, Iran Email: azm.shams@gmail.com

Research Paper

Determination of Flood Potential Using CART, GLM and GAM Machine Learning Models (Case Study: Kashkan Basin)

Hossein Yosefi¹,Hojjatallah Yonesi^{2*},Davoud Davoudimoghadam³Azadeh Arshia^{4*},Zahra Shamsi^{5*}¹ Associate Professor, Department of New Energy and Environment, University of Tehran, Iran² Assistant Professor, Department of Water Engineering, Lorestan University, Khorramabad, Iran³ PhD in Watershed Management, Lorestan University, Khorramabad, Iran⁴ PhD student in Hydraulic Structures, Lorestan University, Khorramabad, Iran⁵ PhD student in Hydraulic Structures, Lorestan University, Khorramabad, Iran

10.22125/IWE.2021.293391.1529.

Received:
July.12.2010Accepted:
September.29.2021Available online:
June.01.2022

Keywords:
AUC, flood map,
ROC, Topographic
moisture index, TSS

Abstract

Flood is a phenomenon that causes a lot of environmental and socio-economic damage. The purpose of this study is to evaluate the efficiency of CART, GLM and GAM machine learning models in identifying flood risk areas in the Kashkan basin. Lorestan province and especially Kashkan basin, including: Selseleh, Delfan, Doreh, Khorramabad, Poldakhtar and Kuhdasht, is flooded and has suffered flood damage many times and in April 2019, experienced the largest flood of the last 200 years. In this regard, various factors including: height, slope direction, land curvature, slope percentage, distance from the river, drainage density, soil, lithology, land use and topographic moisture index were used. The digital map of all the mentioned factors was prepared in ArcGIS10.5 software and in the form of a database. The location of 123 flood events recorded in recent years in this basin was collected and randomly used in two categories of model training (86 cases) and model validation (37 cases) in modeling. Using machine learning models and environmental factors, flood potential prediction maps were prepared and then validated using AUC characteristic curve methods and TSS index. The results of model validation showed that CART machine learning model with AUC = 0.91 and TSS = 0.88 index was the most accurate model in predicting flood risk potential, followed by GAM model with AUC = 0.87 and TSS index = 0.84 and GLM model with AUC = 0.83 and TSS index = 0.88. Accuracy 0.91 CART model indicates the excellent

* **Corresponding Author:** Hojjatallah Yonesi

Address: Assistant Professor, Department of Water Engineering, Lorestan University, Khorramabad, Iran

Email: yonesi.h@lu.ac.ir**Tel:** 09166613480

accuracy of this model for the Kashkan basin. This model shows a larger area of the basin under high potential and moderate flood risk conditions, which include most of the western areas as well as the central areas of the basin (Kuhdasht, Khorramabad and Poldakhtar), which are exactly parts. The same areas were flooded in the great flood of 2019 and it is necessary to be in the first priority of flood risk planning and management in this basin.

1. Introduction

Flood is a phenomenon that causes a lot of environmental and socio-economic damage. The purpose of this study was to evaluate the efficiency of CART, GLM and GAM machine learning models in identifying critical and flood risk areas in the Kashkan basin. Lorestan province and especially Kashkan basin, including: Selseleh, Delfan, Doreh, Khorramabad, Poldakhtar and Kuhdasht, is very flood-prone and has suffered flood damage many times, and in April 2019, the biggest flood in the last 200 years. Has experienced. Areas of the country in recent years have always been affected by floods with high damage, and these areas need to be prioritized in studies and management. Kashkan Basin is one of the areas that has been affected by frequent and frequent floods. Various studies are always important and necessary for flood management in this region. The purpose of this study is to provide a flood zoning map to reduce flood hazards in this flood-prone basin using CART, GLM and GAM machine learning models and using ArcGIS10 software. 5. To improve decision-making and flood management in the region.

2. Materials and Methods

Lorestan province is located in western Iran and its area is about 2855 square kilometers. Kashkan River with a length of 290 km is the most flooded sub-basin of Karkheh. In terms of distribution of tectonic formations, about 50% of the surface of this basin is composed of new alluvial formations, Amiran and Asmari-Shahbazan joint, which have a very high potential for runoff production. The physiological condition of this basin indicates the fact that the factors of area, concentration time, compaction coefficient of the basin have a very high potential for runoff production in a short time and increase its flooding capacity. According to scientific sources and according to the location of the region in this study, altitude factors, slope direction, land curvature, slope percentage, distance from the river, drainage density, soil, lithology, land use and topographic moisture index to predict potential Floods in the Kashkan basin were selected. 123 points which are among the flood points of this basin were used in this research. In each modeling, a group consisting of 86 flood events (70%) was randomly selected for training and a group consisting of 37 flood events (30%) for Validation split. The CART model is a way to create predictive models of data. This model distributes the input data to itself recursively and is able to process predictive variables and numerical and batch targets. The GLM model is one of the most common statistical methods used for spatial modeling of environmental variables. This model is a parametric method and is an extension of linear models. The Generalized General Model (GAM) is a nonparametric model that is an extension of generalized linear models (GLM) and has itself been developed from linear models. In this model, unlike the linear regression model, the data are allowed to determine the shape of the response curve.

3. Results

In identifying flood and flood sensitive areas in the Kashkan watershed, the CART model has excellent accuracy and the GAM and GLM models have very good accuracy. According to the results, the most important environmental parameters affecting the occurrence of floods in the Kashkan basin are distance from the river, altitude and slope.

4. Discussion and Conclusion

Comparative analyzes between the location of past floods and environmental parameters affecting the occurrence of floods were performed and maps to predict the potential for flood risk and flooding of

Kashkan watershed were prepared using CART, GLM and GAM models. Based on various sources, using the Quantile technique in ArcGIS 10.5 software, they were classified into five classes of very low, low, medium, high and very high flood risk potential. The results showed that the classes with very low, low, medium, high and very high risk potentials have different percentages of different areas of Kashkan basin in different models. In order to evaluate the accuracy of flood and flood potential prediction maps, two methods of ROC curve and TSS index were used, using 37 flood events related to previous periods that were not used in the modeling. The accuracy of flood potential prediction in different models is calculated based on the ROC curve method and TSS index, the results of which can be seen in the table below. Validation results showed that the CART machine learning model with a sub-curve area of 0.91 and a TSS index of 0.88 was the most accurate model in predicting flood risk potential, followed by the GAM model with a sub-curve area of 0.87 and a TSS index of 0.84. The GLM model has a curve below 0.83 and a TSS index of 0.80.

5. Six important references

1. Arianpour, M. and Jamali, A. A. 2015. Flood Hazard Zonation using Spatial Multi-Criteria Evaluation (SMCE) in GIS (Case Study: Omidieh-Khuzestan). *European Online Journal of Natural and Social Sciences*. 4(1): 39 – 49.
2. Bates PD. 2012. Integrating remote sensing data with flood inundation models: how far have we got *Hydrol. Process.* (26):2515–2521.
3. Chapi, K., Singh, V. P., Shirzadi, A., Shahabi, H., Bui, D. T., Pham, B. T., Khosravi, K. 2017. A novel hybrid artificial intelligence approach for flood susceptibility assessment. *Environmental Modelling & Software*. 95, 229-245.
4. Khosravi, K., Pourghasemi, H. R., Chapi, K., Bahri, M. 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.
5. Loh, W.Y. 2011. Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. 1(1):14-23.
6. Tehrany MS, Pradhan B, Jebur MN. 2014. Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. *J. Hydrol.* 512(1):332–343.

Conflict of Interest

Authors declared no conflict of interest.