



## مدیریت آب و آبیاری (نشریه علمی)

دوره ۱۱ ■ شماره ۲ ■ تابستان ۱۴۰۰

صفحه‌های ۲۳۵-۲۲۳

DOI: 10.22059/jwim.2021.317527.856

مقاله پژوهشی:

### پهنه‌بندی احتمال وقوع سیل با استفاده از بررسی مقایسه‌ای دو مدل شناخته‌شده جنگل تصادفی و ماشین بردار پشتیبان در شمال ایران

محمد رضا طهماسبی<sup>۱</sup>، سعید شهبانلو<sup>۲\*</sup>، احمد رجبی<sup>۳</sup>، فریبرز یوسفوند<sup>۳</sup>

۱. دانشجوی دکتری، گروه مهندسی آب، دانشکده کشاورزی، واحد کرمانشاه، دانشگاه آزاد اسلامی، کرمانشاه، ایران.

۲. دانشیار، گروه مهندسی آب، دانشکده کشاورزی، واحد کرمانشاه، دانشگاه آزاد اسلامی، کرمانشاه، ایران.

۳. استادیار، گروه مهندسی آب، دانشکده کشاورزی، واحد کرمانشاه، دانشگاه آزاد اسلامی، کرمانشاه، ایران.

تاریخ پذیرش مقاله: ۱۴۰۰/۰۵/۰۴

تاریخ دریافت مقاله: ۱۳۹۹/۱۱/۱۳

#### چکیده

هدف از پژوهش پیش رو، پهنه‌بندی احتمال وقوع سیل در حوزه آبخیز سالیان تپه، واقع در استان گلستان می‌باشد. بدین منظور، از دو مدل معروف و شناخته‌شده داده‌کاوی یعنی مدل جنگل تصادفی (RF) و مدل ماشین بردار پشتیبان (SVM) به‌عنوان پنج‌مارک و به لحاظ الگوریتم محاسباتی توانمند در زمینه ارزیابی فرایند وقوع سیلاب استفاده شد. شواهد سیلاب با استفاده از بازدیدهای میدانی، گزارش‌ها و اطلاعات سازمانی موجود ثبت و در سامانه اطلاعات جغرافیایی (GIS) در قالب نقشه تهیه شد. هم‌چنین، با توجه به مرور منابع گسترده، سیزده عامل زمینه‌ساز شامل فاصله از آبراهه، واحدهای سنگ‌شناسی، درصد شیب، بافت خاک، جهت شیب، کاربری اراضی، انحنای طولی و عرضی دامنه، شاخص رطوبت، شاخص توان فرسایشی آبراهه و طبقات ارتفاعی به‌عنوان عوامل مؤثر بر وقوع سیل در منطقه مورد مطالعه انتخاب و لایه‌های مذکور در سامانه اطلاعات جغرافیایی تهیه شدند. در این مطالعه بعد از آماده‌سازی لایه‌ها، برای آنالیز این داده‌ها و بررسی هم‌خطی آن‌ها از نرم‌افزار SPSS استفاده شد. به‌منظور ارزیابی نتایج مدل‌ها، از مقدار مساحت زیر منحنی تشخیص عملکرد نسبی (ROC) استفاده شد. سه سری متفاوت از نقاط وقوع خطر سیل (S1, S2, S3) شامل ۷۰ درصد برای آموزش مدل و ۳۰ درصد برای اعتبارسنجی به‌صورت تصادفی آماده شد تا دقت و صداقت مدل مورد ارزیابی قرار بگیرد. نتایج نشان داد نقشه میانگین حاصل از مدل جنگل تصادفی در مرحله اعتبارسنجی با مساحت زیر منحنی ۹۶ درصد و صداقت ۰/۰۰۱ کارایی بهتری نسبت به مدل ماشین بردار پشتیبان در پهنه‌بندی سیلاب در حوضه مورد مطالعه دارد.

**کلیدواژه‌ها:** استان گلستان، جنگل تصادفی، سیل، صداقت مدل، مدل ماشین بردار پشتیبان.

### Flood probability zonation using a comparative study of two well-known random forest and support vector machine models in northern Iran

Mohammad Reza Tahmasebi<sup>1</sup>, Saeid Shabanlou<sup>2\*</sup>, Ahmad Rajabi<sup>3</sup>, Fariborz Yosefvand<sup>3</sup>

1. Ph.D. Candidate, Department of Water Engineering, College of Agriculture, Kermanshah Branch, Islamic Azad University, Kermanshah, Iran.

2. Associate Professor, Department of Water Engineering, College of Agriculture, Kermanshah Branch, Islamic Azad University, Kermanshah, Iran.

3. Assistant Professor, Department of Water Engineering, College of Agriculture, Kermanshah Branch, Islamic Azad University, Kermanshah, Iran.

Received: February, 01, 2021

Accepted: July, 26, 2021

#### Abstract

The current study is aimed to zoning flood probability map in the Saliantapeh catchment is located in the Golestan Province. To this aim, two well-known data mining models namely Random Forest (RF) and Support Vector Machine (SVM) were applied due to their robust computational algorithm. Flood inventories were gathered through several field surveys using local information and available organizational resources and corresponding map was created in the geographic information system. Reviewing several worldwide studies, 13 predisposing variables including proximity to stream, soil texture, lithological units, land use/cover, slope percent, elevation/DEM, slope aspect, plan curvature, profile curvature, stream power index and topographic wetness index were chosen and the corresponding maps were generated in the geographic information system. In this study, after preparing the predictor maps, SPSS software was used to analyze this data and testing Multi-collinearity. In order to evaluate models' results the area under the receiver operating were used. Three different sample data sets (s1, s2, s3) including 70% for training and 30% for validation were randomly gathered to evaluate the robustness of the applied models. Results showed that the RF model with the area under curve value of 0.96 and robustness of 0.001 in validation step had better performance on flood probability zonation over the study area.

**Keywords:** Flood, Golestan Province, Random forest, Robustness, Support vector machine.

## مقدمه

یکی از این بلاهای طبیعی که منجر به خسارت‌های زیادی در دهه‌های اخیر شده است، سیل می‌باشد. روند افزایش سیل در پنج دهه گذشته نشان می‌دهد که تعداد وقوع سیل در دهه ۸۰ نسبت به دهه ۴۰ کمابیش ده برابر شده است و به بیان دیگر ۹۰۰ درصد افزایش داشته است (Abdi, 2006). سیل جریان آب بیش از حد طبیعی است؛ چنین حالتی در صورتی که جان انسان‌ها و زندگی بشر را به خطر بیندازد بلا به‌شمار می‌رود. از نظر هیدرولوژیست‌ها اندازه سیل بهترین نحو با حداکثر جریان تشریح می‌شود (Hall, 1981). تجارب کشورهای مختلف نشان می‌دهد که اولین مرحله جهت کاهش آثار زیان‌بار سیل، شناخت مناطق سیل‌گیر و پهنه‌بندی این مناطق از لحاظ میزان خطر سیل‌گیری است تا براساس نتایج حاصله بتوان در رابطه با نحوه استفاده از اراضی و کاربری‌های مختلف از جمله کشاورزی، شهری و صنعتی تصمیم‌گیری نمود و زیان‌های ناشی از سیل را به حداقل ممکن کاهش داد.

فرایند وقوع این خطر بسیار پیچیده بوده و عوامل مختلفی در رخداد آن‌ها نقش دارند. از این‌رو، پیش‌بینی وقوع آن نیز مشکل است. در سال‌های گذشته، ارزیابی حساسیت<sup>۱</sup> برای جوامع بین‌المللی موضوع مهم پژوهش شده است و تبیین و به‌کارگیری مدل‌های مختلف برای پیش‌بینی مکانی مناطق مستعد وقوع خطرات طبیعی توسط پژوهش‌گران مختلفی مورد بررسی قرار گرفته است. پیش‌بینی مکانی رخداد خطرات طبیعی از طریق به‌کارگیری مدل‌هایی که به‌واسطه داده‌های مکانی به‌وجود آمده و خروجی آن‌ها منجر به تهیه نقشه‌های حساسیت به هرکدام از این خطرات می‌شود، مناسب‌ترین راه‌کار برای برنامه‌ریزی مدیریت اراضی در حوزه‌های آبخیز جهت پیش‌گیری از رخداد این وقایع است (Rahi, 2018). مدل‌های ماشین یادگیری به لحاظ به‌کارگیری

الگوریتم‌های تخمین توزیع، تکرار بالای فرایند مدل‌سازی و طبیعت داده محور، توانایی بالایی در شناسایی رفتار وقوع خطر سیل داشته و در پژوهش‌های متعددی برتری نسبی خود را نسبت به مدل‌های آماری دو متغیره و چند متغیره اثبات نموده‌اند. هم‌چنین داده‌کاوی را می‌توان فرایند و تکنولوژی تشخیص ناشناخته‌های گذشته برای رسیدن به مزیت رقابتی نامید. در داده‌کاوی بر جستجو در میان انبوه داده برای استخراج الگوهایی که تشخیص آن‌ها برای انسان بسیار پیچیده و ظریف است تأکید می‌گردد (Nouri & Eskandi, 2009).

از این‌رو، مطالعات مختلفی با استفاده از مدل‌های مختلف از جمله در بحث سیل در نقاط مختلف ایران صورت گرفته است؛ Rahmati et al. (2016a, 2015)، Khosravi et al. (2016)، Bui et al. (2019) و Kheyrizadeh et al. (2012)، با استفاده از مدل ANP به پهنه‌بندی پتانسیل خطر وقوع سیلاب در حوزه آبخیز مردق چای پرداختند. نتایج نشان داد پنج درصد از سطح حوضه که مربوط به قسمت‌های پایین‌دست حوضه است در خطر وقوع سیلاب بسیار زیاد و حدود ۲۹ درصد از سطح حوضه در خطر وقوع سیلاب زیاد است.

Rahmati & Phourghasemi (2017) به‌منظور مدل‌سازی مکانی مناطق مستعد سیل در استان گلستان، از سه مدل تابع شواهد قطعی (EBF)، جنگل تصادفی (RF) و درخت رگرسیونی پیشرفته (BRT) استفاده کردند. نتایج نشان داد که هر سه مدل قابلیت بالایی در تجزیه و تحلیل ارتباط میان عوامل محیطی و رخداد‌های سیل دارند. هم‌چنین نتایج ارزیابی دقت مدل‌ها نشان داد که مدل تابع شواهد قطعی دقت بالاتری نسبت به دو مدل دیگر دارد. هم‌چنین در دهه‌های اخیر پژوهش‌های مختلفی در سراسر دنیا برای پیش‌بینی و شناسایی مناطق مستعد سیل با استفاده از این الگوریتم‌ها و مدل‌ها انجام شده است، از

ارزیابی دقت مدل، صداقت و ربوستنس مدل نیز ملاک ارزیابی قرار بگیرد. برای رفع شکاف‌های مطالعه، این مدل‌های شناخته‌شده به‌عنوان پنج‌مارک در حوضه آبریز سالیان‌تپه واقع در استان گلستان مورد استفاده قرار گرفتند. منطقه مورد مطالعه به دلیل تغییرات کاربری اراضی (به‌ویژه شهرنشینی) در جنگل‌ها و تغییرات آب‌وهوایی در برابر پدیده سیلاب بسیار مستعد بوده است. سیل از ویران‌کننده‌ترین خطرات طبیعی در این منطقه است. بنابراین، پهنه‌بندی مناطق مستعد سیل و اقدامات پیش‌بینی خطر سیل می‌تواند، در کاهش تلفات آن به کشاورزی و زندگی بشر کمک کند. از این‌رو، از اهداف اصلی کار فعلی استفاده از دو الگوریتم شناخته‌شده یادگیری ماشین RF و SVM به‌عنوان روش‌های هوشمند و مقایسه مهارت آن‌ها در پهنه‌بندی احتمال وقوع سیل در حوضه آبریز سالیان‌تپه و در نهایت ارزیابی پایداری و صداقت روش‌های ذکرشده با ایجاد سه مجموعه داده جداگانه (S1, S2, S3) است.

## مواد و روش‌ها

### منطقه مورد مطالعه

حوضه آبریز سالیان‌تپه (با مساحت ۴۵۱۵/۱۷ کیلومتر مربع) در استان گلستان در شمال شرقی ایران واقع شده است که در محدوده طول جغرافیایی ۳۶ درجه و ۳۰ دقیقه تا ۳۷ درجه و ۲۰ دقیقه شمالی و عرض جغرافیایی ۵۴ درجه و ۲۰ دقیقه تا ۵۵ درجه و ۴۰ دقیقه شرقی قرار گرفته است (شکل ۱). منطقه مورد مطالعه به دلیل تغییرات کاربری اراضی از جمله جنگل‌زدایی و توسعه بی‌رویه، در سال‌های اخیر به شدت تحت تأثیر وقوع سیل قرار گرفته است. تسریع خطر سیل در اثر فعالیت‌های غیرمسئولانه بشر ایجاد شده است. از نظر توپوگرافیکی میانگین ارتفاع این منطقه بین ۲۰- تا ۳۶۵۲ متر می‌باشد.

جمله: مدل درخت تصمیم (DT)، مدل ترکیبی دومتغیره و چندمتغیره، مدل ماشین‌بردار پشتیبان (SVM) (Tehrany et al., 2013, 2015)؛ نسبت فراوانی (FR) و ماشین بردار پشتیبان (SVM) (Mojaddadi et al., 2017)؛ مدل‌های درخت بایس ساده<sup>۲</sup> (NBTree)، درخت تصمیم‌گیری متناوب<sup>۳</sup> (ADTree) و مدل جنگل تصادفی (RF) (Chen, 2020). (Albers et al., 2016) از رویکرد مدل‌سازی جنگل تصادفی برای تجزیه و تحلیل سیل در سیستم مخزن تنظیم شده در بخشی از کانادا استفاده کردند. Lee & Pradhan (2007) در مطالعه‌ای به پیش‌بینی مکانی نقشه حساسیت به سیلاب با استفاده از مدل‌های جنگل تصادفی و درخت رگرسیون تقویت‌شده در شهر سئول، کره جنوبی پرداختند. ارزیابی صحت مدل جنگل تصادفی، براساس الگوریتم‌های رگرسیون و طبقه‌بندی، دقت بالاتری را نسبت به مدل رگرسیون تقویت‌شده نشان داد. Woznicki et al. (2019) به تهیه نقشه کامل دشت سیلاب در ایالات متحده موقت با استفاده از جنگل تصادفی اقدام نمودند. عملکرد مدل جنگل تصادفی توانایی آن را برای نشان دادن موفقیت سایر مناطق سیلابی نقشه‌برداری نشده را تأیید کرد، درحالی‌که فقط از داده‌های عمومی استفاده می‌شود.

مروری بر پژوهش‌های انجام‌شده نشان می‌دهد که SVM و RF به‌عنوان روش‌های کارآمد برای مدل‌سازی خطر سیل هستند. به‌عبارت دیگر، از این مدل‌های شناخته‌شده به‌طور گسترده برای بررسی و مدل‌سازی وقایع با دقت و کارایی خوب استفاده شده است. تازگی اصلی در مطالعه حاضر این است که برای انتخاب مدل برتر برای پهنه‌بندی مناطق مستعد سیل در منطقه سالیان‌تپه، سه سری متفاوت از نقاط وقوع خطر سیل (S1, S2, S3) شامل ۷۰ درصد برای آموزش مدل و ۳۰ درصد برای اعتبارسنجی به‌صورت تصادفی آماده شد تا علاوه بر

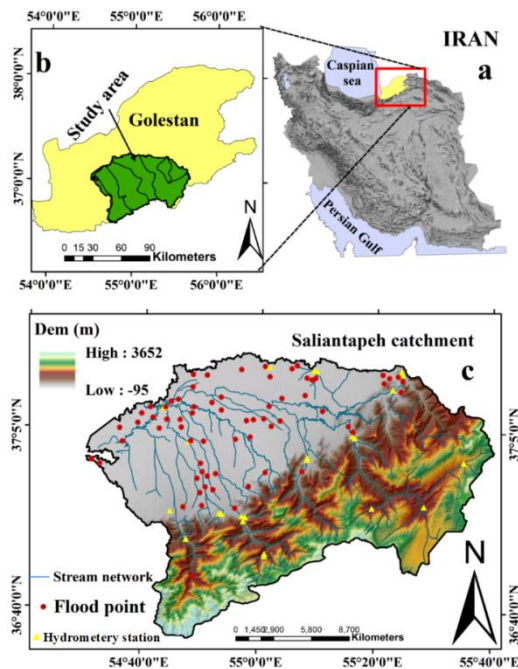


Figure 1. Location of the study area. a) Iran, b) Golestan province, c) Digital elevation model of Saliantapeh catchment

ملی می‌شود که توسط تکنیک‌های متعددی از جمله تفسیر عکس‌های هوایی، مطالعات ژئومورفولوژی، تصاویر ماهواره-ای، وقوع رویدادهای گذشته و پیمایش‌های میدانی (صحرائی) تهیه می‌شود (Guzzetti *et al.*, 2000). در منطقه مورد مطالعه حدود ۷۱ نقطه وقوع سیلاب با استفاده مطالعات صحرائی، استفاده از اطلاعات محلی و گزارش‌های موجود در منطقه و اداره آب منطقه‌ای استان گلستان تهیه شد. در هنگام اجرای مدل‌های داده‌کاوی باید نقاط مورد استفاده برای آموزش مدل<sup>۴</sup> از نقاط اعتبارسنجی<sup>۵</sup> متفاوت باشند (Lee & Pradhan, 2007). بنابراین الگوریتم تقسیم‌بندی تصادفی<sup>۶</sup> (Pourtaghi & Pourghasemi, 2014; Rahmati *et al.*, 2016b) برای جدا کردن نقاط آموزشی و نقاط اعتبارسنجی مورد استفاده قرار گرفت. که در این پژوهش ۷۰ درصد از این نقاط برای آموزش مدل‌ها کالیبراسیون مدل و ۳۰ درصد برای اعتبارسنجی مدل استفاده شد. برای ارزیابی صداقت مدل<sup>۷</sup> و حساسیت داده‌ها (Conoscenti *et al.*, 2014) سه مجموعه داده آماده شد.

میانگین میزان بارش سالانه تقریباً بین ۳۴۸-۸۱۰ میلی‌متر است. کاربری اراضی غالب منطقه مطالعاتی، کشاورزی (۴۳/۶۹ درصد) با مساحت ۱۹۷۳/۲۴ کیلومترمربع است. متوسط حداقل و حداکثر دما به ترتیب ۱۱ و ۱۸/۵ درجه سانتی‌گراد است. منطقه مورد مطالعه شامل ۳۰۰ روستا است و شهرها تا حدی می‌توانند در معرض انحراف و قایع سیل قرار بگیرند. در این پژوهش پس از تهیه نقشه پراکنش سیلاب و انتخاب ۱۳ عامل پیش‌بینی‌کننده پهنه‌بندی احتمال وقوع سیل در حوضه سالیان‌تپه با استفاده از دو مدل شناخته‌شده RF و SVM، در نهایت ارزیابی براساس شاخص AUC و معیار ربوستنس مدل برتر انجام شد. نمودار جریان‌ی مراحل انجام تحقیق در شکل (۲) ارائه شده است.

### تهیه نقشه پراکنش سیلاب

نقشه پراکنش خطرات طبیعی بسته به هدف و قابلیت دسترسی به داده‌ها شامل دو دسته مقیاس محلی و مقیاس

پهنه‌بندی احتمال وقوع سیل با استفاده از بررسی مقایسه‌ای دو مدل شناخته‌شده جنگل تصادفی و ماشین بردار پشتیبان در شمال ایران

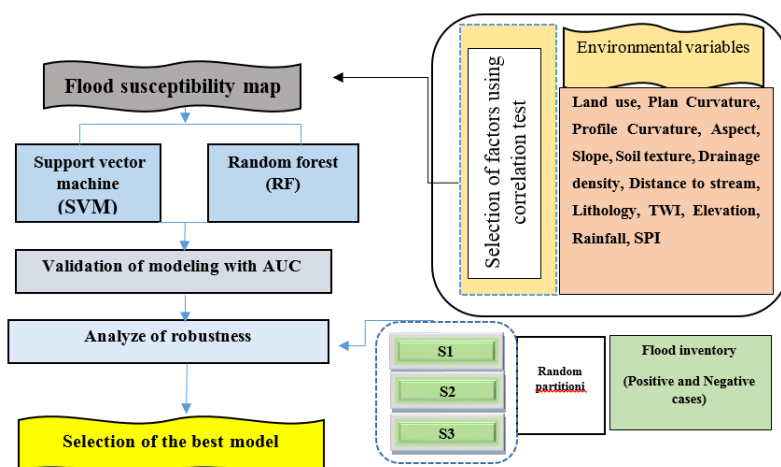


Figure 2. Flow diagram of used method for current research

کنترل جریان‌های سطحی، رواناب و رسوب دارد ( Dickie & Parsons, 2012). نقشه کاربری اراضی و نقشه بافت خاک از اداره منابع طبیعی استان گلستان تهیه شد. نقشه کاربری اراضی با استفاده از تصاویر گوگل ارث اصلاح شد. بافت‌های لومی شنی، لومی رسی، لومی رسی شنی، رسی سیلتی، لومی رسی سیلتی و لومی سیلتی از بافت‌های خاک اراضی این منطقه است. فاصله از شبکه آبراهه یکی از مهم‌ترین متغیرهای محیطی است که روی بزرگی و سرعت سیل تأثیر می‌گذارد ( Glenn *et al.*, 2012). این عامل پایداری یک دامنه را از طریق اشباع مواد بر روی دامنه کنترل می‌کند (Yalcin *et al.*, 2008). نقشه فاصله از شبکه آبراهه با استفاده از تابع Euclidean distance در نرم‌افزار ArcGIS 10.5 تهیه شد. فاصله از شبکه آبراهه از ۰-۱۲۹۴۱ متر متغیر است. به‌منظور تبدیل شبکه آبراهه به یک نقشه و کمیت قابل اندازه‌گیری، نقشه تراکم زه‌کشی با ابزار Line density تهیه شد. نقشه بارندگی سالانه حوزه آبخیز مورد مطالعه با استفاده از اطلاعات بارش ۲۰ ایستگاه اندازه‌گیری که از اداره آب منطقه‌ای استان گلستان جمع‌آوری شد، و با روش IDW<sup>۱۴</sup> درون‌یابی شد. این نقشه از ۳۸۴ تا ۸۱۰ میلی‌متر بر سال

### تهیه عوامل کنترل‌کننده پیش‌بینی‌کننده خطر سیل

در این پژوهش براساس مرور منابع و با بررسی‌های انجام‌شده در منطقه مورد مطالعه حدود سیزده عامل شامل واحدهای سنگ‌شناسی، کاربری اراضی، فاصله از شبکه آبراهه، بافت خاک، درصد شیب، جهت شیب، طبقات ارتفاعی<sup>۸</sup>، بارش، شاخص رطوبت توپوگرافیکی (TWI)، انحنا طولی دامنه، انحنا عرضی دامنه، شاخص قدرت جریان<sup>۹</sup> (SPI) و تراکم شبکه زه‌کشی انتخاب شدند و با استفاده از نرم‌افزار ArcGIS 10.5 و SAGA<sup>۱۰</sup> به لایه‌های رستری با اندازه‌های سلولی ۳۰m×۳۰m تبدیل شدند. در این مطالعه بعد از آماده‌سازی لایه‌ها، برای آنالیز این داده‌ها و بررسی هم‌خطی<sup>۱۱</sup> آن‌ها از نرم‌افزار SPSS استفاده شد (Daoud, 2017). با کنترل‌کردن شاخص‌های آستانه تحمل<sup>۱۲</sup> و عامل تورم واریانس<sup>۱۳</sup> (VIF) هیچ‌کدام از دارای هم‌خطی نبوده و برای مدل‌سازی استفاده شد. لایه-مدل رقومی ارتفاعی (DEM) منطقه مورد مطالعه از اداره منابع طبیعی استان گلستان و نقشه شیب، جهت شیب انحنا عرضی و طولی دامنه از روی نقشه DEM تهیه شده است. کاربری اراضی نقش مهمی در عملکرد فرایندهای هیدرولوژیکی و ژئومورفولوژیکی از طریق

مطالعه حاضر، نقشه واحدهای سنگ‌شناسی منطقه مورد مطالعه با توجه به نقشه‌های زمین‌شناسی موجود در مقیاس ۱:۱۰۰,۰۰۰ از بخش مطالعات زمین‌شناسی ایران، تهیه شده است. حوزه آبخیز مورد مطالعه تحت پوشش انواع مختلف سازندهای سنگ‌شناسی قرار دارد و در ۱۹ گروه طبقه‌بندی شده است. نقشه‌های مربوط به فاکتورهای کنترل‌کننده (لایه‌های پیش‌بینی‌کننده) در شکل (۳) ارائه شده است.

متغیر است. شاخص SPI که بیانگر توان فرسایشی آبراهه بوده و بر ناپایداری دامنه‌ها تأثیر می‌گذارد (Poudyal et al., 2010) از روی مدل رقومی ارتفاعی با استفاده از نرم‌افزار SAGA تهیه شده است. عامل رطوبت توپوگرافیکی با اثرگذاری بر وضعیت رطوبت خاک و توپوگرافی سطح و پاسخ هیدرولوژیکی در وقوع خطر سیلاب نقش دارد (Wilson et al., 2000)، به طوری که رطوبت خاک یک نقش مهم در ناپایداری دامنه دارد.

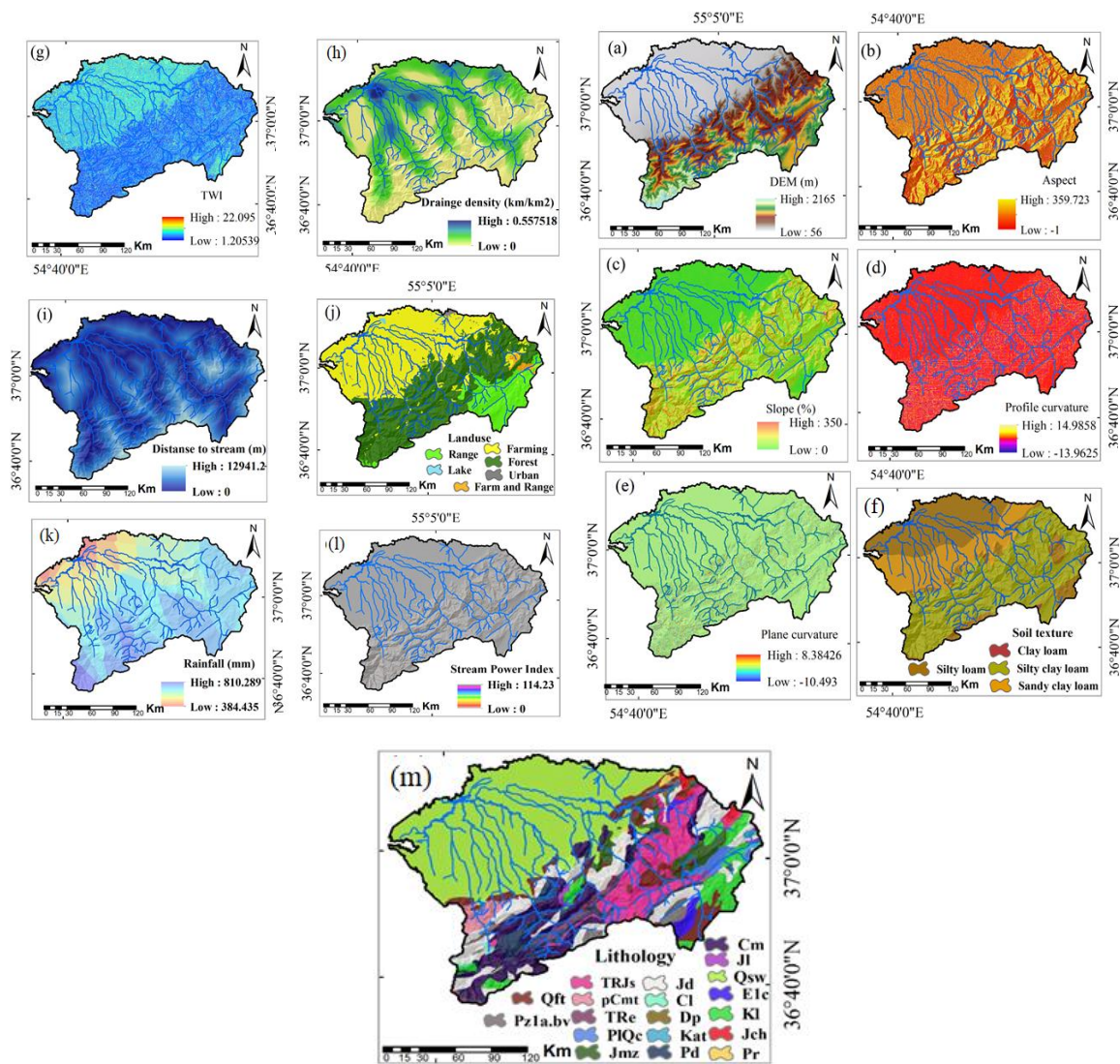


Figure 3. Flood predictor maps [(a) DEM (m), (b) Aspect, (c) Slope, (d) Profile curvature, (e) Plan curvature, (f) Soil texture (g) Topographic Wetness Index, (h) Drainage density, (i) Distance to stream (m), (j) Land use, (k) Rainfall (mm), (l) Stream Power Index, (m) lithology.]

### مدل جنگل تصادفی<sup>۱۰</sup> (RF)

مدل جنگل تصادفی یکی از روش‌های مدل‌سازی درخت تصمیم است (Jafarian & Kargar, 2017) به‌گونه‌ای که از تلفیق نتایج به‌دست‌آمده از درخت‌های مختلف، نتیجه یا برآورد نهایی را راجع به ارتباط میان خطرات طبیعی و عوامل محیطی ارائه می‌دهد. در مسائل مربوط به رگرسیون، جنگل تصادفی تعداد اختیاری از درختان ناآزموده (زیرمجموعه‌ای از متغیرهای مستقل) است که نتایج واکنش این درختان به‌صورت ترکیبی جهت تخمین صحیح استعداد وقوع خطر (متغیر وابسته) مورد استفاده قرار می‌گیرد (Servati et al., 2014). این یک ابزار تحلیلی بسیار قدرتمند جهت جستجوی اطلاعات و اصلاح تعداد بهینه درختان در مدل جنگل تصادفی است. به‌منظور به‌کارگیری این مدل در مدل‌سازی خطرات طبیعی از نرم‌افزار R و پکیج random forest استفاده می‌شود. کار ساخت جنگل با استفاده از درخت‌ها اغلب به‌روش کیسه‌گذاری انجام می‌گیرد. ایده اصلی روش کیسه‌گذاری آن است که ترکیبی از مدل‌های یادگیری، نتایج و عملکرد کلی مدل را افزایش می‌دهد. به بیان ساده، جنگل تصادفی چندین درخت تصمیم می‌سازد و آن‌ها را با یکدیگر ادغام می‌کند تا پیش‌بینی‌های صحیح و پایدارتری حاصل شوند. در این روش بردار تصادفی  $\Theta_k$  که مستقل از بردارهای تصادفی  $\Theta_1 \dots \Theta_{k-1}$  بوده، برای درخت  $K$ ام، تولید می‌شود. هم‌چنین همه بردارها دارای توزیع یکسانی می‌باشند. درخت رگرسیونی با استفاده از مجموعه داده‌های آموزش و  $\Theta_k$  رشد می‌کند و نتیجه مجموعه درخت‌های  $K$  برابر است با:

$$K = \{h_1(x), h_2(x) \dots h_k(x)\} \quad (1)$$

که در اینجا:

$$X = \{x_1, x_2 \dots x_p\} \quad (2)$$

$$h_k(x) = h(x, \Theta_k) \quad (3)$$

این بردارها یک بردار ورودی  $P$  بعدی هستند که یک جنگل را تشکیل می‌دهند. خروجی‌های  $k$  تولیدشده گروهی مربوط به هر درخت برابر است با:

$$\widehat{y}_1 = h_1(x), \widehat{y}_2 = h_2(x), \dots, \widehat{y}_k = h_k(x) \quad (4)$$

که  $y_k$  خروجی درخت  $K$ ام می‌باشد. برای به‌دست‌آوردن خروجی نهایی، متوسط همه پیش‌بینی‌های درخت‌ها محاسبه می‌شود (Chen et al., 2020).

### مدل ماشین بردار پشتیبان<sup>۱۱</sup> (SVM)

مدل ماشین بردار پشتیبان یکی از مدل‌های ماشینی یادگیری نظارت‌شده است که جهت طبقه‌بندی تفکیک گروه‌ها به‌کار می‌رود (Wang, 2005). ماشین بردار پشتیبان، الگوریتمی پیشرفته است که براساس شناسایی شرایط مختلف و تعیین توابع مختلف ناپارامتری، حداکثر تفکیک را بین داده‌های محیطی ایجاد می‌کند؛ به‌گونه‌ای که بتوان روابط مختلف میان عوامل محیطی و وقوع خطرات ایجاد نمود. این مدل قادر به ارائه ماشین‌های خطی برای تولید یک سطح بهینه، داده‌ها را با خطای کم و حداکثر فاصله میان صفحه و نزدیک‌ترین نقاط آموزشی (بردارهای پشتیبان)، تفکیک می‌نماید (Pourghasemi et al., 2013). اگر نقاط آموزشی را به‌صورت  $[x_i, y_i]$  و بردار ورودی  $X_i \in \{-1, 1\}$  تعریف کنیم و  $i = 1, \dots, l$  آن‌گاه در حالتی که داده‌ها به‌صورت خطی قابل تفکیک هستند، قواعد تصمیم‌گیری که تعریف می‌شود و توسط یک صفحه بهینه که طبقات تصمیم‌گیری باینری را تفکیک می‌کند، به‌صورت معادله زیر خواهد بود:

$$Y = \text{Sign}(\sum_{i=1}^N y_i a_i (X, X_i) + b) \quad (5)$$

که در آن،  $Y$  نتیجه معادله،  $y_i$  ارزش طبقه نمونه آموزشی است. هم‌چنین  $X$  و  $X_i$  نیز به‌ترتیب نشان‌دهنده بردار داده‌های ورودی و بردارهای پشتیبان هستند. در این معادله پارامترهای  $\alpha_i$  و  $b$  تعیین‌کننده ابر صفحه هستند. اگر

حساسیت و تشخیص را در نظر می‌گیرد که در این صورت نواقص هریک از معیارهای حساسیت و تشخیص رفع می‌شود. درحقیقت محورهای منحنی ROC براساس معیارهای حساسیت و تشخیص به دست می‌آید. هم‌چنین برای تعیین تأثیر تغییر مجموعه داده‌ها (نقاط وقوع) و آنالیز حساسیت داده‌ها، نقاط وقوع سیل‌گیری به سه گروه (S1, S2, S3) تقسیم و ارزیابی صداقت (Robustness) مدل براساس معادله زیر انجام شد (Angileri et al., 2016):

$$RAUC-ROC = AUC-ROC_{max} - AUC-ROC_{min} \quad (V)$$

RAUC-ROC صداقت و صحت مدل است که حاصل اختلاف بیش‌ترین و کم‌ترین مقدار RAUC-ROC است.

## نتایج و بحث

### اجرای مدل

به‌طورکلی، تعداد ۷۱ نقطه وقوع سیل در حوزه آبخیز مورد مطالعه ثبت شد. هفتاد درصد از این نقاط برای کالیبراسیون (آموزش مدل) و بقیه برای ارزیابی مدل‌ها به شکل تصادفی انتخاب شدند. پس از انجام مدل‌سازی مکانی، پهنه‌بندی احتمال وقوع سیلاب با استفاده از متغیرهای پیش‌بینی‌کننده و هر گروه داده آموزشی (S1, S2, S3) در مدل‌های جنگل تصادفی و مدل ماشین بردار پشتیبان اجرا شد و نقشه پهنه‌بندی برای هر گروه تهیه شد. سپس نقشه نهایی پهنه‌بندی از میانگین این سه نقشه به دست آمد. برای ساختن نقشه‌های پتانسیل سیل از روش طبقه‌بندی شکستن طبیعی<sup>۱۹</sup> در چهار کلاس کم، متوسط، زیاد و خیلی زیاد استفاده شد (Akgün & Bulut, 2007). شکل‌های (۴) و (۵) کلاس‌های استعداد خطر سیل‌گیری را برای نقشه میانگین حاصل از اجرای سه گروه داده را به ترتیب برای مدل جنگل تصادفی و مدل ماشین بردار پشتیبان نشان می‌دهد. شکل (۶) توزیع نسبی میانگین کلاس‌های احتمال وقوع سیل را برای گروه S3 نشان می‌دهد. در مورد مناطق طبقه‌بندی‌شده در

به‌صورت خطی قابل تفکیک نباشند، معادله فوق به صورت معادله (۶) تغییر می‌یابد.

$$Y = \text{Sign}(\sum_{i=1}^N y_i a_i K(X, X_i) + b) \quad (6)$$

که تابع  $K(X, X_i)$  تابع کرنلی است که برای ایجاد ماشین‌هایی با انواع مختلفی از سطوح تصمیم‌گیری غیرخطی در فضای داده‌ها، ضرب‌های داخلی تولید می‌کند (Vapnik, 1995). به‌منظور به‌کارگیری این مدل در مدل‌سازی خطرات طبیعی، ابتدا تمام لایه‌های رقومی عوامل تأثیرگذار بر وقوع هر خطر که به‌عنوان متغیرهای مستقل در مدل‌سازی انتخاب شده‌اند، به ساختار رستری (با اندازه پیکسل برابر) تبدیل می‌شود. سپس موقعیت خطرات گروه آموزش (واسنجی) که در حقیقت به‌عنوان متغیر وابسته مدنظر قرار دارند، به‌صورت ساختار رستری در محیط نرم‌افزار ArcGIS10.5 تهیه می‌شود. سپس برای اجرای مدل ماشین بردار پشتیبان، تمام لایه‌های رستری ساخته‌شده وارد نرم‌افزار R می‌شود و مدل ماشین بردار پشتیبان اجرا می‌شود.

### ارزیابی عملکرد پیش‌بینی

در این مرحله، نتایج حاصل از پهنه‌بندی احتمال وقوع خطر سیلاب، براساس نقاط وقوع گروه اعتبارسنجی (۳۰ درصد وقایع سیل گذشته) و در قالب روش منحنی مشخصه عامل گیرنده<sup>۱۷</sup> (ROC) اعتبارسنجی می‌شود و مساحت زیر این منحنی<sup>۱۸</sup> (AUC) به‌عنوان معیار کمی برای اعتبارسنجی استفاده می‌شود (Felicísimo et al., 2013). هرگاه مقدار مساحت زیر منحنی بیش‌تر از ۹۰ درصد باشد، دقت مدل عالی خواهد بود (Hosmer, 2000). اگرچه در مدل‌سازی‌های ژئومورفولوژیک بایستی از معیارهای ارزیابی مختلفی برای اعتبارسنجی دقت مدل استفاده کرد، اما با توجه با پیشینه پژوهش در این زمینه، روش منحنی مشخصه عامل گیرنده به‌عنوان جامع‌ترین روش درنظر گرفته می‌شود (Walter, 2002)؛ زیرا این روش به‌طور هم‌زمان دو معیار



پهنه‌بندی احتمال وقوع سیل با استفاده از بررسی مقایسه‌ای دو مدل شناخته‌شده جنگل تصادفی و ماشین بردار پشتیبان در شمال ایران

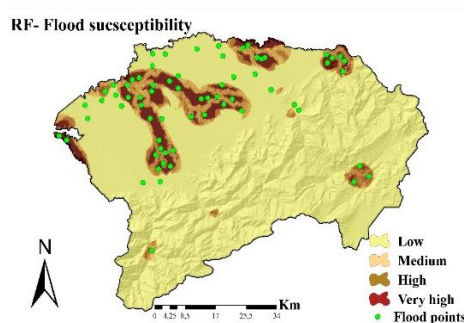
**Table 1. Statistical features of the probabilistic forecasting**

MLe models	Data set	Probability values	
		Mean	St. dev.
SVM	S1	0.241	0.235
	S2	0.251	0.237
	S3	0.250	0.237
RF	S1	0.188	0.232
	S2	0.156	0.226
	S3	0.188	0.232

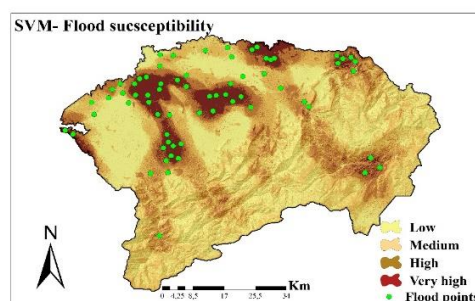
### ارزیابی نقشه‌ی پهنه‌بندی احتمال وقوع سیل

نقشه پهنه‌بندی احتمال وقوع سیل براساس منحنی مشخصه عامل گیرنده (ROC) اعتبارسنجی شد، هم‌چنین مساحت زیر این منحنی‌ها (AUC) برای هر دو مدل محاسبه شد که نتایج آن برای هر سه مجموعه داده در جدول (۲) نشان داده شده است. هم‌چنین شکل (۷) نمودار منحنی ROC را برای مجموعه S3 برای هر دو مدل نشان می‌دهد. همان‌گونه که مشاهده می‌گردد، دقت هر دو مدل بالای ۹۰ محاسبه شد. براساس منابع علمی (Hosmer, 2000) در زمینه تحلیل روش منحنی ROC و اعتبارسنجی مدل‌های پیش‌بینی مکانی می‌توان بیان کرد که مدل‌های با دقت بیش از ۹۰ درصد دارای کارایی عالی در پیش‌بینی مناطق مستعد سیلاب دارد. تطابق نقاط سیلاب‌های مشاهداتی روی پهنه‌بندی صورت گرفته در نقشه‌های خروجی (شکل‌های ۴ و ۵) تأییدکننده صحت نتایج ارزیابی بوده و نشان می‌دهد که مدل‌های مورد استفاده در تعیین محل‌های در معرض سیل موفق عمل کرده‌اند. هم‌چنین شکل (۸) نتایج آنالیز حساسیت داده‌ها (صداقت مدل) را نشان می‌دهد. مقدار ربوستنس در مرحله اعتبارسنجی برای مدل جنگل تصادفی ۰/۰۰۱ و برای مدل ماشین بردار ۰/۰۰۸ به دست آمد، که نشان‌دهنده پایداری و ربوستنس بالای مدل RF با تغییر مجموعه داده‌ها می‌باشد (Marmion et al., 2008). ارزیابی صداقت مدل در مرحله آموزش برای هر دو مدل یکسان بود.

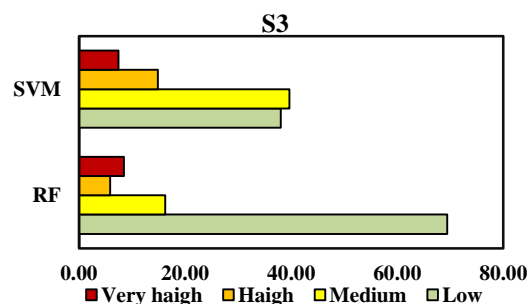
کلاس احتمال وقوع زیاد و بسیار زیاد، در مناطق با درصد شیب کم و نزدیک به رودخانه‌ها قرار گرفتند. طبق مدل RF، حدود ۱۵ درصد از منطقه موردنظر در پهنه خطر سیل زیاد و بسیار زیاد قرار گرفت. مطابق نقشه پیش‌بینی شده توسط مدل SVM نزدیک به ۲۲/۷ درصد از منطقه مورد مطالعه در کلاس‌های خطر زیاد و پتانسیل سیل بسیار زیاد قرار گرفتند. ویژگی‌های آماری پیش‌بینی احتمالی برای همه مجموعه داده در جدول (۱) نشان داده شده است.



**Figure 4. The flood-susceptibility-map (FSM) of the target area based on Random Forest model (RF)**



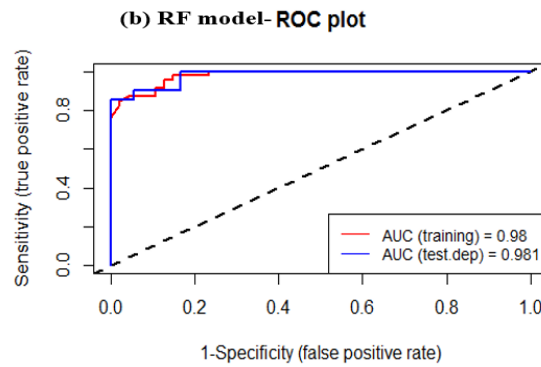
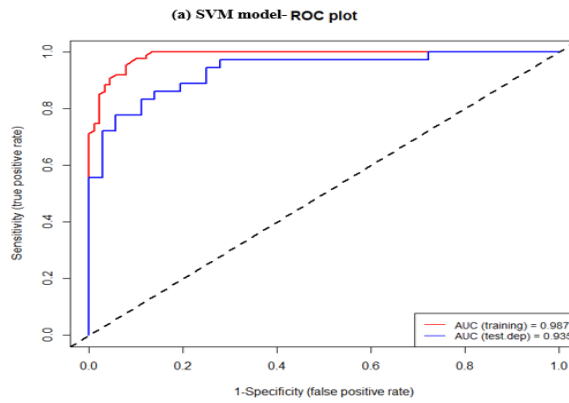
**Figure 5. The Flood-Susceptibility-Map (FSM) of the target area based on Support Vector Machine model (SVM)**



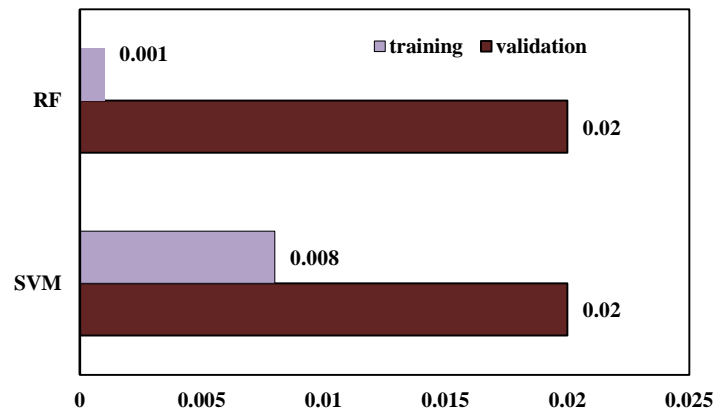
**Figure 6. Relative distributions of flood-probability - classes for RF and, SVM models based on S3**

**Table 2. Evaluation results of RF model and SVM method based on AUC -ROC**

AUC-ROC	Data set	models	
		SVM	RF
Training	s1	0.982	0.997
	s2	0.979	0.998
	s3	0.987	0.988
	Mean	0.982	0.994
Validation	s1	0.96	0.95
	s2	0.97	0.97
	s3	0.93	0.98
	Mean	0.95	0.96



**Figure 7. ROC curves of the (a) SVM model and (b) RF model for S3**



**Figure 8. Robustness of the applied models in training phase and validation step based on AUC**

## نتیجه‌گیری

در این پژوهش، کارایی مدل‌های جنگل تصادفی (RF) و ماشین بردار پشتیبان (SVM) برای شناسایی مناطق مستعد سیل، در حوزه آبخیز سالیان تپه و ثبت موقعیت مکانی ۷۱ نقطه سیل در منطقه بررسی شد و نقشه‌های پیش‌بینی مناطق مستعد خطرات سیل با استفاده از این مدل‌ها به‌دست آمد. براساس نقشه‌های خروجی، بخش‌های آبرفتی و هموار مرکزی که در مجاورت جریان‌های سطحی قرار دارند استعداد وقوع متوسط و زیادی برای ایجاد خطر سیل را به‌خود اختصاص داده‌اند. براساس نتایج به‌دست‌آمده، می‌توان بیان نمود که مدل‌های جنگل تصادفی و ماشین بردار پشتیبان دارای کارایی عالی (Hosmer, 2000) برای پیش‌بینی مناطق مستعد سیل هستند. در مطالعه‌ای که توسط Gayen et al. (2019) در مناطقی از هند با استفاده از الگوریتم‌های مختلف یادگیری ماشینی انجام شد، مدل جنگل تصادفی با دقت ۹۶/۲ درصد براساس معیار AUC دارای بهترین عملکرد و بالاترین دقت پیش‌بینی را دارا بود که نتایج حاصل از این پژوهش را تأیید می‌کند. هم‌چنین نتایج حاصل از این پژوهش با نتایج Lee & Pradhan, (2016) Albers et al. (2016) و (2007) Woznicki et al. و (2019) Chen et al. (2020) مطابقت دارد.

در پیش‌بینی و پهنه‌بندی مناطق مستعد سیلاب هستند. برای اطمینان از موفقیت و عملکرد مدل<sup>۲</sup>، انجام آنالیز حساسیت داده‌ها و ارزیابی صداقت، مدل با تعداد سه سری داده اجرا شد. در نهایت نقشه خروجی پهنه‌بندی برای هر مدل از میانگین نقشه حساسیت وقوع برای هر سه سری مجموعه داده‌ها به‌دست آمد. Angileri et al. (2016) برای تهیه نقشه حساسیت فرسایش آبی و ارزیابی صداقت مدل درخت گرادیان آماری پیشرفته<sup>۱</sup>، از چهار سری مجموعه داده استفاده کردند و در این مطالعات برای هر بار اجرای

مدل یک مجموعه متفاوت از نقاط وقوع و عدم‌وقوع (مجموعه داده‌ها)<sup>۲</sup> به تعداد مساوی و به‌صورت تصادفی انتخاب شد (Rotigliano et al., 2018).

به‌طورکلی نقشه‌های پهنه‌بندی احتمال وقوع سیل در مرحله اولیه برنامه‌ریزی شهری بسیار مهم و ضروری می‌باشند. تازگی اصلی در مطالعه حاضر این است که کدام مدل عملکرد بهتری برای پهنه‌بندی مناطق مستعد سیل دارد. برای رفع شکاف‌های مطالعه، این مدل‌های شناخته‌شده به‌عنوان بنچ‌مارک در حوزه آبریز سالیان‌تپه واقع در استان گلستان مورداستفاده قرار گرفتند. منطقه مورد مطالعه به‌دلیل تغییرات کاربری اراضی (به‌ویژه شهرنشینی) در جنگل‌ها و تغییرات آب‌وهوایی در برابر پدیده سیلاب بسیار مستعد بوده است. نتایج این پژوهش، یک راهنما و بهترین مثال برای مدیریت خطرات طبیعی و برنامه‌ریزی مکانی برای اولویت‌بندی توسعه اقتصادی آینده را ارائه می‌دهد و می‌تواند به‌عنوان ابزاری باارزش برای حفاظت از منابع طبیعی و محیط‌زیست و هم‌چنین برنامه‌ریزی پایدار برای کاربری اراضی در مناطق مستعد سیلاب استان گلستان مورداستفاده قرار گیرد و از نظر انتخاب مکان و کاربری اراضی برای سیاست‌گذاران به‌منظور برنامه‌ریزی و اجرای پروژه‌های اصلی دولت مهم خواهد بود.

## پی‌نوشت‌ها

1. Susceptibility
2. Naïve Bayes tree
3. Alternating decision tree
4. Training
5. validation
6. randomly partition algorithm
7. robustness
8. Digital Elevation Model
9. Stream Power Index
10. System for automated geoscientific analyses
11. Multi-collinearity test
12. Tolerance
13. Variance Inflation Factor (VIF)

7. Conoscenti, C., Angileri, S., Cappadonia, C., Rotigliano, E., Agnesi, V., & Märker, M. (2014). Gully erosion susceptibility assessment by means of GIS-based logistic regression: a case of Sicily (Italy). *Geomorphology*, 204, 399-411.
8. Dickie, J.A., & Parsons, A.J. (2012). Eco-geomorphological processes within grasslands, shrublands and badlands in the semi-arid Karoo, South Africa. *Land Degradation Dev.*, 23(6), 534-547.
9. Daoud, J.I. (2017). Multicollinearity and regression analysis. *J. Phy*, Conference Series (949(1), 012009). IOP Publishing.
10. Felicísimo, Á., Cuartero, A., Remondo, J., & Quirós, E. (2013). Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: a comparative study. *Landslides*, 10, 175-189.
11. Gayen, A., Pourghasemi, H.R., Saha, S., Keesstra, S., & Bai, S. (2019). Gully erosion susceptibility assessment and management of hazard-prone areas in India using different machine learning algorithms. *Science of the Total Environment*, 668, 124-138.
12. Guzzetti, F., Cardinali, M., Reichenbach, P., & Carrara, A. (2000). Comparing landslide maps: A case study in the upper Tiber River Basin, central Italy. *Environmental Management*, 25(3), 247-263.
13. Glenn, E., Morino, K., Nagler, P., Murray, R., Pearlstein, S., & Hultine, K. (2012). Roles of saltcedar (*Tamarix* spp.) and capillary rise in salinizing a non-flooding terrace on a flow-regulated desert river. *Journal of Arid Environment*, 79, 56-65.
14. Hall, A. J. (1981). Flash flood forecasting. World Meteorological Organization (WMO (Series); no. 577.), Operational hydrology report (World Meteorological Organization); 18, 48.
15. Hosmer, D.W., & Lemeshow, S. (2000). Multiple Logistic Regression. Hoboken, NJ: John Wiley & Sons, Inc. doi: 10.1002/0471722146.ch2.
16. Jafarian, Z., & Kargar, M. (2017). Comparison of Random Forest (RF) and Boosting Regression Tree (BRT) For Prediction of Dominant Plant Species Presence in Polour Rangelands, Mazandaran Province. *Iranian Journal of Applied Ecology*, 6(1), 41-55.
17. Kheyrizadeh, M., J. Maleki and H. Amonia. 2012. Flood hazard zoning using ANP model in watershed, case study: Mardaghchay Watershed. *Quantitative Geomorphological Researches*, 3(2), 39-56. (in Persian)
14. Inverse Distance Weight
15. Random forest
16. Support vector machine
17. Receiver operating characteristic
18. Area under curve
19. Natural break
20. Goodness of fit
21. Stochastic Gradient Treeboost
22. Data sets

## تعارض منافع

هیچ گونه تعارض منافع توسط نویسندگان وجود ندارد.

## منابع

1. Abdi, P. (2006). Investigation of flood potential of Zanjan River basin by SCS method and GIS. National Irrigation and Drainage Committee. Technical workshop on coexistence with floods. (In Persian)
2. Akgün, A., & Bulut, F. (2007). GIS-based landslide susceptibility for Arsin-Yomra (Trabzon, North Turkey) region. *Environment Geology*, 51(8), 1377-1387.
3. Albers, S. J., Déry, S. J., & Petticrew, E. L. (2016). Flooding in the Nechako River Basin of Canada: A random forest modeling approach to flood analysis in a regulated reservoir system. *Canadian Water Resources Journal/Revue canadienne des ressources hydriques*, 41(1-2), 250-260.
4. Angileri, S.E., Conoscenti, C., Hochschild, V., Märker, M., Rotigliano, E., & Agnesi, V. (2016). Water erosion susceptibility mapping by applying Stochastic Gradient Treeboost to the Imera Meridionale River basin (Sicily, Italy). *Geomorphology*. 262, 61-76.
5. Bui, D.T., Khosravi, K., Shahabi, H., Daggupati, P., Adamowski, J.F., Melesse, A., Pham, B.T., Pourghasemi, H.R., Mahmoodi, M., Bahrami, S., Pradhan, B., Shirzadi, A., Chapi, K., & Lee, S. (2019). Flood Spatial Modeling in Northern Iran Using Remote Sensing and GIS: A Comparison between Evidential Belief Functions and Its Ensemble with a Multivariate Logistic Regression Model. *Remote Sensing*, 11(13), 1589.
6. Chen, W., Li, Y., Xue, W., Shahabi, H., Li, S., Hong, H., & Ahmad, B.B. (2020). Modeling flood susceptibility using data-driven approaches of naïve bayes tree, alternating decision tree, and random forest methods. *Science of The Total Environment*, 701, 134-979.

18. Khosravi, K., Nohani, E., Maroufinia, E., & Pourghasemi, H.R. (2016). A GIS-based flood susceptibility assessment and its mapping in Iran: a comparison between frequency ratio and weights-of-evidence bivariate statistical models with multi-criteria decision-making technique. *Natural Hazards*, 83(2), 947-987.
19. Lee, S., & Pradhan, B. (2007). Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. *Landslides*, 4(1), 33-41.
20. Marmion, M., Hjort, J., Thuiller, W., & Luoto, M. (2008). A comparison of predictive methods in modelling the distribution of periglacial landforms in Finnish Lapland. *Earth Surface Processes and Landforms*, 33(14), 2241-2254.
21. Mojaddadi, H., Pradhan, B., Nampak, H., Ahmad, N., & Ghazali, A.H.B. (2017). Ensemble machine-learning-based geospatial approach for flood risk assessment using multi-sensor remote-sensing data and GIS. *Geomatics, Natural Hazards and Risk*, 8(2), 1080-1102.
22. Nouri Boroujerdi, P., & Eskandi, V. (2009) Introduction to Quantitative Studies in Management (Case Study: Data Mining in Management Studies). *Quarterly Journal of Quantitative Studies in Management*, 3(2) 1-13 (In Persian)
23. Poudyal, C.P., Chang, C., Oh, H.J., & lee, S. (2010). Landslide susceptibility maps comparing frequency ratio and artificial neural networks: a case study from the Nepal Himalaya. *Environmental Earth Sciences*, 61(5), 1049-1064.
24. Pourghasemi, H.R., Jirandeh, A.G., Pradhan, B., Xu, C., & Gokceoglu, C. (2013). Landslide susceptibility mapping using support vector machine and GIS at the Golestan Province, Iran. *Journal of Earth System Science*, 122(2), 349-369.
25. Pourtaghi, Z.S., & Pourghasemi, H.R. (2014). GIS-based groundwater spring potential assessment and mapping in the Birjand Township, southern Khorasan Province, Iran. *Hydrogeology Journal*, 22(3), 643-662
26. Rahi, G.h. (2018). *Prediction of trench erosion sensitivity using spatial data mining methods*. Ph.D. thesis, Faculty of Natural Resources Engineering. Sari University of Agricultural, Sciences and Natural Resources. (In Persian).
27. Rahmati, O., Pourghasemi, H. R., & Zeinivand, H. (2015). Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the Golastan Province, Iran. *Geocarto International*, 31(1), 42-70
28. Rahmati, O., Zeinivand, H., & Besharat, M. (2016a). Flood hazard zoning in Yasooj region, Iran, using GIS and multi-criteria decision analysis. *Geomatics, Natural Hazards and Risk*, 7(3), 1000-1017.
29. Rahmati, O., Pourghasemi, H. R., & Melesse, A. M. (2016b). Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: a case study at Mehran Region, Iran. *Catena*, 137, 360-372.
30. Rahmati, O., & Pourghasemi, H. R. (2017). Identification of critical flood prone areas in data-scarce and ungauged regions: A comparison of three data mining models. *Water Resources Management*, 31(5), 1473-1487
31. Rotigliano, E., Martinello, C., Agnesi, V., & Conoscenti, C. (2018). Evaluation of debris flow susceptibility in El Salvador (CA): a comparison between Multivariate Adaptive Regression Splines (MARS) and Binary Logistic Regression (BLR). *Hungarian Geogr. Bull.*, 67, 361-373.
32. Servati, M.R., Ghahrodi Tali, M., Golkarami, A., & Njafi, E. (2014). Geomorphological thresholds for gully erosion in Kchick watershed, NE Golestan Province. *Applied researches in geographical sciences*, 32, 231-249, (in Persian)
33. 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. *Journal of Hydrology*, 504, 69-79.
34. 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.
35. Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. New York, Springer-Verlag, pp. 122.
36. Wilson, J.P., & Gallant, J.C. (Eds). (2000). *Terrain analysis: principles and applications*. John Wiley and Sons.
37. Walter, S.D. (2002). Properties of the summary receiver operating characteristic (SROC) curve for diagnostic test data. *Stat Med.* 21, 1237-1256.
38. Wang, L. (2005). *Support Vector Machines: Theory and Applications*. New York, Springer-Verlag, pp.412.
39. Woznicki, S.A., Baynes, J., Panlasigui, S., Mehaffey, M., & Neale, A. (2019). Development of a spatially complete floodplain map of the conterminous United States using random forest. *Science of the total environment*, 647, 942-953.
40. Yalcin, A. (2008). GIS-based landslide susceptibility mapping using analytical hierarchy process and bivariate statistics in Ardesen (Turkey): comparisons of results and confirmations. *Catena*, 72(1), 1-12.