



## "مقاله پژوهشی"

# ارزیابی پتانسیل وقوع سیل با استفاده از مدل‌های داده کاوی ماشین بردار پشتیبان، چاید و جنگل تصادفی (مطالعه موردی: حوزه آبخیز فریزی)

مهدی زارعی<sup>۱</sup>، رحمان زندی<sup>۲</sup> و مهناز ناعمی تبار<sup>۳</sup>

۱- استادیار مرکز پژوهشی علوم جغرافیایی و مطالعات اجتماعی-دانشگاه حکیم سبزواری، سبزواری، ایران، (نویسنده مسول: m.zarei@hsu.ac.ir)

۲- استادیار دانشکده جغرافیا و علوم محیطی، دانشگاه حکیم سبزواری، سبزواری، ایران

۳- دانشجو دکتری ژئومورفولوژی، دانشکده جغرافیا و علوم محیطی، دانشگاه حکیم سبزواری، سبزواری، ایران

تاریخ دریافت: ۱۳۹۹/۱۲/۶ تاریخ پذیرش: ۱۴۰۰/۲/۱۳

صفحه: ۱۳۳ تا ۱۴۴

## چکیده مبسوط

**مقدمه و هدف:** سیل مانند سایر پدیده‌های هیدرولوژیکی یک پدیده غیر قطعی است که در هر زمان و مکان احتمال وقوع آن وجود داشته و تحت تاثیر عوامل مختلف اقلیمی، مشخصات فیزیکی حوضه، وضعیت پوشش گیاهی و کاربری اراضی و دخالت‌های انسانی است. تعیین سهم هر یک از پارامترها بر میزان بروز سیل مهم است. در حال حاضر، با ارتقای سامانه‌های اطلاعات جغرافیایی (GIS)، روش‌های سنجش از دور (RS) و ماشین یادگیری (ML) می‌توان مدل سازی بسیار دقیقی از احتمال وقوع سیلاب انجام داد. با این وجود، ساخت این مدل‌ها نیازمند دانش دقیق و اصولی از فرآیند وقوع سیل، بررسی پارامترهای موثر در شکل گیری سیل، درک چگونگی تاثیر هر پارامتر بر ایجاد سیل و انتخاب و توسعه مدل مناسب و ارزیابی آن است. با توجه به اهمیت تعیین مناطق مستعد وقوع سیل در مناطق مختلف به خصوص حوضه‌های واقع در مناطق خشک و نیمه‌خشک همچون منطقه مورد مطالعه، تحقیق حاضر به منظور ارزیابی خطرپذیری سیل با استفاده از مدل‌های داده کاوی ماشین بردار پشتیبان، چاید و جنگل تصادفی در این منطقه هدف گذاری شده است.

**مواد و روش‌ها:** در پژوهش حاضر جهت بررسی خطر وقوع سیل از مدل‌های داده کاوی ماشین بردار پشتیبان، چاید و جنگل تصادفی استفاده شده است. به طور کلی هدف از ارائه مدل‌های داده کاوی رسیدن به یک تخمین معقول و دقیق از پیش بینی مکانی وقوع سیل، مقایسه کارایی مدل‌ها و انتخاب مناسب ترین روش برای تهیه نقشه ارزیابی حساسیت سیل می‌باشد. در این پژوهش از اطلاعات مختلف همچون نقشه توپوگرافی مقیاس ۱:۵۰۰۰۰ جهت استخراج خطوط تراز، نقشه زمین شناسی مقیاس ۱:۱۰۰۰۰۰، نقشه خاک تهیه شده توسط اداره کل منابع طبیعی و آبخیزداری استان خراسان رضوی، تصویر مدل رقومی ارتفاع (DEM) با قدرت تفکیک مکانی ۱۲/۵ متر، تصاویر ماهواره‌ای گوگل ارث، داده‌های هواشناسی باران سنجی دوره آماری ۲۰ ساله (۹۸-۷۸) ایستگاه‌های اندرخ، اولنگ اسدی، سد کارده، مارشک، بلغور، گوش بالا، آل، چناران، مغان، چکنه علیا، آبقد فریزی، تلغور، قدیر آباد و کبکان استفاده شده است. طبقات ارتفاعی، شیب، جهت شیب، شبکه‌های زهکشی و آبراهه‌های اصلی، تحدب سطح زمین از روی تصویر DEM و خطوط تراز استخراج گردیدند. کاربری اراضی منطقه از روی تصاویر ماهواره‌ای گوگل ارث مربوط به سال ۲۰۲۰ و به روش طبقه بندی نظارت شده تهیه گردید. نقشه پوشش گیاهی منطقه نیز بر اساس شاخص NDVI و از روی تصاویر ماهواره‌ای لندست ۸ سال ۲۰۱۸ تهیه گردید.

**یافته‌ها:** عامل ارتفاع نقش کلیدی در کنترل جهت حرکت سیل و عمق سطح آب دارد. در ارتفاع ۲۰۰۰ متر و بیشتر با افزایش ارتفاع پتانسیل سیل خیزی در منطقه مورد مطالعه افزایش می‌یابد. بر اساس نتایج در بین کاربری‌های حوضه مورد مطالعه، اراضی آبی و باغی به علت نفوذ بیشتر، رواناب کمتری تولید نموده و کمتر مستعد سیل خیزی می‌باشند. در منطقه مطالعاتی در شیب ۶۰ درجه به دلیل افزایش شیب، زمان تأخیر حوضه کم، میزان نفوذ آب در خاک کم و در نتیجه حجم سیلاب و رواناب سطحی افزایش خواهد یافت. کلاس ۰/۱۲۰-۰/۷۴۴ بیشترین تاثیر را در وقوع سیلاب حوضه دارد. دامنه‌های شمالی، شمال غربی و غربی به دلیل دریافت بارش زیاد، ماندگاری طولانی مدت برف و داشتن رطوبت دارای پتانسیل سیل خیزی می‌باشد. در منطقه مورد مطالعه بارش بیش از ۲۵۰ میلی متر بیشترین تاثیر را در رخداد سیلاب دارا است. در منطقه مورد مطالعه خاک مزیک به دلیل نفوذپذیری نسبتاً کم باعث تولید رواناب بیشتر و ایجاد سیل می‌شود. بر اساس نتایج طبقات شاخص رطوبت توپوگرافی در منطقه مورد مطالعه کلاس ۳۵۹/۹۹-۲۶۸/۳۸ تاثیر زیادی در وقوع سیلاب داشته است. در منطقه مورد مطالعه مناطق مقعر در ایجاد سیل تاثیر زیادی دارد به این دلیل که مهمترین و موثرترین فاکتورها در وقوع سیل شیب و انحنا زمین می‌باشد. در حال حاضر قابلیت پیش بینی حساسیت سیلاب منطقه مطالعاتی با استفاده از سطح زیر منحنی مورد بررسی قرار گرفت و از مقدار AUC، درصد نرخ موفقیت و قابلیت پیش بینی هر دو در پژوهش حاضر ۱۶ سناریو با ترکیب پارامترهای مختلف به عنوان ورودی مدل ایجاد شدند. نتایج این تحقیق نشان می‌دهد که در مدل ماشین بردار پشتیبان خطی با بهترین سناریو M3 با بالاترین ضریب همبستگی ۰/۹۷۲ و کمترین مقدار MAE=۰/۵۳۸، در مدل جنگل تصادفی بهترین سناریو M10 با بالاترین ضریب همبستگی ۰/۹۶۱ و کمترین مقدار خطا MAE=۰/۶۸۵، در مدل درخت تصمیم نوع چاید بهترین سناریو M8 با بالاترین ضریب همبستگی ۰/۹۵۴ و کمترین مقدار خطا MAE=۰/۷۲۳ بوده است.

**نتیجه گیری:** به طور کلی بر اساس نتایج تحقیق حاضر طبقات با پتانسیل سیل خیزی کم و متوسط بیشتر در بخش‌های شرقی و جنوبی حوضه قرار گرفته‌اند به طوری که در بخش شرقی حوضه به علت شیب کم و نفوذپذیری مناسب خطر سیل خیزی متوسط می‌باشد. بر اساس نتایج به دلیل وجود کاربری‌های مرتعی فقیر در نیمه غربی و شمال غربی حوضه بیشترین پتانسیل سیل خیزی مشاهده شده قرار دارند. همچنین نتایج نشان داد قسمت‌های شمالی و غربی حوضه که از نظر زمین شناسی و لیتولوژی دارای سازندهای سطحی مانند مارن، رس و سیلت بوده و ضریب نفوذپذیری آن‌ها بسیار کم و پوشش گیاهی ناچیز دارند، از پتانسیل بالایی در بروز سیلاب برخوردار هستند. در این پژوهش ارزیابی مدل‌ها با استفاده از ضریب همبستگی (R) و مقادیر میانگین قدر مطلق خطا (MAE) صورت گرفت. بررسی نتایج مدل‌ها نشان داد که به ترتیب مدل‌های ماشین بردار پشتیبان، چاید و جنگل تصادفی با سناریو M3، M8 و M10 با بیشترین میزان همبستگی و کمترین خطای میانگین، از دقت بالایی در برآورد خطر وقوع سیل در منطقه مطالعاتی برخوردار هستند. علاوه بر این، به منظور ارزیابی مدل‌های ارائه شده از مساحت زیر منحنی (ROC) استفاده شد. بر این اساس این مقادیر هم در داده‌های تعلیمی و هم در داده‌های آموزشی در الگوریتم (SVM) و مدل الگوریتم نوین جنگل تصادفی دارای نتایج دقیق تری می‌باشد. این نتیجه بیانگر این مطلب است که هر دو مدل از نظر صحت و اعتبار مدل سازی مورد تأیید قرار گرفته‌اند.

**واژه‌های کلیدی:** پهنه بندی، حوزه آبخیز فریزی، سیل، مدل‌های داده کاوی، منحنی ROC

## مقدمه

ارتفاع مجاور را غرقاب نماید سیل گفته می‌شود (۱۲). علت اصلی سیلاب‌ها بارش‌های سنگین و طولانی مدتی است که منجر به وقوع رواناب بالا در رودخانه‌ها یا مناطق هموار

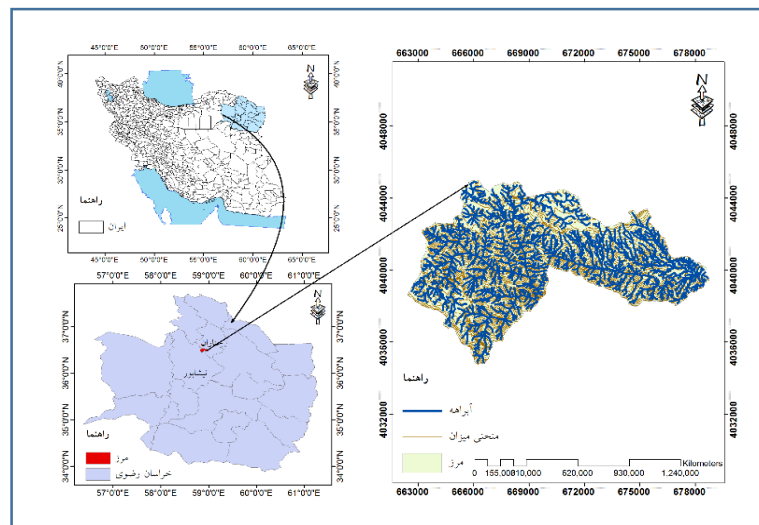
به هر گونه دبی جریانی که بیش از ظرفیت حمل طبیعی بستر رودخانه بوده، از کناره‌های آن سرریز و اراضی کم

می‌باشد. شهاب و همکاران (۳۵) در حوضه آبخیز کلوان در همدان با مدل‌های الگوریتم یادگیری ماشین (ML)، درخت رگرسیون تقویت شده (BRT)، جنگل تصادفی (RF)، جنگل تصادفی موازی (PRF)، جنگل تصادفی منظم (RRF) و درختان تصادفی (ERT) به این نتیجه رسیدند که مدل ERT با مقدار AUC ۰/۸۲ مناسب ترین مدل است و مقدار AUC برای مدل‌های PRF، RF و BRT به ترتیب ۰/۷۹، ۰/۷۸ و ۰/۷۵ است. موبلی و همکاران (۳۸) در جنوب شرق تگزاس با استفاده از مدل‌های جنگل تصادفی و FEMA پرداختند. طبق بررسی صورت گرفته نتایج نشان‌دهنده آن است که مدل جنگل تصادفی با حساسیت بالا ۰/۸۹ AUC در مقایسه با مدل FEMA، تقریباً ۶۴۹۰۰۰ سازه با حداقل ۱ درصد احتمال سالانه جاری شدن سیل را شناسایی می‌کند که سه برابر بیشتر از آنچه در حال حاضر توسط مدل FEMA به عنوان مناطق مستعد سیل شناسایی شده است. توفیکول و همکاران (۴) در حوضه‌ای واقع در کشور بنگلادش با استفاده از دو مدل گروه ترکیبی جدید، Dagging، Random Subspace (RS)، Random مصنوعی (ANN)، جنگل تصادفی (RF) و ماشین بردار پشتیبان (SVM) نشان دادند که مقدار سطح زیر منحنی AUC برای همه مدل‌ها بالاتر از ۰/۸۰ بوده است. برای مدل حساسیت به سیل، مدل Dagging برتر عمل می‌کند و پس از آن مدل‌های RF، ANN، SVM و RS دارای عملکرد خوبی است. با توجه به اهمیت تعیین مناطق مستعد وقوع سیل در مناطق مختلف به خصوص حوضه‌های واقع در مناطق خشک و نیمه‌خشک همچون منطقه مورد مطالعه، تحقیق حاضر به منظور ارزیابی خطرپذیری سیل با استفاده از مدل‌های داده کاوی ماشین بردار پشتیبان، چاید و جنگل تصادفی در این منطقه هدف‌گذاری شده است.

#### منطقه مورد مطالعه

آبخیز فریزی، جزئی از آبخیز کویر مرکزی است که در جنوب غربی شهرستان چناران و شرق نیشابور و در موقعیت طول‌های جغرافیایی  $46^{\circ} 48' 58''$  تا  $45^{\circ} 59' 58''$  و عرض‌های جغرافیایی  $36^{\circ} 26' 41''$  تا  $36^{\circ} 32' 10''$  قرار گرفته است. حوضه مورد مطالعه  $80/572$  کیلومتر مربع وسعت داشته و بر اساس اقلیم‌نمای آمبرژه دارای اقلیم نیمه خشک و نیمه مرطوب می‌باشد (۳۹). در نقشه شماره ۱ موقعیت حوضه آبخیز نسبت به شهرستان‌های استان خراسان رضوی نشان داده شده است.

می‌شوند (۱۲). آنچه مسلم است وقوع سیلاب‌ها از یک سو با شدت بارندگی، وقوع طوفان‌ها و ویژگی‌های هیدرولوژیکی حوضه (درصد اشباع خاک، درصد نفوذپذیری، سرعت ذوب برف و میزان فرسایش) و از سوی دیگر با فعالیت‌های انسانی و تغییرات اقلیم ناشی از آن رابطه مستقیم دارد (۳۴). سیل مانند سایر پدیده‌های هیدرولوژیکی یک پدیده غیر قطعی است که در هر زمان و مکان احتمال وقوع آن وجود داشته و تحت تاثیر عوامل مختلف اقلیمی، مشخصات فیزیکی حوضه، وضعیت پوشش گیاهی و کاربری اراضی و دخالت‌های انسانی است (۳۱). تعیین سهم هر یک از پارامترها بر میزان بروز سیل مهم است. در حال حاضر، با ارتقای سامانه‌های اطلاعات جغرافیایی (GIS)، روش‌های سنجش از دور (RS) و ماشین یادگیری (ML) می‌توان مدل سازی بسیار دقیقی از احتمال وقوع سیلاب انجام داد. با این وجود، ساخت این مدل‌ها نیازمند دانش دقیق و اصولی از فرآیند وقوع سیل، بررسی پارامترهای موثر در شکل‌گیری سیل، درک چگونگی تاثیر هر پارامتر بر ایجاد سیل و انتخاب و توسعه مدل مناسب و ارزیابی آن است (۲۱). هرچند کنترل کامل خسارات ناشی از سیلاب امری غیرممکن است اما می‌توان با تحلیل‌های مناسب و انجام اقدامات پیشگیری تا حد زیادی از اثرات مخرب آن جلوگیری نمود که یکی از این اقدامات تهیه نقشه پتانسیل سیل‌خیزی است (۲۳). مطالعات زیادی در زمینه پهنه‌بندی سیلاب در مناطق مختلف انجام شده به عنوان مثال، راستگو و همکاران (۳۲) طی مطالعه‌ای به پهنه‌بندی سیلاب با استفاده از الگوریتم ماشین بردار پشتیبان در شهر جناح پرداختند، نتایج ایشان حاکی از عملکرد خوب مدل در ارزیابی وقوع سیل می‌باشد. در تحقیق دیگری شریفی گرم دره و همکاران (۳۶) به ارزیابی کارایی سامانه‌های ماشین بردار پشتیبان و شبکه عصبی مصنوعی در حوضه آبخیز دریاچه نمک پرداختند و نشان دادند که روش مدل‌سازی ماشین بردار پشتیبان نسبت به روش شبکه عصبی مصنوعی از دقت بیشتری در پیش‌بینی دبی‌های حداکثر لحظه‌ای برخوردار است. آوند و همکاران (۹) با استفاده از مدل‌های یادگیری ماشین جنگل تصادفی و خطی تعمیم یافته بیزین در حوضه آبخیز تجن به این نتیجه رسیدند که سه فاکتور ارتفاع، فاصله از رودخانه و شیب بیشترین تاثیر را در سیل‌گیری منطقه مورد مطالعه دارند. همچنین نتایج ارزیابی خروجی مدل‌ها نشان‌دهنده برتری مدل RF و دقت بیشتر این مدل در تهیه نقشه حساسیت به وقوع سیل در منطقه مورد مطالعه



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

## مواد و روش‌ها

در پژوهش حاضر جهت بررسی خطر وقوع سیل از مدل‌های داده‌کاوی ماشین بردار پشتیبان، چاید و جنگل تصادفی استفاده شده است. به طور کلی هدف از ارائه مدل‌های داده کاوی رسیدن به یک تخمین معقول و دقیق از پیش‌بینی مکانی وقوع سیل، مقایسه کارایی مدل‌ها و انتخاب مناسب ترین روش برای تهیه نقشه ارزیابی حساسیت سیل می‌باشد. بنابراین با توجه به دقت آماری بالای مدل‌های مذکور در این مطالعه به مقایسه تطبیقی و ارزیابی کارایی آنها در بررسی وقوع سیل و ارزیابی دقت آنها در پهنه‌بندی سیلاب در منطقه مورد مطالعه پرداخته شد. در این پژوهش از اطلاعات مختلف همچون نقشه توپوگرافی مقیاس ۱:۵۰۰۰۰ جهت استخراج خطوط تراز، نقشه زمین شناسی مقیاس ۱:۱۰۰۰۰۰، نقشه خاک تهیه شده توسط اداره کل منابع طبیعی و آبخیزداری استان خراسان رضوی، تصویر مدل رقومی ارتفاع (DEM) با قدرت تفکیک مکانی ۱۲/۵ متر، تصاویر ماهواره‌ای گوگل ارث، داده‌های هواشناسی باران سنجی دوره آماری ۲۰ ساله (۹۸-۷۸) ایستگاه‌های اندرخ، اولنگ اسدی، سد کارده، مارشک، بلغور، گوش بالا، آل، چناران، مغان، چکنه علیا، آبد فریزی، تلغور، قدیر آباد و کبکان استفاده شده است. طبقات ارتفاعی، شیب، جهت شیب، شبکه‌های زهکشی و آبراهه‌های اصلی، تحذب سطح زمین از روی تصویر DEM و خطوط تراز استخراج گردیدند. کاربری اراضی منطقه از روی تصاویر ماهواره‌ای گوگل ارث مربوط به سال ۲۰۲۰ و به روش طبقه‌بندی نظارت شده تهیه گردید. نقشه پوشش گیاهی منطقه نیز بر اساس شاخص NDVI و از روی تصاویر ماهواره‌ای لندست ۸ سال ۲۰۱۸ تهیه گردید. شاخص توان آبراهه (SPI) و شاخص رطوبت توپوگرافی (TWI) از رابطه ۱ و ۲ محاسبه می‌شود. لایه تراکم زهکشی بر مبنای لایه آبراهه‌ها استخراج و لایه واحد اراضی منطقه، از اداره کل منابع طبیعی و آبخیزداری استان خراسان رضوی تهیه شد.

$$SPI = A_s \times \tan \beta \quad (1)$$

$A_s$ : مساحت ویژه حوضه بر حسب متر مربع بر متر و  $\beta$ :

زاویه شیب به درجه است.

$$TWI = (A_s / \tan(\beta)) \quad (2)$$

$A_s$ : مساحت ویژه بالا دست یک پیکسل (متر مربع به متر)،

$\beta$ : شیب پیکسل به رادیان است

ماشین بردار پشتیبان (SVM)

ساختار مدل ماشین بردار از اصل استقرایی کمینه‌سازی خطای ساختاری بهره می‌جوید (۳۵). ماشین بردار پشتیبان یک سیستم یادگیری کارآمد بر مبنای تئوری بهینه‌سازی مقید است که از اصول استقرایی کمینه‌سازی خطای ساختاری استفاده و منجر به یک جواب بهینه کلی می‌شود. در یک مدل رگرسیونی SVM لازم است وابستگی تابع متغیر وابسته ( $y$ ) به مجموعه‌ای از متغیرهای مستقل ( $x$ ) تخمین زده شود.

$$y = f(x) + \text{noise} \quad (3)$$

در نهایت، تابع تصمیم رگرسیون بردار پشتیبان غیرخطی به صورت رابطه ۴ خواهد بود که کنترل کننده میزان نوسان تابع گوسی همچنین کنترل کننده نتایج پیش‌بینی و تعمیم دهنده مدل SVM است.

$$f(x_i) = \sum_{i=1}^k (-\partial_i - \partial_i^*) k(X_i, X_j) + b \quad (4)$$

سپس ۶۴ نقطه به صورت وقوع سیلاب پس از بررسی سیلاب‌های اخیر شناسایی و برداشت شد. به طور تصادفی، ۷۰ درصد نقاط برای واسنجی، ۳۰ درصد برای ارزیابی تقسیم شد. سپس، احتمال وقوع سیل با استفاده از مدل نامبرده شده مدل‌سازی شد. پس از آن، مدل با استفاده از نقاط سیلی که در اجرا و واسنجی مدل شرکت داده نشده، توسط معیارهای مختلف ارزیابی همچون منحنی مشخصه عملکرد سیستم ROC اعتبار سنجی شد

## مدل جنگل تصادفی

مدل جنگل تصادفی برای بسیاری از مجموعه داده‌ها، دسته‌بندی را با سرعت بالایی انجام می‌دهد. این الگوریتم از جمله دسته‌بندی‌هایی است که روش Bagging را به کار

به مناطق پایین‌تر جریان می‌یابد و در مناطق مرتفع سیلاب سریع‌تر از مناطق کم ارتفاع و مسطح جریان پیدا می‌کند (۱۳). در ارتفاع ۲۰۰۰ متر و بیشتر با افزایش ارتفاع پتانسیل سیل‌خیزی در منطقه مورد مطالعه افزایش می‌یابد. تغییرات کاربری اراضی هنگامی که با تغییرات آب و هوایی همراه باشد، می‌تواند منجر به افزایش خطر سیل شود (۱۹). بر اساس نتایج در بین کاربری‌های حوضه مورد مطالعه، اراضی آبی و باغی به علت نفوذ بیشتر، رواناب کمتری تولید نموده و کمتر مستعد سیل‌خیزی می‌باشند. ولی کاربری مرتع به دلیل پوشش گیاهی فقیر و نفوذپذیری کمتر، رواناب بیشتری تولید نموده و از لحاظ سیل‌خیزی مستعدتر می‌باشد. شیب اراضی، تعیین کننده نحوه و میزان جریان و سرعت حرکت آب می‌باشد. به این ترتیب که در اراضی با شیب شدید، تحت تأثیر نیروی ثقل آب دارای سرعت زیادی می‌شود (۸). در منطقه مطالعاتی در شیب ۶۰ درجه به دلیل افزایش شیب، زمان تأخیر حوضه کم، میزان نفوذ آب در خاک کم و در نتیجه حجم سیلاب و رواناب سطحی افزایش خواهد یافت. اراضی مخلوط و اراضی متفرقه در قسمت‌های کم ارتفاع‌تر حوضه با ارتفاعی کمتر از ۲۰۰۰ متر و اغلب در دامنه‌هایی با جهت جنوبی و یا متمایل به آن قرار گرفته است. الگوی پراکنش گونه‌های غالب و فراوان در اراضی این تیپ نامنظم می‌باشد و از لحاظ نفوذپذیری و ایجاد رواناب مستعد ایجاد سیلاب می‌باشد. تراکم زهکشی به طور موثری به لیتولوژی، شیب، توپوگرافی و شرایط اقلیمی مربوط می‌باشد و نقش اساسی در زمان تمرکز و اوج سیلاب دارد (۳). کلاس ۰/۱۲۰-۰/۰۷۴ بیشترین تأثیر را در وقوع سیلاب حوضه دارد. سنگ‌شناسی منطقه و خاک حاصل از آن تا حدود زیادی تعیین کننده شدت و ظرفیت نفوذپذیری بوده و رواناب را تحت تأثیر قرار می‌دهد. رسوبات آبرفتی رودخانه‌ای به دلیل میزان نفوذپذیری کمتر رواناب زیادی تولید کرده و از لحاظ سیل‌خیزی مستعدتر می‌باشد. رواناب، مبنایی برای اهداف کنترل رواناب در مناطق بالا دست می‌باشد (۱). در منطقه مورد مطالعه مناطقی با ضریب رواناب بالا پتانسیل سیل‌خیزی بالاتری دارند. نقشه نفوذپذیری حوضه با توجه به ویژگی خاک‌شناسی و زمین‌شناسی تهیه شد. مناطقی که دارای نفوذپذیری زیاد است منطبق بر سازندهای آبرفتی، آهکی، آهک رسی، شیل آهکی و مارن می‌باشد که عموماً فاقد پوشش خاکی و دارای خاک خیلی کم عمق است که در محدوده ارتفاعی ۲۰۰۰ متر و دارای شیب بیش از ۴۰ درجه و در جهات مختلف می‌باشند. درجه شیب یک پارامتر کنترل کننده ضروری در ارزیابی پایداری شیب است. به‌طور کلی با افزایش زاویه شیب، خطر سیل نیز افزایش می‌یابد و انتظار می‌رود میزان وقوع سیل در دامنه‌های پرشیب بیشتر از دامنه‌های با شیب ملایم باشد (۳۲). دامنه‌های شمالی، شمال غربی و غربی به دلیل دریافت بارش زیاد، ماندگاری طولانی مدت برف و داشتن رطوبت دارای پتانسیل سیل‌خیزی می‌باشد. آبراه‌ها به عنوان بستر عبور آب‌های ناشی از بارندگی فاکتور مهمی هستند که شناسایی تراکم و میزان فاصله آن‌ها در پیش‌بینی و کاهش خسارات سیلاب‌ها موثر است (۳۴). در منطقه مطالعاتی کلاس کمتر از ۴۰۰ متر

می‌گیرد. این روش ترکیبی از چندین درخت تصمیم است که در ساخت آن چندین نمونه بوت استرپ از داده‌ها شرکت دارند و در ساخت هر درخت به طور تصادفی تعدادی از متغیرهای ورودی شرکت می‌کنند. با استفاده از روش بوت‌استرپ تعداد زیادی نمونه‌های  $n$  تایی از مجموعه داده‌های مشاهداتی اولیه، نمونه‌برداری می‌شوند. هدف از اجرای بوت‌استرپ، پیدا کردن خطای (واریانس) برآوردگر با استفاده از تکرار مراحل نمونه‌گیری و برآوردیابی است. در هر بار تکرار بر اساس یک عمل باز نمونه‌گیری از داده‌ها، برآوردگر مورد نظر محاسبه می‌شود. در طی فرآیند ساخت درخت در هر شاخه، از بین تمام  $M$  متغیر مستقل به صورت تصادفی  $m$  متغیر برای تقسیم شدن انتخاب می‌شود. برای حالت رگرسیونی نسبت  $\frac{m}{M}$  برابر با یک سوم است و برای کلاسه‌بندی برابر با  $m = \sqrt{M}$  پیشنهاد شده است. پس از ساخت تمام درخت داده‌های تست به درخت معرفی شده و به تعداد درخت‌ها برای بردار ورودی یک خروجی به دست می‌آید. با میانگین‌گیری این خروجی‌ها، خروجی نهایی مدل و با در نظر گرفتن توزیع تجربی خروجی‌ها مقادیر صدک‌ها و دامنه عدم قطعیت محاسبه می‌شود (۲۷). محاسبات مدل جنگل تصادفی در نرم‌افزار Matlab2018 انجام شد. به منظور انتخاب مناسب‌ترین خط برازش یافته توسط این روش در احتمالات متفاوت، متغیر اندازه گره که نشان‌دهنده تعداد برگ‌ها در هر شاخه است با آزمون و خطا تعیین شد.

#### الگوریتم CHAID

چاید مخفف مجذور مربع شناسایی کننده تعاملات خودکار است. این روش آماری بسیار موثر برای بخش‌بندی و رشد درخت، به وسیله کلاس توسعه یافته است. این الگوریتم مقادیری را که به صورت آماری همگن در نظر گرفته شده‌اند (شباهت) با توجه به متغیر هدف و حفظ تمام مقادیری که ناهمگن (غیر مشابه) هستند، ادغام می‌کند. آزمون آماری مورد استفاده بستگی به سطح اندازه‌گیری میدان هدف دارد. اگر صفت هدف پیوسته باشد، یک آزمون  $F$  صورت می‌گیرد. اگر صفت هدف دسته‌ای باشد، یک آزمون مجذور مربع به کار برده می‌شود. چاید یک روش باینری نیست و می‌تواند بیش از دو دسته در هر سطح خاص از درخت ایجاد کند. بنابراین گرایش به ساخت درخت نسبت به روش رشد باینری گسترده‌تر است. این الگوریتم برای تمام انواع متغیرها پاسخگو است و دو متغیر وزنی و فراوانی را می‌پذیرد. مدل مذکور همچنین مقادیر مفقودی را با یک طبقه بندی معتبر ایجاد می‌کند (۱۸).

#### نتایج و بحث

شکل‌گیری پدیده سیل در هر منطقه بستگی به تعدادی از عوامل مختلف هیدرولوژیکی و ژئومورفیکی موثر در وقوع سیل دارد. این عوامل با توجه به خصوصیات حوضه‌ها می‌تواند تغییر کند، این نکته بسیار حائز اهمیت است که در این میان باید معقول‌ترین و مهم‌ترین پارامترها انتخاب شوند. عامل ارتفاع نقش کلیدی در کنترل جهت حرکت سیل و عمق سطح آب دارد. جریان سیلاب به طور مستمر از مناطق بالاتر

بارش با نتایج نگارش و همکاران (۲۵)، انحنای زمین با نتایج نوری و همکاران (۲۶) مطابقت دارد.

#### متغیرهای ورودی به مدل‌ها

برای انتخاب مناسب‌ترین مدل بر اساس معیار خطا، از روش رتبه‌بندی برای داده‌های تست استفاده می‌گردد. بدین معنا که به هر یک از مدل‌ها بر اساس معیارهای ذکر شده رتبه‌ای داده می‌شود و در پایان پس از مقایسه مجموع رتبه‌های به دست آمده، مناسب‌ترین مدل برای تخمین بهترین معادله نفوذ انتخاب می‌شود. شیوه رتبه‌بندی این مدل‌ها بدین صورت است که در هر روش، مدلی که کمترین میزان خطای MAE را دارد، پایین‌ترین رتبه، یعنی رتبه ۱ و به مدل‌ها یا ترکیبات ورودی دیگر در هر یک از دسته‌های ورودی بر حسب میزان معیار خطاهای مذکور، رتبه‌های ۲ الی آخر تعلق می‌گیرد. اما در مورد معیارهای ضریب همبستگی (R) مدل یا ترکیبی که بیشترین مقدار این معیارها را دارا است، رتبه ۱ اختصاص می‌یابد. همچنین به مدل‌هایی که میزان معیار خطای آن‌ها یکسان باشند، رتبه یکسان تعلق می‌گیرد. در نهایت پس از انتخاب مدل‌ها یا ترکیبات مناسب هر روش و مقایسه بین آنها، روش برتر در منطقه مورد مطالعه انتخاب می‌گردد. بررسی مقایسه مدل‌های ماشین بردار پشتیبان، چایلد و جنگل تصادفی نشان می‌دهد که می‌توان یک الگوی مشخصی برای منطقه مطالعاتی را با دقت مناسب برای برآورد سیلاب معرفی نمود و باید برای هر مدل، الگوی برتر را معرفی کرد. در نهایت پس از انتخاب مدل‌ها یا ترکیبات مناسب هر روش و مقایسه بین آنها، روش برتر انتخاب می‌گردد. جدول ۱ ضریب همبستگی و خطای سنایوهای مختلف در چهار مدل را نشان می‌دهد. همچنین در این تحقیق تحلیل حساسیت به منظور شناسایی و تعیین درجه اهمیت هر یک از پارامترها در سری داده‌های ورودی بر عملکرد دینامیکی سیستم صورت گرفته است. به منظور آنالیز حساسیت در هر سری داده ورودی یک پارامتر حذف و مدل با همان داده‌های تست، اجرا و مقادیر ضریب همبستگی و خطا تعیین گردید. با حذف هر پارامتر هر چه خطای مدل افزایش بیشتری نشان دهد، بیانگر تأثیر بیشتر آن پارامتر بر فرآیند شبیه‌سازی می‌باشد. جهت انجام تحلیل حساسیت برای سری داده‌های ورودی از مدل‌های RF، linear SVM و CHAID به دلیل دقیق‌تر بودن نتایج آن استفاده شده است.

بیشترین تأثیر را در وقوع سیلاب دارد. بارندگی را می‌توان مهم‌ترین عاملی دانست که به طور مستقیم در چرخه هیدرولوژی دخالت دارد (۵). در منطقه مورد مطالعه بارش بیش از ۲۵۰ میلی متر بیشترین تأثیر را در رخداد سیلاب دارا است. ویژگی‌های پوشش گیاهی و خاک می‌توانند بر واکنش سیلاب در حوضه زهکشی تأثیر داشته باشند. در منطقه مورد مطالعه خاک مزیک<sup>۱</sup> به دلیل نفوذپذیری نسبتاً کم باعث تولید رواناب بیشتر و ایجاد سیل می‌شود. شاخص رطوبت توپوگرافی یکی از شاخص‌های تأثیرگذار بر وقوع و پتانسیل سیل در حوضه‌های آبخیز به شمار می‌رود (۲۹). بر اساس نتایج طبقات شاخص رطوبت توپوگرافی در منطقه مورد مطالعه کلاس ۲۶۸/۳۸-۳۵۹/۹۹ تأثیر زیادی در وقوع سیلاب داشته است. شاخص پوشش گیاهی NVDI که هرچه به سمت منفی تمایل داشته باشد تأثیر زیادی بر رخداد سیل دارد زیرا شامل مناطقی است که از لحاظ پوشش گیاهی ضعیف می‌باشد و باعث کم شدن نفوذ آب و جاری شدن رواناب می‌شود. در منطقه مورد مطالعه مناطق مقعر در ایجاد سیل تأثیر زیادی دارد به این دلیل که مهم‌ترین و موثرترین فاکتورها در وقوع سیل شیب و انحنای زمین می‌باشد. سیل در مناطقی رخ می‌دهد که دارای شیب زیاد و از نظر انحنای زمین مقعر باشند. توان آبراهه بیانگر توان فرساینده‌گری جریان بوده که با درجه شیب و مساحت حوزه آبخیز رابطه مستقیم دارد. بنابراین وقتی سرعت جریان سطحی افزایش یابد در نتیجه شاخص آبراهه افزایش می‌یابد (۲۸). هر چه شاخص توان آبراهه بیشتر باشد سیل خیزی افزایش می‌یابد. در رابطه با عامل ارتفاع نتایج پژوهش با نتایج فیضی کوشکی و همکاران (۱۱)، کاربری اراضی با نتایج حسینی و همکاران (۱۵)، عامل شیب با نتایج اصغری و همکاران (۳)، عامل واحد اراضی با نتایج رضوی‌زاده و همکاران (۳۳)، عامل تراکم زهکشی با نتایج اصغری و همکاران (۷)، شاخص پوشش گیاهی NDVI با نتایج جهان تیغ و همکاران (۱۶)، نفوذپذیری با نتایج زهتابیان و همکاران (۴۰)، فاصله از آبراهه با نتایج عرب عامری و همکاران (۶)، شاخص رطوبت توپوگرافی با نتایج یوسف‌زاده و همکاران (۳۹)، شاخص توان آبراهه با نتایج معروفی نیا و همکاران (۲۰)، زمین شناسی با نتایج فیض‌نیا و همکاران (۱۰)، رواناب با نتایج مصفايي و همکاران (۲۲)، خاک با نتایج جوادی و همکاران (۱۷)، جهت شیب با نتایج قنبرزاده و همکاران (۳۰)،

جدول ۱- ضریب همبستگی و خطای سناریوهای مختلف در مدل ماشین بردار پشتیبان خطی، درخت تصمیم نوع چاید و جنگل تصادفی در ایستگاه

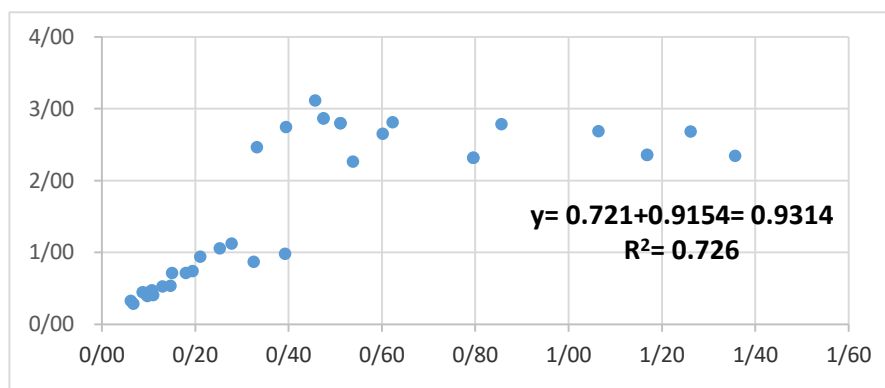
Table 1. Correlation coefficient and error of different scenarios in linear support vector machine model, chaid type decision tree and random forest at the station

| سناریو | SVM-Linear |       | Chaid |       | Random Forest |       |
|--------|------------|-------|-------|-------|---------------|-------|
|        | R          | MAE   | R     | MAE   | R             | MAE   |
| M1     | ۰/۹۳۱      | ۲/۱۵۵ | ۰/۸۵۱ | ۱/۱۱۳ | ۰/۸۵۴         | ۲/۱۲۳ |
| M2     | ۰/۹۱۱      | ۱/۳۹۹ | ۰/۸۶۳ | ۱/۶۳۱ | ۰/۶۲۳         | ۳/۵۶۱ |
| M3     | ۰/۹۷۲      | ۰/۵۳۸ | ۰/۹۲۱ | ۱/۰۵۱ | ۰/۷۲۴         | ۱/۵۵۵ |
| M4     | ۰/۸۰۳      | ۱/۴۷۱ | ۰/۹۱۲ | ۱/۲۰۱ | ۰/۸۹۱         | ۱/۶۱۷ |
| M5     | ۰/۸۵۴      | ۱/۳۲۴ | ۰/۸۹۱ | ۱/۳۶۱ | ۰/۷۲۱         | ۱/۳۰۲ |
| M6     | ۰/۹۰۳      | ۱/۶۵۶ | ۰/۷۶۸ | ۲/۲۱۱ | ۰/۸۷۶         | ۱/۶۵۴ |
| M7     | ۰/۷۶۵      | ۰/۹۲۳ | ۰/۹۲۳ | ۱/۶۲۳ | ۰/۸۹۸         | ۱/۲۱۳ |
| M8     | ۰/۹۱۵      | ۱/۳۲۱ | ۰/۹۵۴ | ۰/۷۲۳ | ۰/۷۶۱         | ۱/۴۲۳ |
| M9     | ۰/۸۹۱      | ۱/۷۴۱ | ۰/۷۸۵ | ۲/۳۵۳ | ۰/۸۹۹         | ۱/۱۳۶ |
| M10    | ۰/۹۱۱      | ۰/۸۲۳ | ۰/۸۰۵ | ۲/۸۲۱ | ۰/۹۶۱         | ۰/۶۸۵ |
| M11    | ۰/۹۰۵      | ۱/۴۴۲ | ۰/۹۱۱ | ۱/۸۷۱ | ۰/۷۵۵         | ۱/۷۵۸ |
| M12    | ۰/۸۲۱      | ۱/۳۴۱ | ۰/۸۶۳ | ۱/۵۶۳ | ۰/۸۴۴         | ۱/۳۶۷ |
| M13    | ۰/۸۷۸      | ۰/۹۲۱ | ۰/۸۹۷ | ۱/۳۱۰ | ۰/۸۴۷         | ۱/۳۷۵ |
| M14    | ۰/۹۱۲      | ۱/۱۳۳ | ۰/۹۳۱ | ۰/۹۶۳ | ۰/۷۶۳         | ۱/۸۷۱ |
| M15    | ۰/۹۲۱      | ۰/۸۵۴ | ۰/۸۹۶ | ۱/۶۱۲ | ۰/۶۹۸         | ۱/۴۶۹ |
| M16    | ۰/۸۹۹      | ۱/۳۱۲ | ۰/۷۹۷ | ۱/۵۱۴ | ۰/۷۹۱         | ۱/۷۸۹ |

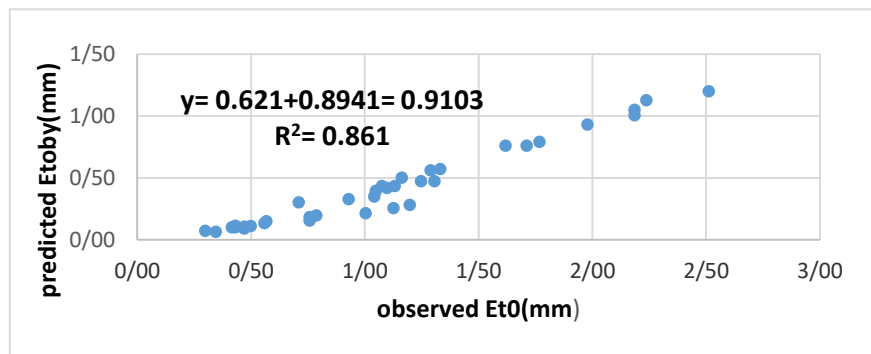
بردار پشتیبان به عنوان بهترین الگوها با بیشترین ضریب همبستگی و کمترین خطا به عنوان الگوی برتر شناخته شدند. نتایج آنالیز حساسیت‌ها نشان می‌دهد در مدل ماشین بردار پشتیبان خطی به ترتیب پارامترهای ارتفاع، شیب، لیتولوژی، خاک و در مدل جنگل تصادفی به ترتیب پارامترهای کاربری اراضی، جهت شیب، رطوبت توپوگرافی و تراکم زهکشی به عنوان پارامترهای تأثیرگذار در مدل شناخته شدند. در مدل درخت تصمیم نوع چاید پارامترهای رواناب، خاک، فاصله از آبراهه و بارش به عنوان مهمترین پارامترها، اهمیت نسبی بالایی در برآورد سیلاب در منطقه مورد مطالعه دارند. شکل‌های (۲، ۳، ۴) پراکنش داده‌های مشاهداتی و برآورد شده توسط بهترین الگو از هر مدل نشان داده شده است. علاوه بر این، به منظور ارزیابی مدل‌های ارائه شده از مساحت زیر منحنی (ROC) استفاده شد (شکل ۵، ۶، ۷). الگوریتم نوین جنگل تصادفی کارایی بهتری در پهنه‌بندی حساسیت سیلاب در حوضه مورد مطالعه داشته‌اند. خروجی‌های مدل‌ها در جداول (۲ و ۳) ارائه شده است. سپس نقشه‌های پهنه بندی خطر سیلاب منطقه مطالعاتی با سه مدل نام برده تهیه شدند (شکل‌های ۸، ۹، ۱۰).

### اعتبار سنجی مدل‌ها

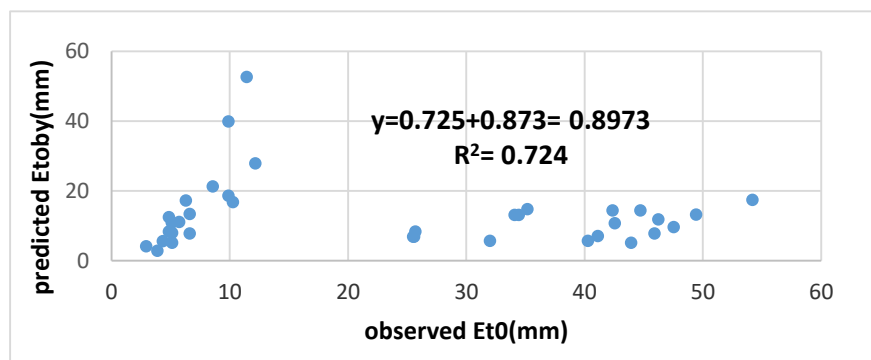
یک استراتژی مهم در مدل‌های پیش‌بینی اعتبارسنجی نتایج است که می‌تواند تفسیر معناداری نتایج را ارائه دهد. برای اعمال این اعتبارسنجی موقعیت نقاط سیلاب‌های موجود با نقشه‌های حساسیت سیلاب مقایسه گردید و حساسیت سیلاب با داده‌های تعلیمی و صحت‌سنجی با استفاده از شاخص‌های نرخ موفقیت و نرخ پیش‌بینی مورد اعتبارسنجی قرار گرفت. در حال حاضر قابلیت پیش‌بینی حساسیت سیلاب منطقه مطالعاتی با استفاده از سطح زیر منحنی مورد بررسی قرار گرفت و از مقدار AUC، درصد نرخ موفقیت و قابلیت پیش‌بینی هر دو در پژوهش حاضر ۱۶ سناریو با ترکیب پارامترهای مختلف به عنوان ورودی مدل ایجاد شدند. نتایج این تحقیق نشان می‌دهد که در مدل ماشین بردار پشتیبان خطی با بهترین سناریو M3 با بالاترین ضریب همبستگی ۰/۹۷۲ و کمترین مقدار MAE=۰/۵۳۸، در مدل جنگل تصادفی بهترین سناریو M10 با بالاترین ضریب همبستگی ۰/۹۶۱ و کمترین مقدار خطا MAE=۰/۶۸۵، در مدل درخت تصمیم نوع چاید بهترین سناریو M8 با بالاترین ضریب همبستگی ۰/۹۵۴ و کمترین مقدار خطا MAE=۰/۷۲۳ بوده است. در نهایت مدل جنگل تصادفی و ماشین



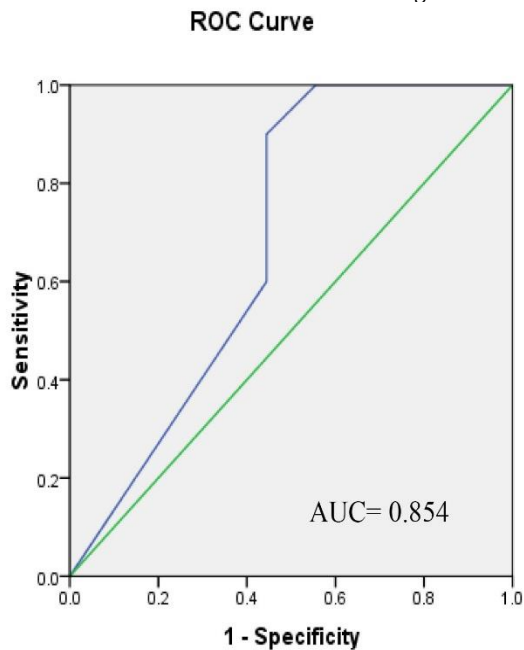
شکل ۲- پراکنش مدل ماشین پشتیبان بردار  
Figure 2. Vector backup machine model distribution



شکل ۳- پراکنش مدل جنگل تصادفی  
Figure 3. Distribution of a random forest model

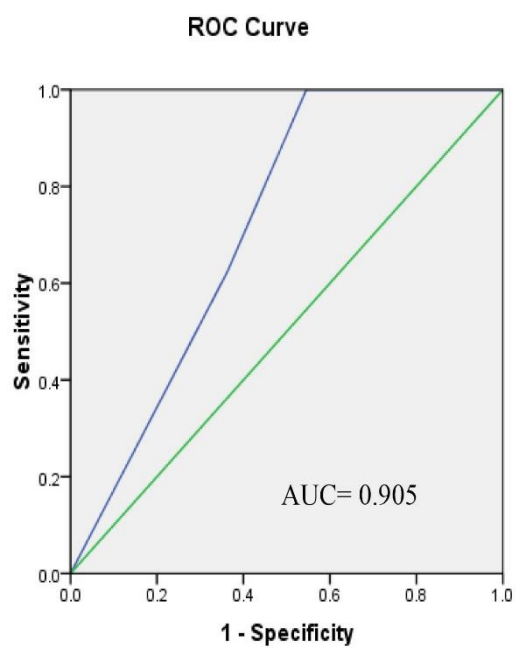


شکل ۴- پراکنش مدل چاید  
Figure 4. Chaid model distribution



Diagonal segments are produced by ties.

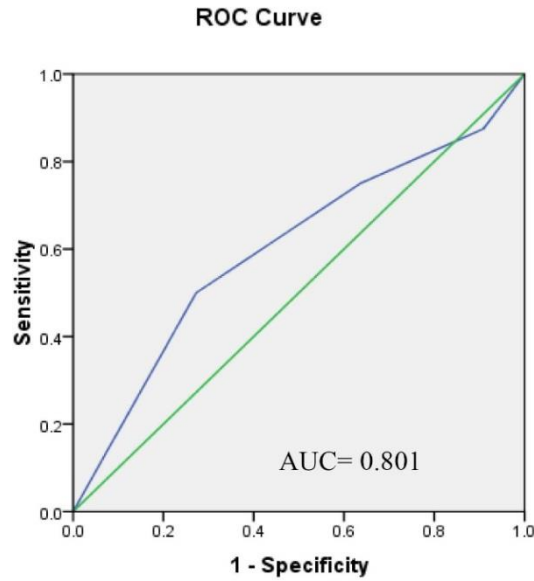
شکل ۶- منحنی ROC الگوریتم نوین جنگل تصادفی  
Figure 6. ROC curve of the new random forest algorithm



Diagonal segments are produced by ties.

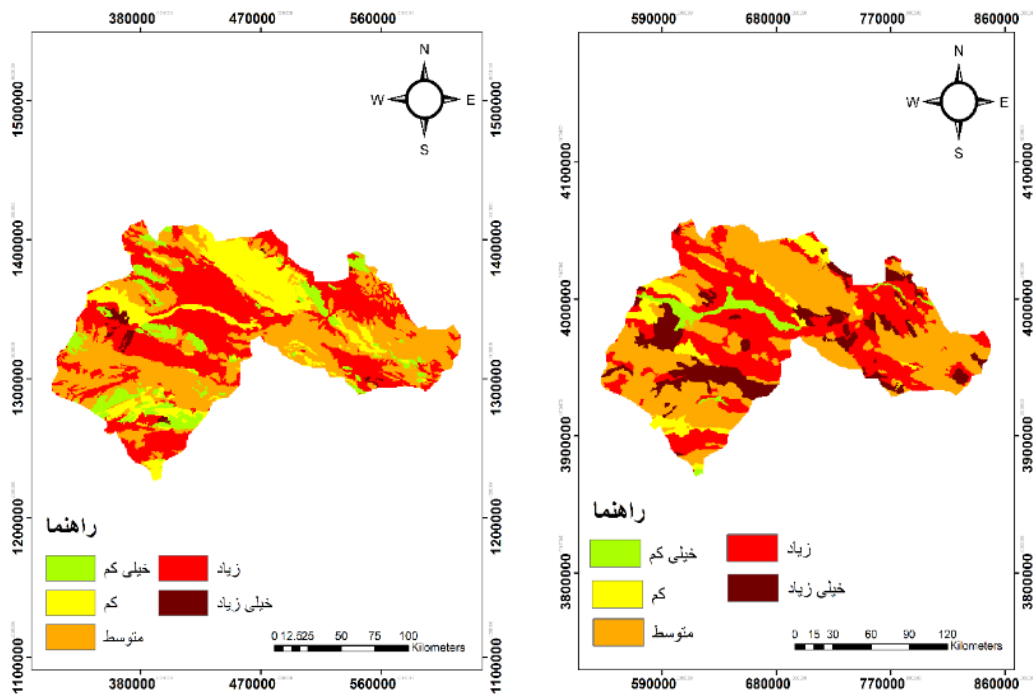
شکل ۵- منحنی ROC الگوریتم SVM-SIGMOID  
Figure 5. ROC curve of SVM-SIGMOID algorithm





Diagonal segments are produced by ties.

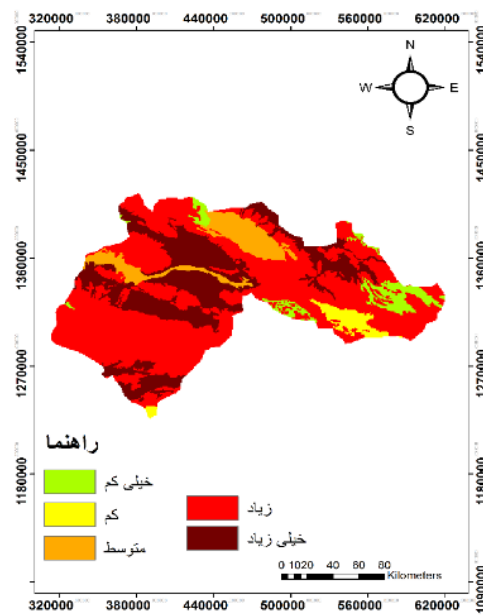
شکل ۷- منحنی ROC چاید  
Figure 7. Chaid ROC curve



شکل ۹- پهنه‌بندی با مدل جنگل تصادفی  
Figure 9. Zoning with random forest model

شکل ۸- پهنه‌بندی با مدل ماشین بردار پشتیبان  
Figure 8. Zoning with vector support machine model





شکل ۱۰- پهنه‌بندی با مدل چاید  
Figure 10. Zoning with Chaid model

جدول ۲- خروجی‌های مدل‌ها در محیط R

| درصد هر طبقه            | طبقات حساسیت                                    | نوع تابع                   |
|-------------------------|-------------------------------------------------|----------------------------|
| ۰/۰۷۱<br>۸۱/۴۴<br>۰/۲۲۳ | پارامتر سیگما<br>مقدار تابع هدف<br>خطای یادگیری | الگوریتم ماشین بردار (SVM) |
| ۰/۰۸۵<br>۹۳/۱۲<br>۰/۴۵۱ | پارامتر سیگما<br>مقدار تابع هدف<br>خطای یادگیری | مدل چاید                   |
| ۰/۰۷۹<br>۷۹/۲۴<br>۰/۳۲۱ | پارامتر سیگما<br>مقدار تابع هدف<br>خطای یادگیری | الگوریتم نوین جنگل تصادفی  |

جدول ۳- مقایسه مقادیر خروجی آزمون ROC بین مدل‌ها

| سطح اطمینان ۹۵ درصد     | معناداری                | خطای انحراف معیار       | مساحت زیر منحنی ROC     | مدل‌های حساسیت                                                      |
|-------------------------|-------------------------|-------------------------|-------------------------|---------------------------------------------------------------------|
| باند پایین<br>باند بالا |                         |                         |                         |                                                                     |
| ۰/۸۵۴<br>۰/۷۷۶<br>۰/۷۴۳ | ۰/۵۲۱<br>۰/۶۷۹<br>۰/۶۵۸ | ۰/۰۰۵<br>۰/۰۰۱<br>۰/۰۰۵ | ۰/۰۴۳<br>۰/۰۶۱<br>۰/۰۵۲ | ۰/۹۰۵<br>۰/۸۰۱<br>۰/۸۵۴                                             |
|                         |                         |                         |                         | الگوریتم ماشین بردار (SVM)<br>مدل چاید<br>الگوریتم نوین جنگل تصادفی |

### نتیجه‌گیری کلی

به‌طور کلی بر اساس نتایج تحقیق حاضر طبقات با پتانسیل سیل‌خیزی کم و متوسط بیشتر در بخش‌های شرقی و جنوبی حوضه قرار گرفته‌اند به طوری که در بخش شرقی حوضه به علت شیب کم و نفوذپذیری مناسب خطر سیل‌خیزی متوسط می‌باشد. بر اساس نتایج به دلیل وجود کاربری‌های مرتعی فقیر در نیمه غربی و شمال غربی حوضه بیشترین پتانسیل سیل‌خیزی مشاهده شده قرار دارند. عدم وجود پوشش گیاهی مناسب در این مناطق نارسایی در نفوذپذیری نزولات جوی را افزایش و در نتیجه سیل‌خیزی حوضه را افزایش می‌دهد. عوامل طبیعی مانند شیب زیاد و ارتفاع بالا و پوشش گیاهی

فقیر باعث تشدید در سیل‌خیزی می‌شود. همچنین نتایج نشان داد قسمت‌های شمالی و غربی حوضه که از نظر زمین‌شناسی و لیتولوژی دارای سازندهای سطحی مانند مارن، رس و سیلت بوده و ضریب نفوذپذیری آن‌ها بسیار کم و پوشش گیاهی ناچیز دارند، از پتانسیل بالایی در بروز سیلاب برخوردار هستند. در این پژوهش ارزیابی مدل‌ها با استفاده از ضریب همبستگی (R) و مقادیر میانگین قدر مطلق خطا (MAE) صورت گرفت. بررسی نتایج مدل‌ها نشان داد که به ترتیب مدل‌های ماشین بردار پشتیبان، چاید و جنگل تصادفی با سناریوی M3، M8 و M10 با بیشترین میزان همبستگی و کمترین خطای میانگین، از دقت بالایی در برآورد

گرفته‌اند. از نتایج این پژوهش می‌توان به منظور شناسایی مناطق مستعد و حساس در برابر سیل و در نتیجه تعیین مکانهای مناسب و امن جهت توسعه‌های عمرانی و زیرساختی، و همچنین در تعیین زمان تقریبی رخداد سیل استفاده و تا حدودی از خسارات و تلفات ناشی از این مخاطره جلوگیری نمود.

خطر وقوع سیل در منطقه مطالعاتی برخوردار هستند. علاوه بر این، به منظور ارزیابی مدل‌های ارائه شده از مساحت زیر منحنی (ROC) استفاده شد. بر این اساس این مقادیر هم در داده‌های تعلیمی و هم در داده‌های آموزشی در الگوریتم (SVM) و مدل الگوریتم نوین جنگل تصادفی دارای نتایج دقیق‌تری می‌باشد. این نتیجه بیانگر این مطلب است که هر دو مدل از نظر صحت و اعتبار مدل‌سازی مورد تأیید قرار

## منابع

1. Abdollahzadeh, A., M. Ownegh, A. Sadoddin and R. Mostafazadeh. 2015. Technical Note: Constraints to residential land use development arising from flood and runoff coefficient in a land use planning framework, case study: Ziarat Watershed, Golestan Province, Watershed Engineering and Management, 8(2): 221-235 (In Persian).
2. Abedini, M. and M.H. Fathi. 2015. Flood Risk Mapping and Evaluation by using the Analytic Network Process Case Study: (Khiav Chai Catchment), Journal of Hydrogeomorphology, 2(3): 99-120.
3. Abedini, M. and R. Fathi Jokadan. 2016. Flood Risk zoning in the Karganroud's Catchment Basin Using ArcGIS, Hydrogeomorphology, 7: 1-17.
4. Towfiqul Islam, A.R., S. Talukdar, S. Mahato, S. Kundu, K. UddinEibek, Q. Bao Pham, A. Kuriqi, Nguyen, and T. ThuyLinh. 2021. Flood susceptibility modelling using advanced ensemble machine learning models, Geoscience Frontiers, 12(3): 101075.
5. Alizadeh, A. 2011. Principles of Applied Hydrology, 33rd Edition. Ferdowsi University of Mashhad. 990 (In Persian).
6. Arab Ameri, A.L., H.R. Pourghasemi and K. Shirani. 2017. Flood susceptibility zonation using new ensemble Bayesian-AHP methods (Case study: Neka Watershed, Mazandaran Province), Echo Hydrology, 4(2): 462-447.
7. Asgari, Sh., A.A. Safari and H.A. Fathi. 2018. Investigation of flood potential in Jafarabad catchment. Journal of Applied Research in Geographical Sciences, 50: 76-90 (In Persian).
8. Asghari Saraskanrood, S., E. Pirooz and B. Zeinali. 2015. Flood risk zoning in Aq Laqanchai watershed using Vickor model, Quantitative Geomorphological Research, 4(3): 231-245.
9. Avand, M.T., H.R. Moradi and M. Ramazanzadeh. 2020. Flood Susceptibility Mapping Using Random Forest Machine Learning and Generalized Bayesian Linear Model, Environment and Water Engineering, 6(1): 83-95.
10. Faiz Nia, S., Z.Z. Mosaviyan, A. Abdollahiyan dehkordi and Kh. Ebrahimi Darcheh. 2016. Investigation of the effect of geology on flooding a case study of the Junqan and Aqa watersheds in the city. You have pasture and watershed, Iranian Journal of Natural Resources, 69(4): 1017-1029 (In Persian).
11. Faizi Kushki, F.F., D. Rasooli and H. Haji mohammadi. 2019. Investigation of synoptic patterns governing the occurrence of floods in the Karun Basin, Quarterly Journal of Crisis Prevention and Management Knowledge, 9(1): 22-34.
12. Fathalizadeh, B., M. Abedini and M. Rajabi. 2020. Investigating the Causes and Hazards of flood in Zunuzchay Watershed using HEC-HMS Hydrological Model and Fuzzy Logic, Quantitative Geomorphology Research, 9(1): 134-155.
13. Fernandez, D. and M. Lutz. 2010. Urban flood hazard zoning in Tucuman Province, Argentina, using GIS and multi criteria decision analysis, Engineering Geology, 111(1): 90-98.
14. Gholami, M., E.A. Ghanavati and A. Ahmadabadi. 2019. Landslide susceptibility mapping of Kan using index of Entropy and SVM, Quantitative Geomorphological Research, 8(1): 16-33.
15. Hosseini, Y., J. Ramezani Moghaddam and Z. Abdolalizadeh. 2019. Evaluating the Impact of Land Use Changes on Flooding and Flood Runoff in Amuqin Drainage Basin, Journal of Natural Environmental Hazards, 8(22): 145-162.
16. Jahantigh, M. and M. Jahantigh. 2019. Study effect of flood productivity on vegetation changes using field work and Landsat satellite images (Case study: Shandak of Sistan region), RS & GIS for Natural Resources, 10(4): 57-72.
17. Javadi, M.R., M. Baghery, M. Vafakhah and Sh.A. Gholami. 2012. Effect of Flood Spreading on Physical Soil Properties (A Case Study: Delijan Flood Spreading), Journal of Watershed Management Research, 5(9): 119-129 (In Persian).
18. Kisi, O. and Y. Kilic. 2015. An investigation on generalization ability of artificial neural networks and M5 model tree in modeling reference evapotranspiration, Theor Appl Climatol, 1-13.
19. Leander, R., T.A. Buishand, B.J.J.M. van den Hurk, and M.J.M. de Wit. 2008. Estimated changes in flood quantiles of the river Meuse from resampling of regional climate model output, Journal Hydrology, 351(3-4): 331-343.

20. Merufinia, E., E. Nohani, K. Khosravi and K. Chapi. 2016. Evaluation of Statistical Index Method in Flood Susceptibility Mapping, *Journal of Soil and Water Knowledge*, 26(2/4): 201-214.
21. Mojaddadi Rizeei, H., M. Habibnezhad Roshan, K. Shahedi and B. Pradhan. 2020. The Efficiency of an Ensemble Frequency Ratio-Support Vector Machine Model in the Detection of Flood-Prone Areas of the Kalat Basin. *Echo Hydrology*, 7(1): 77-95.
22. Mosafaei, J., F. Sarfaraz and A. Saleh Por. 2018. Investigation of the effect of water and soil protection projects on runoff coefficient and characteristics of Parroud watershed, 13th National Conference on Watershed Management Science and Engineering of Iran and 3rd National Conference on Natural Resources and Environment Protection, 1-8.
23. Mosaffaie, J., M. Kamali, A. Salehpour Jam, K. Shahedi, K. Soleiman and A. Gomrokchi. 2020. Evaluation the Efficiency of AHP Model in Prioritizing of Barajin Sub Watersheds from Flood Potential Viewpoint, *Journal of Watershed Management Research*, 11(21): 48-58 (In Persian).
24. Nadiri, M. 2019. Flood Risk Zoning Using TOPSIS-AHP Fuzzy Logic, *Journal of Geography (Regional Planning)*, 9(3): 293 -363.
25. Neghresh, H. and J. Veisi. 2013. Analysis of the effects of rainfall changes on flooding of the Ravand river basin in the west of Islamabad, Kermanshah province, *Journal of Regional Planning*, 11: 79-98.
26. Nouri, H., K. Shahedi, M. Habibnezhad Roshan, A. Kaviani and M. Faramarzi. 2019. Susceptibility to flooding in the Razavar Watershed using analytical hierarchy process method, *Journal of Natural Hazards*, 8(19): 35-50.
27. Pahlavan Rad, M.R., N. Toomanian, F. Khormali, C. Brungard, C.B. Komaki and P. Bogaert. 2014. Updating soil survey maps using random forest and conditioned Latin hypercube sampling in the loess-derived soils of northern Iran, *Journal of Geoderma*, 232: 97-106.
28. Pourghasemi, H.R., H.R. Moradi, S.M.F. Aghda, C. Gokceoglu and B. Pradhan. 2014. GIS-based landslide susceptibility mapping with probabilistic likelihood ratio and spatial multi-criteria evaluation models (North of Tehran, Iran). *Arab Journal Geosci*, 7(5): 1857-1878.
29. Qanbarzadeh, H., A. Behniyafar, M.R. Servati, M. Mosavi and A.M. Noormohammady. 2015. The Impact of Geomorphic Variables of Fluvial on Flood Hazard of Mountain steep Towns (Case Study: Town Torghabeh, Khorasan Razavi province), *Environmental Planning Quarterly*, 31: 77-98.
30. Rajabi shahbandi, E., M. Masoudian and R. Fazl Oula. 2020. The Effect of Intercity Retarding Reservoirs on Flood Peak Discharge and Reduce the Dimensions of Flood Collection Network (Case Study: Gozarkhan Subcatchment, Sari), *Journal of Watershed Management Research*, 11(21): 72-83 (In Persian).
31. Rastgoo, A., A.R. Ghanbari and A. Ansari Lari. 2019. Evaluation and potential assessment of flood risk in Jinnah city using SVM support vector machine algorithm, *Quarterly Journal of Natural Geography*, 12(45): 107-125.
32. Razavizadeh, S. and K. shahedi. 2016. Prioritizing flooding in Taleghan watershed using a combination of TOPSIS and AHP, *Iranian Natural Ecosystems Quarterly*, 7(4): 33-46.
33. Roustaei, Sh., F. Ayaseh and M.H. M. Rezaei Moghadam. 2020. Quasi-two-dimensional simulation of the flood flow of Liqvan River with emphasis on the flood plain, *Quantitative Geomorphological Research*, 9(1): 28-21.
34. Shahab, S., P. Subodh Chandra, S. Asish, R. Chakraborty, M. Assefa and M. Melesse Amirhosein. 2020. Flash Flood Susceptibility Modeling using New Approaches of Hybrid and Ensemble Tree-Based Machine Learning Algorithms, *Remote Sensing*, Published. 31(12): 3568, doi:10.3390/rs12213568: 2-23.
35. Shrif Garmdareh, E., M. Vafakhah and S. Eslamian. 2019. Assessment the Performance of Support Vector Machine and Artificial Neural Network Systems for Regional Flood Frequency Analysis (A Case Study: Namak Lake Watershed), *Journal of Water and Soil Science (Science and Technology of Agriculture and Natural Resources)*, 23(1): 351-365.
36. Siasar, H. and T. Honar. 2019. Application of Support vector machine, CHAID and Random forest models, in estimated daily Reference evapotranspiration in northern Sistan and Baluchestan province, *Iranian Journal of Irrigation and Drainage*, 2(13): 378-388.
37. Mobley, W., A. Sebastian, R. Blessing, E. Wesley, L. Highfield and D. Samuel. 2020. Quantification of Continuous Flood Hazard using Random Forrest Classification and Flood Insurance Claims at Large Spatial Scales: A Pilot Study in Southeast Texas Natural Hazards and Earth System Sciences. <https://doi.org/10.5194/nhess-2020-347>: 1-22.
38. Velayati, S.I., A. Ghayyur Hassan and S. Shafi'a. 2004. Investigation of the hydrological regime of Frizi river (Kashfarud sub-basin) and its role in feeding the aquifer of Mashhad plain, *Geography and development*, 2(3): 47-72 (In Persian).
39. Yousefzadeh, A., B. Zeynali, Kh. Valizadeh Kamran and S. Asghari SAR Eskinrood. 2019. The Extraction of Flood Potential of Simineh River Basin Applying Satellite Images. Topographic Wetness Index and Morphological Features, *Geography and Sustainability of Environment*, 32: 49-61.
40. Zahradyan, Gh.R. and P. Rahimzadeh. 2010. Investigation of the effect of flood spreading on soil permeability. Case study: Mosian plain of Ilam province, *Quarterly Journal of Natural Geography*, 9: 1-14.

## Assessment of Flood Occurrence Potential using Data Mining Models of Support Vector Machine, Chaid and Random Forest (Case study: Frizi watershed)

Mahdi Zarei<sup>1</sup>, Rahman Zandi<sup>2</sup> and Mahnaz Naemi Tabar<sup>3</sup>

1- Assistant Professor, Research Center of Social Studies and Geographical Sciences, Hakim Sabzevari University, Sabzevar, Iran, (Corresponding author: m.zarei@hsu.ac.ir)

2- Assistant Professor, Faculty of Geography and Environmental Sciences, Hakim Sabzevari University, Sabzevar, Iran

3- PhD Student in Geomorphology, Faculty of Geography and Environmental Sciences, Hakim Sabzevari University, Sabzevar, Iran

Received: 24 February, 2021 Accepted: 3 May, 2021

### Extended Abstract

**Introduction and Objective:** Flood, like other hydrological phenomena, is an uncertain phenomenon that can occur at any time and place and is influenced by various climatic factors, physical characteristics of the basin, vegetation status and land use, and human intervention. Determining the contribution of each parameter to the flood incidence is important. At present, with the development of GIS, remote sensing (RS), and machine learning (ML) methods, very accurate modeling of flood probability can be performed. However, the construction of these models requires accurate and principled knowledge of the flood occurrence process, the study of effective parameters in flood formation, understanding of how each parameter affects flood generation, and the selection and development of appropriate models and their evaluation. Due to the importance of determining flood-prone areas in different areas, especially basins located in arid and semi-arid areas such as the study area, the present study was conducted to assess flood risk using vector mining data mining models. Random support, grass, and forest are targeted in this area.

**Material and Methods:** In the present study, to support the risk of flooding, data support models of support vector machine, field, and random forest have been used. In general, the purpose of presenting data mining models is to achieve a reasonable and accurate estimate of spatial prediction of flood occurrence, compare the efficiency of the models and select the most appropriate method for preparing a flood sensitivity assessment map. In this study, from various information such as the topographic map of scale 1: 50000 to extract level lines, a geological map of scale 1: 100000, a soil map prepared by the General Department of Natural Resources and Watershed Management of Khorasan Razavi province, digital elevation model (DEM) image with Spatial resolution of 12.5 m, Google Earth satellite imagery, meteorological data, rain gauge, statistical period of 20 years (98-78), Andarkh stations, Olang Asadi, Kardeh Dam, Marshak, Bulgur, Bala Gosh, Al, Chenaran, Moghan, Chekneh Olya, Abqad Frizi, Talgur, Qadirabad, and Kabkan have been used. Elevation, slope, slope direction, drainage networks, main waterways, and convexity of the ground surface were extracted from the DEM image and level lines. Land use of the region was prepared from Google Earth satellite images related to 2020 and in a supervised classification method. The vegetation map of the region was also prepared based on the NDVI index and from satellite images of Landsat 8 in 2018.

**Results:** The elevation factor plays a key role in controlling the direction of flood movement and water surface depth. At an altitude of 2000 m and more, with increasing altitude, the flood potential in the study area increases. According to the results, among the uses of the studied basin, irrigated and garden lands produce less runoff due to more infiltration and are less prone to flooding. In the study area, at a slope of 60 degrees, due to the increase in slope, the latency of the basin is low, the amount of water infiltration into the soil is low, and as a result, the volume of floods and surface runoff will increase. Class 0/0074-0/0120 has the greatest impact on the occurrence of basin floods. The northern, northwestern, and western slopes have the potential for flooding due to heavy rainfall, long-term snow retention, and moisture. In the study area, more than 250 mm of rainfall has the greatest impact on the occurrence of floods. In the study area, due to the relatively low permeability of the soil, the soil produces more runoff and floods. Based on the results of topographic moisture index classes in the study area, classes 268/38-359/99 had a great impact on the occurrence of floods. In the study area, concave areas have a great impact on floods because the most important and effective factors in the occurrence of floods are the slope and curvature of the earth. At present, the predictability of flood sensitivity in the study area was investigated using the area under the curve. The results of this study show that in the linear support vector machine model with the best scenario M3 with the highest correlation coefficient of 0.972 and the lowest value of MAE = 0.538, and in the random forest model the best scenario M10 with the highest correlation coefficient 961 0.0 and the lowest error value MAE = 0.685, in the Chaid decision tree model, the best scenario was M8 with the highest correlation coefficient of 0.954 and the lowest error value was MAE = 0.723.

**Conclusion:** In general, according to the results of the present study, the floors with low and medium flooding potential are more located in the eastern and southern parts of the basin, so in the eastern part of the basin due to low slope and good permeability flood risk, It is average. According to the results, due to the existence of poor rangeland land uses in the western and northwestern half of the basin, the highest flood potential has been observed. The results also showed that the northern and western parts of the basin, which in terms of geology and lithology have surface formations such as marl, clay, and silt, and their permeability coefficient is very low and vegetation is low, have a high potential for occurrence. They have flooded. In this study, the models were evaluated using correlation coefficient (R) and mean absolute error values (MAE). Examination of the results of the models showed that the support vector machine, chad, and random forest models with scenarios M3, M8, and M10 with the highest correlation and the lowest mean error, respectively, have high accuracy in estimating the risk of floods in the study area. In addition, the area under the curve (ROC) was used to evaluate the proposed models. Accordingly, these values have more accurate results in both educational data and training data in the algorithm (SVM) and the new random forest algorithm model. This result indicates that both models have been validated in terms of modeling accuracy and validity.

**Keywords:** Data mining models, Flood, Friesian watershed, ROC curve, Zoning