

ارزیابی روش‌های داده‌کاوی و آمار دومتغیره در پهنه‌بندی خطر وقوع گسل (مطالعه موردی: حوضه قره‌قوم)

مهدی بشیری^{۱*}، سیده‌مائده کاوسی‌داودی^۲ و علی افضلی^۳

۱- نویسنده مسئول، استادیار، گروه مهندسی طبیعت و گیاهان دارویی، دانشکده کشاورزی و منابع طبیعی، دانشگاه تربت‌حیدریه، خراسان رضوی، ایران،
پست الکترونیک: m.bashiri@torbath.ac.ir

۲- کارشناس ارشد آبخیزداری، شرکت سامانه آب کومش، خراسان رضوی، ایران

۳- کارشناس گروه آموزشی مدیریت مناطق بیابانی، مرکز تحقیقات بین‌المللی بیابان، دانشگاه تهران، ایران

تاریخ پذیرش: ۹۸/۰۴/۱۰

تاریخ دریافت: ۹۷/۱۰/۰۸

چکیده

گسل از فرایندهای اصلی زمین‌ریختی است که با پدیده‌هایی همانند زمین‌لرزه ارتباط دارد. از این رو آگاهی از اثر عوامل محیطی در بروز گسل و شناخت مناطق پرخطر اهمیت زیادی دارد که این اهداف در این پژوهش بررسی شده‌اند. عرصه پژوهش حوضه قره‌قوم است که پس از تعیین گسل‌های آن، متغیرهای ارتفاع، شیب، جهت، اقلیم، کاربری، ژئومورفولوژی، سنگ‌شناسی، فرسایش، بارش، پهنه‌های سیلابی، خاک‌شناسی، شاخص‌های قدرت جریان، ناهمواری زمین و فاصله از جاده، آبراهه، طاق‌دیس و ناودیس آن استخراج گردید. سپس الگوریتم‌های داده‌کاوی شامل درخت‌تصمیم، جنگل‌تصادفی، تجمیعی‌بوستینگ، ماشین‌بردار پشتیبان، رگرسیون لوجیستیک و شبکه‌عصبی در نرم‌افزار R برای شناسایی ارزش‌متغیرها و روش‌های آماری دومتغیره شامل ارزش‌اطلاعات و تراکم سطح برای شناسایی ارزش‌کلاس‌های هر متغیر در وقوع گسل برآزش شد. ارزیابی دقت طبقه‌بندی الگوریتم‌ها با منحنی ROC نشان داد که الگوریتم‌های جنگل‌تصادفی و ماشین‌بردار پشتیبان به ترتیب با سطح زیر منحنی ۸۸ و ۸۶ درصد بهترین عملکرد را در طبقه‌بندی وقوع گسل بر مبنای متغیرهای ورودی دارند. در نهایت با ضرایب جینی در الگوریتم جنگل‌تصادفی، نقشه‌های پهنه‌بندی به‌دست آمده از ترکیب این الگوریتم با روش‌های آماری دومتغیره تهیه و اعتبارسنجی شد. طبق این الگوریتم، به ترتیب متغیرهای ارتفاع، خاک‌شناسی و شاخص ناهمواری زمین مهمترین پارامترها در وقوع گسل شناخته شدند. طبق ارزیابی نقشه‌های پهنه‌بندی، در روش ارزش‌اطلاعات و تراکم سطح به ترتیب حدود ۵۲ و ۳۵ درصد گسل‌ها در کلاس خطر خیلی‌زیاد قرار گرفتند. از این رو روش ارزش‌اطلاعات در تعیین پهنه‌های حساس به وقوع گسل دقت بیشتری داشت. با توجه به نتایج، روش‌های داده‌کاوی به‌عنوان ابزاری مفید در مدیریت خطر گسل معرفی گردید. همچنین لزوم توجه به متغیرهای محیطی به‌ویژه توپوگرافی در مراحل مدیریت و تغییر کاربری حوضه ضرورت می‌یابد.

واژه‌های کلیدی: ارزش‌اطلاعات، الگوریتم طبقه‌بندی، تراکم سطح، منحنی ROC.

مقدمه

زمین‌لغزش از پرخسارت‌ترین آنهاست و همگام با دست‌کاری بشر در سیستم‌های طبیعی شتاب فزاینده‌ای یافته است (Ilanlou et al., 2009). عرصه‌های طبیعی به‌دلیل

توسعه و عمران شهری و روستایی موجب بروز برخی ناهنجاری‌های طبیعی شده است که گسل، حرکات توده‌ای و

تحولات اقتصادی-اجتماعی و رشد صنعتی به صورت کمی و کیفی مورد تهدید و تخریب است (Gohardoust *et al.*, 2017) و پیش‌بینی شده است که به علت ادامه تغییرات انسان در طبیعت، رخداد گسل و خسارتهای حاصل از آن ابعاد گسترده‌تری پیدا کند (Jedari-Eyvazi & mahmoudi, 2001). بررسی پیامدهای برخی عوامل محیطی همانند کاربری اراضی، زمین‌شناسی و ژئومورفولوژی بر وقوع گسل، کمک شایانی به برنامه‌ریزی می‌کند و اطلاعات ارزشمندی را در زمینه مدیریت پایدار ارائه می‌دهد. هدف از تعیین الگوی گسل‌ها، پیش‌بینی مکان وقوع پدیده‌های ناشی از آن است و شناسایی عوامل مؤثر بر وقوع و پهنه‌بندی حساسیت آن برای تشخیص مناطق مستعد خطر و انتخاب مکان‌های مساعد توسعه، بسیار مفید می‌باشد (Anbalagan, 1992). پیش‌بینی زمان رخداد گسل از توان دانش فعلی بشر خارج است، از این رو با شناسایی مناطق حساس به گسل می‌توان تا حدودی از خطر ناشی از بروز آن جلوگیری نمود. این موضوع در آبخیزداری با توجه به هدف آن در اجرای مدیریت جامع منابع طبیعی موجود در حوضه و استفاده بهینه از آنها بررسی می‌شود (Ilanlou *et al.*, 2009). زیان‌های اجتماعی و اقتصادی گسل را می‌توان با برنامه‌ریزی و مدیریت مؤثر کاهش داد، این رویکردها شامل محدودیت توسعه در مناطق مستعد، رعایت اصول مهندسی و توسعه سیستم‌های هشدار است. اما شناخت مناطق حساس نیازمند استفاده از علوم جدید می‌باشد. در این راستا، علم داده‌کاوی، با هدف شناسایی الگوها و ارتباطات قابل فهم، معتبر، جدید و به‌طور بالقوه سودمند در داده‌ها ابداع گردید. مفهوم داده‌کاوی شامل الگوریتم‌ها و روش‌هایی است که اطلاعات را از داده‌ها استخراج می‌کند (Fayyad *et al.*, 1996). به بیان دیگر این روش‌ها راه‌های پیاده‌سازی عملیات داده‌کاوی هستند. از آنجاکه هر الگوریتم دارای نقاط ضعف و قوت است، در بین این الگوریتم‌ها بهترین وجود ندارد و با توجه به داده‌ها و کارایی موردنظر، باید ابزار یا ابزارهایی انتخاب شده و مدل مناسب طراحی و اجرا شود (Ghazanfari *et al.*, 2008). مزایای عمده تکنیک‌های داده‌کاوی شامل کاربرد آن در علوم مختلف، امکان یادگیری و پیش‌بینی و کشف الگوهای پنهان در داده‌ها می‌باشد. محدودیت‌های اصلی این تکنیک‌ها نیز تطبیق‌پذیری کم، نیاز به دانش قابل توجه، پیچیدگی زیاد و عدم قابلیت تعمیم به مسائل مختلف هستند. اما به‌طور کلی کاربرد این روش‌ها موجب کاهش وابستگی به عملیات مفصل میدانی و کاهش حجم، زمان و هزینه عملیات می‌شود (Mahdavi *et al.*, 2017). همچنین پژوهشگران در راستای توسعه روش‌ها و تکنیک‌های پیشرفته داده‌کاوی، برای بهبود دقت طبقه‌بندی، تلاش‌های گسترده‌ای انجام داده‌اند (Fathizad *et al.*, 2017). همسو با این پژوهش در زمینه مدل‌سازی و پهنه‌بندی مخاطرات طبیعی، مطالعات مختلفی انجام شده است. به‌عنوان نمونه در خارج از کشور، کاربرد مدل نسبت فراوانی و رگرسیون لجستیک در تهیه نقشه حساسیت زمین‌لغزش با سیستم اطلاعات جغرافیایی بررسی شد. با ورود متغیرهای شیب، جهت، شکل شیب، پستی‌وبلندی، بافت و مواد خاک، زهکشی و ضخامت مؤثر خاک، نوع جنگل، قطر و سن درخت، تراکم جنگل و کاربری به فرایند مدل‌سازی، مدل رگرسیون لجستیک در مقایسه با نسبت فراوانی صحت بالاتری را نشان داد (Lee, 2004). اثر عوامل ارتفاع، سنگ‌شناسی و شیب بر زمین‌لغزش‌های ژاین از طریق مقایسه دو مدل رگرسیون لجستیک و تحلیل سلسله مراتبی مطالعه شد. نتایج نشان داد که در تحلیل سلسله مراتبی ۷۰ درصد و در رگرسیون لجستیک ۶۳ درصد زمین‌لغزش‌ها در مناطق با حساسیت بالا قرار دارند (Ayalew *et al.*, 2005). مقایسه مدل‌های نسبت فراوانی و رگرسیون لجستیک با استفاده از سیستم اطلاعات جغرافیایی انجام شد. بررسی جهت و شکل شیب، فاصله از جاده، لیتولوژی، فاصله از عارضه خطی و کاربری نشان داد که در دو مدل نسبت فراوانی و رگرسیون لجستیک به ترتیب ۸۶/۹۷ و ۸۶/۳۷ درصد محدوده در مناطق پرخطر هستند (Lee & Sambath, 2006). با ترکیب آنالیز شیء‌گرا و الگوریتم جنگل تصادفی، چهار منطقه در هائیتی، چین، ایتالیا و فرانسه از نظر حساسیت لغزش بررسی و از الگوریتم

داده‌کاوی شامل کاربرد آن در علوم مختلف، امکان یادگیری و پیش‌بینی و کشف الگوهای پنهان در داده‌ها می‌باشد. محدودیت‌های اصلی این تکنیک‌ها نیز تطبیق‌پذیری کم، نیاز به دانش قابل توجه، پیچیدگی زیاد و عدم قابلیت تعمیم به مسائل مختلف هستند. اما به‌طور کلی کاربرد این روش‌ها موجب کاهش وابستگی به عملیات مفصل میدانی و کاهش حجم، زمان و هزینه عملیات می‌شود (Mahdavi *et al.*, 2017). همچنین پژوهشگران در راستای توسعه روش‌ها و تکنیک‌های پیشرفته داده‌کاوی، برای بهبود دقت طبقه‌بندی، تلاش‌های گسترده‌ای انجام داده‌اند (Fathizad *et al.*, 2017). همسو با این پژوهش در زمینه مدل‌سازی و پهنه‌بندی مخاطرات طبیعی، مطالعات مختلفی انجام شده است. به‌عنوان نمونه در خارج از کشور، کاربرد مدل نسبت فراوانی و رگرسیون لجستیک در تهیه نقشه حساسیت زمین‌لغزش با سیستم اطلاعات جغرافیایی بررسی شد. با ورود متغیرهای شیب، جهت، شکل شیب، پستی‌وبلندی، بافت و مواد خاک، زهکشی و ضخامت مؤثر خاک، نوع جنگل، قطر و سن درخت، تراکم جنگل و کاربری به فرایند مدل‌سازی، مدل رگرسیون لجستیک در مقایسه با نسبت فراوانی صحت بالاتری را نشان داد (Lee, 2004). اثر عوامل ارتفاع، سنگ‌شناسی و شیب بر زمین‌لغزش‌های ژاین از طریق مقایسه دو مدل رگرسیون لجستیک و تحلیل سلسله مراتبی مطالعه شد. نتایج نشان داد که در تحلیل سلسله مراتبی ۷۰ درصد و در رگرسیون لجستیک ۶۳ درصد زمین‌لغزش‌ها در مناطق با حساسیت بالا قرار دارند (Ayalew *et al.*, 2005). مقایسه مدل‌های نسبت فراوانی و رگرسیون لجستیک با استفاده از سیستم اطلاعات جغرافیایی انجام شد. بررسی جهت و شکل شیب، فاصله از جاده، لیتولوژی، فاصله از عارضه خطی و کاربری نشان داد که در دو مدل نسبت فراوانی و رگرسیون لجستیک به ترتیب ۸۶/۹۷ و ۸۶/۳۷ درصد محدوده در مناطق پرخطر هستند (Lee & Sambath, 2006). با ترکیب آنالیز شیء‌گرا و الگوریتم جنگل تصادفی، چهار منطقه در هائیتی، چین، ایتالیا و فرانسه از نظر حساسیت لغزش بررسی و از الگوریتم

قره‌قوم که دارای گسل‌های فعال و غیرفعال فراوانی است، تشخیص مناطق پرخطر و اثر عوامل محیطی بر گسل‌ها در کاهش خسارتها ضروریست. آگاهی از نقش عوامل مؤثر، گامی مهم در ارزیابی خطر است و در مدیریت بحران و توسعه پایدار نقش کلیدی دارد. بررسی الگو و عوامل مرتبط با گسل به منظور شناسایی مناطقی که مورد تهدیدند، شناسایی مکان‌های امن برای توسعه سکونت‌گاه یا کاربری‌هایی مانند راه و مسیرهای انتقال نیرو و انرژی مورد توجه برنامه‌ریزان است. آنچه گفته شد اهدافی است که پژوهش پیش‌رو در حوضه قره‌قوم به دنبال دستیابی به آن است.

مواد و روش‌ها منطقه مورد مطالعه

حوضه قره‌قوم یکی از شش حوضه بزرگ آبریز کشور، در خراسان رضوی واقع شده است. این حوضه با مساحت ۴۴۱۲۰ کیلومترمربع بین مختصات ۱۲ دقیقه و ۵۸ درجه تا ۱۱ دقیقه و ۶۱ درجه طول شرقی و ۱۹ دقیقه و ۳۴ درجه تا ۳۳ دقیقه و ۳۷ درجه عرض شمالی قرار دارد. این حوضه از شمال و شمال‌غرب با ترکمنستان و از شرق با افغانستان هم‌مرز است و شامل ۱۳ دشت درگز، کلات‌نادر، گنبدلی، سرخس، آق‌دربند، نریمان، مشهد، سنگ‌بست، صالح‌آباد-جنت‌آباد، فریمان-تربت‌جام، تایباد، شهرنو-باخرز و کرات می‌باشد (Ministry of Energy, 2015). مهمترین رشته کوه‌های آن هزارمسجد در شمال و شمال‌غرب و بینالود در غرب است. در مجموع ۲۱ گسل فعال در خراسان‌رضوی وجود دارد که کشف‌رود، تربت‌جام، بینالود، سبزوار، دشت‌بیاض و سرایان و نیشابور یا کپه‌داغ و درونه از گسل‌های مهم آن است. وضعیت گسل‌های حوضه و موقعیت آن در شکل ۱ آمده است.

تهیه لایه‌های اطلاعاتی حوضه

برای تحلیل، لایه‌های اطلاعاتی کاربری اراضی، شیب،

جنگل تصادفی برای انتخاب ویژگی‌های مفید استفاده شد. صحت نتایج پهنه‌بندی در این مناطق از ۷۳/۳ تا ۸۷/۱ درصد متفاوت بود (Andre & Norman, 2010). همچنین با نسبت وقوع احتمالی و ارزیابی چند متغیره، پهنه‌بندی لغزش در شمال تهران انجام شد که نتایج پژوهش، مدل نسبت وقوع احتمالی را برتر از ارزیابی چند متغیره نشان داد (Pourghasemi *et al.*, 2014). در نهایت در بررسی حساسیت لغزش با AHP در یونان، مهمترین عوامل شامل شیب و لیتولوژی به دست آمدند و مدل AHP با قدرت ۶۹/۴۵ درصد، پهنه‌های با خطر بالا و خیلی بالا را تخمین زد (Papadakis & Karimalis, 2017). در پژوهش‌های داخل کشور، ارزیابی روش‌های ارزش اطلاعات، تراکم سطح، تحلیل سلسله مراتبی و گوپتا و جوشی در پهنه‌بندی خطر لغزش در استان ایلام نشان داد که روش ارزش اطلاعات کارایی بهتری دارد (Naderi & Karimi, 2011). پهنه‌بندی خطر زمین‌لغزش با مدل‌های ارزش اطلاعاتی، تراکم سطح و LNRF در حوضه چالکروند نشان داد که در مدل تراکم سطح، LNRF و ارزش اطلاعاتی به ترتیب ۷۹، ۶۳ و ۶۶ درصد از حوضه در محدوده خطر لغزش بالا قرار دارد (Shadfar *et al.*, 2011). همچنین در پهنه‌بندی خطر زمین‌لغزش با استفاده از روش‌های SMCE و AHP در آبخیز هفشجان چهارمحال و بختیاری، فاکتورهای فاصله از جاده، گسل و آبراهه مهمترین عوامل بودند (Ebrahimi *et al.*, 2017).

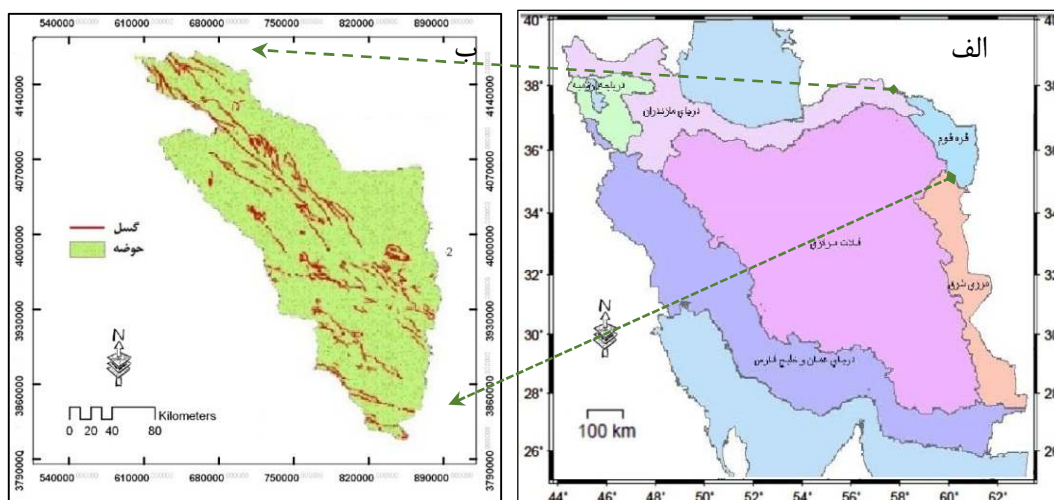
ایران یکی از زلزله‌خیزترین کشورهای دنیا است و شهرهای آن در رابطه با این پدیده طبیعی آسیب‌های فراوان دیده است. ایران دارای ۱۴۰۰ گسل فعال است که از سال ۱۹۳۶ میلادی باعث بروز ۴۶۰ زلزله تاریخی و ثبت ۶۵۰۰ زلزله شده است. مطالعه گسل‌ها از آنجاکه توان لرزه‌زایی دارند، دارای اهمیت است. شناسایی زون‌های گسلی فعال و برآورد توان لرزه‌زایی این گسل‌ها می‌تواند در کاهش خسارت‌های جانی و مالی مهم باشد. بی‌شک بین گسل و زلزله رابطه نزدیکی برقرار است و قسمت اعظم زلزله‌ها روی گسل‌های قدیمی متمرکزند. در مناطقی همانند حوضه

مقادیر TRI نیز بر مبنای روش ارائه شده توسط Riley و همکاران (۱۹۹۹) و مدول‌های GIS قابل استخراج است.

$$\text{SPI} = \ln(As * \tan\sigma) \quad \text{رابطه ۱}$$

مبنای تهیه این لایه‌ها مدل رقومی ارتفاع (DEM: Digital Elevation Model) با دقت پیکسل ۳۰ متر است که طبق لایه شبکه آبراه‌های منطقه، اقدام به مرزبندی حوضه و برش زدن لایه‌ها با این مرز شد. لایه گسل‌ها طبق بانک اطلاعات گسل‌های کشور تهیه و رس‌تری شد و کاربری اراضی حوضه نیز طبق لایه کشوری تهیه شده توسط سازمان جنگل‌ها، مراتع و آبخیزداری کشور استخراج گردید. مبنای تهیه لایه خاک‌شناسی نیز نقشه منابع و استعداد خاک‌های ایران می‌باشد که در آن واحدهای خاک تا حد زیرگروه مشخص گردیده است (Jamshidi et al., 2015).

جهت، ارتفاع، کلاس فرسایشی (با روش MPSIAC: Modified Pacific Southwest Inter Agency Committee)، خاک‌شناسی، اقلیم (با روش آمبرژه)، واحد زمین‌شناسی، تیپ ژئومورفولوژی (تیپ‌های نه‌گانه استاندارد فائو)، فاصله از جاده، فاصله از آبراهه، فاصله از طاق‌دیس، فاصله از ناودیس، کلاس بارشی، پهنه‌های سیلابی، شاخص ناهمواری زمین (TRI: Terrain Ruggedness Index) و شاخص قدرت جریان (SPI: Stream Power Index) در محیط نرم‌افزاری ArcGIS به صورت رس‌تری با اندازه پیکسل ۳۰ متر تهیه شد. TRI اختلاف در مقادیر ارتفاع یک سلول مرکزی و هشت سلول احاطه‌کننده آن (Sarp, 2014) و SPI حاصل ضرب سطح حوضه در تانژانت زاویه شیب آن می‌باشد. رابطه ۱ برای محاسبه SPI پیشنهاد شده است (Wilson & Gallant, 2000) که در آن SPI شاخص قدرت جریان، As مساحت ویژه حوضه (متر) و σ شیب بر حسب درجه می‌باشد.



شکل ۱- موقعیت حوضه قره‌قوم در کشور (الف) و گسل‌های آن (ب)

شامل کاربری اراضی، شیب، جهت، ارتفاع، کلاس فرسایشی، خاک‌شناسی، اقلیم، واحد زمین‌شناسی، تیپ ژئومورفولوژی، کلاس بارشی، فاصله از جاده و آبراهه، فاصله از طاق‌دیس و ناودیس، پهنه‌های سیلابی و شاخص‌های ناهمواری زمین و قدرت جریان، تبدیل به بانک اطلاعاتی گردید. به این ترتیب که به تعداد نقاط

اجرای الگوریتم‌های داده‌کاوی

بخشی از پژوهش به ارزیابی الگوریتم‌های طبقه‌بندی در تفکیک مناطق غیرحساس و حساس به گسل بر اساس متغیرهای پایه حوضه اختصاص یافت. طبق اطلاعات ادغام لایه‌ها که در محیط ArcGIS انجام شد، متغیرهای وابسته وجود یا نبود گسل و متغیرهای مستقل

روش تراکم سطح نیز از طریق رابطه ۳، برای هر پارامتر ارزش طبقات محاسبه گردید (Luca *et al.*, 2015).

$$W_i = \text{Ln} \left[\frac{\text{Densclass}}{\text{Densmap}} \right] = \text{Ln} \left[\frac{\frac{Npix(S_i)}{Npix(N_i)}}{\frac{\sum Npix(S_i)}{\sum Npix(N_i)}} \right] \quad \text{رابطه ۲}$$

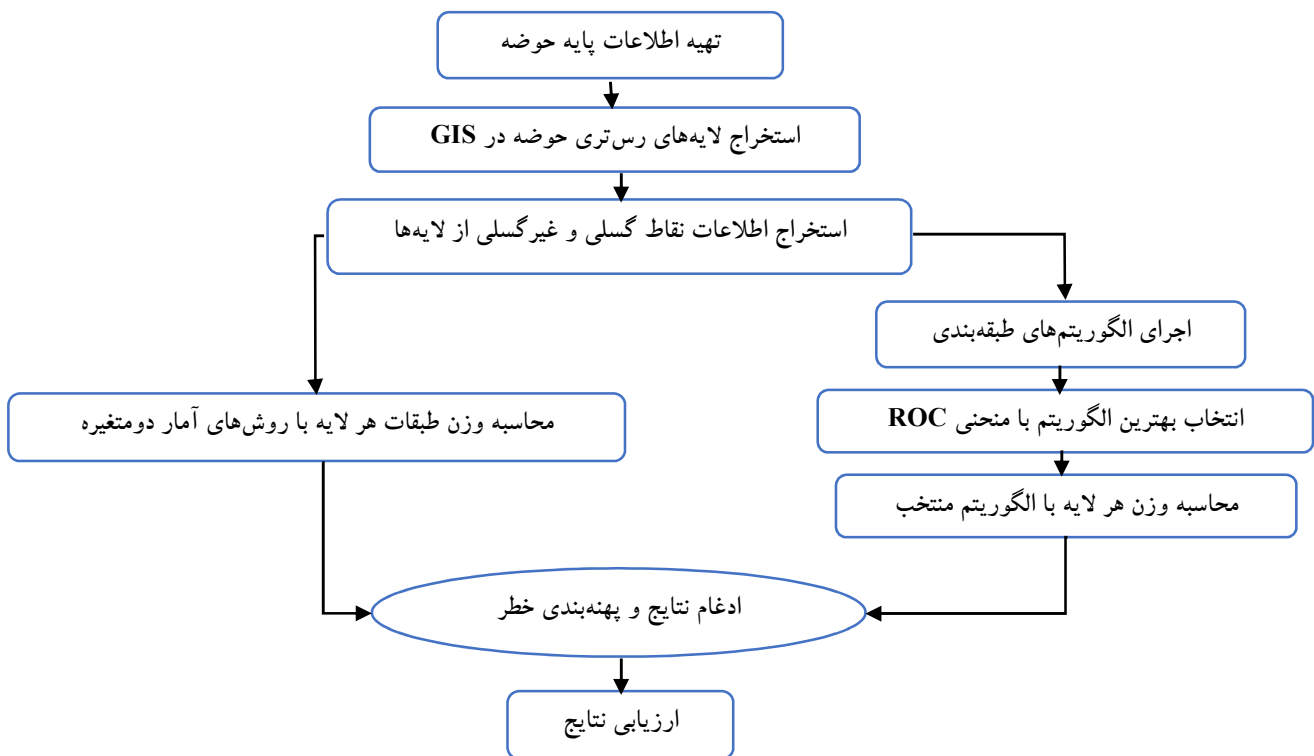
$$W_i = 1000 \left[\frac{NpixSi}{NpixNi} - \frac{\sum NpixSi}{\sum NpixNi} \right] \quad \text{رابطه ۳}$$

در این روابط، W_i : وزن هر طبقه، Densclass: تراکم گسل‌های هر طبقه، Densmap: تراکم کل گسل‌های منطقه، $NpixSi$: تعداد پیکسل گسل‌های هر طبقه، $NpixNi$: تعداد پیکسل‌های هر طبقه، $\sum NpixSi$: مجموع کل پیکسل گسل‌های موجود در منطقه و $\sum NpixNi$: مجموع کل پیکسل‌های منطقه است. پس از محاسبه وزن هر یک از طبقات متغیرها، در زیرمجموعه‌های لایه اطلاعاتی مربوطه، با ایجاد یک فیلد جدید، وزن وارد گردید. سپس با ابزار Spatial Analyst بر اساس وزن‌های وارد شده، یک لایه رس‌تری ایجاد شد. لایه‌های رس‌تری تهیه شده با ابزار Raster Calculator و با توجه به وزن‌های هر طبقه که از دو روش آماری دومتغیره به دست آمده و نیز وزن هر متغیر (لایه) که از الگوریتم برتر داده‌کاوی به دست آمده، با یکدیگر ادغام و لایه پهنه‌بندی خطر ایجاد شد. در پایان با ابزار Reclassify، پنج طبقه احتمال خطر وقوع گسل با استفاده از روش شکستگی‌های طبیعی (Natural Breaks) استخراج شد. در شکل ۲ نمودار جریانی روش اجرای این پژوهش ارائه شده است.

گسلی، نقاط فاقد گسل به صورت کاملاً تصادفی ایجاد و اطلاعات تمامی این نقاط از لایه‌های متغیرهای مستقل در ArcGIS استخراج و وارد محیط Excel شد. در ادامه فرایند تحلیل، متغیرهای کیفی کدگذاری شده و بعد داده‌ها به دو دسته آموزشی و آزمون تقسیم شدند، یعنی ۷۰٪ داده‌ها به آموزش و ۳۰٪ باقی‌مانده به آزمون مدل اختصاص یافت (Esfandiyari-Darabadi & Beheshti-Javid, 2016). در نهایت الگوریتم‌های طبقه‌بندی شامل ماشین بردار پشتیبان (Support Vector Machine)، جنگل تصادفی (Random Forest)، درخت تصمیم (Decision Tree)، رگرسیون لوجستیک (Logistic)، شبکه عصبی مصنوعی (Artificial Neural Network) و تجمعی بوستینگ (Boosting Aggregate) در نرم‌افزار R و افزونه Rattle بر روی داده‌ها برازش شد. ارزیابی الگوریتم‌ها با منحنی مشخصه عملیاتی دریافت‌کننده (ROC: Receiver Operating Characteristic) و بر مبنای سطح زیرمنحنی (Area Under the Curve) انجام شد. ایده‌آل‌ترین مدل، بیشترین AUC را دارد که از نیم تا یک متغیر است و هر چه به یک نزدیک‌تر باشد، بیانگر دقت بهتر تفکیک و طبقه‌بندی است (Pourghasemi *et al.*, 2014).

پهنه‌بندی خطر وقوع گسل

پهنه‌بندی خطر گسل و ارزیابی دقت پهنه‌بندی بخش دیگر پژوهش بود که با توجه به ارزش نسبی هر متغیر در محیط ArcGIS اقدام به پهنه‌بندی بر اساس برآیند ویژگی‌های هر پیکسل شد. برای محاسبه ارزش نسبی طبقات هر لایه، از روش آماری دو متغیره ارزش اطلاعات (Information Value) و تراکم سطح (Density Area) استفاده شد. در روش ارزش اطلاعات، با استفاده از رابطه ۲ وزن هر کلاس محاسبه شد. در

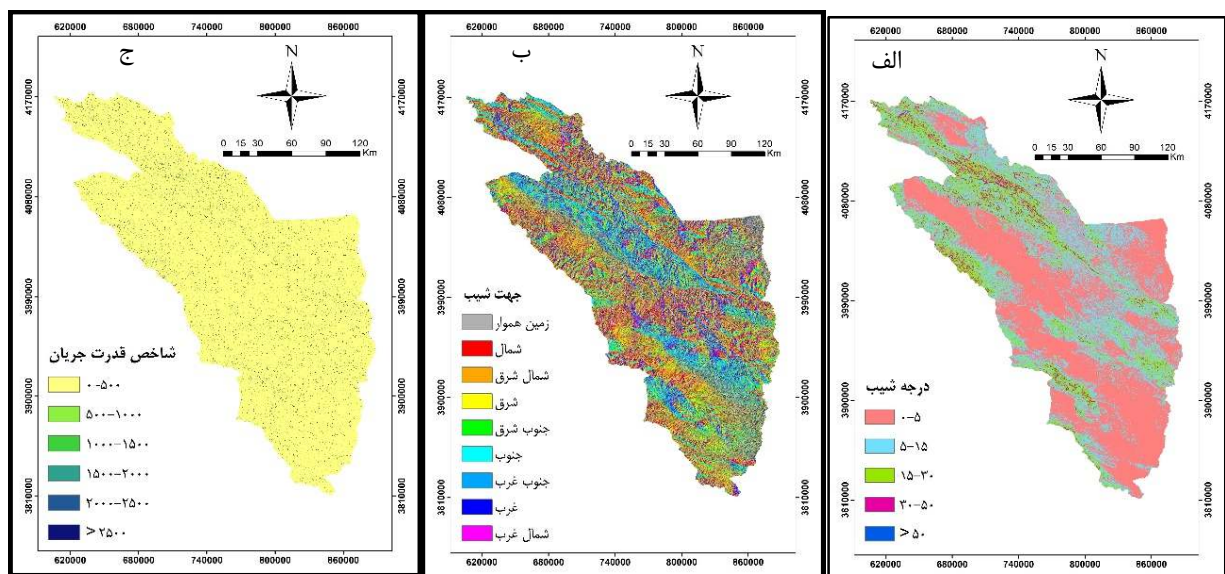


شکل ۲- نمودار جریانی مراحل انجام پژوهش

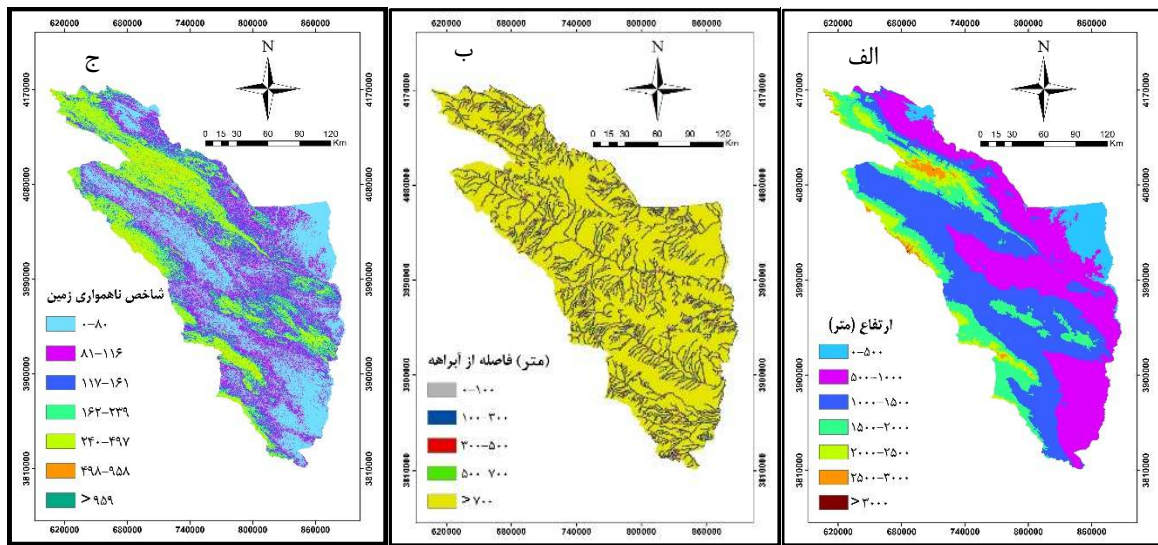
نتایج

جدول ۱ نیز ارزش نسبی طبقات با بالاترین ارزش در هر لایه اطلاعاتی را با دو روش ارزش اطلاعات و تراکم سطح ارائه نموده‌است.

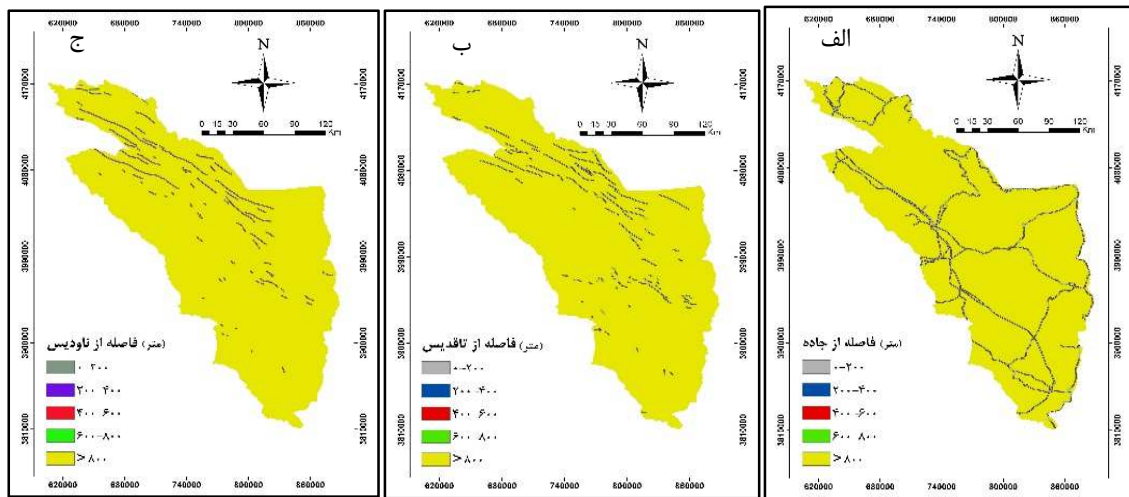
لایه‌های اطلاعاتی استخراج شده حوضه قره‌قوم طبق روش تحقیق، در شکل‌های شماره ۳ تا ۸ ارائه گردیده است.



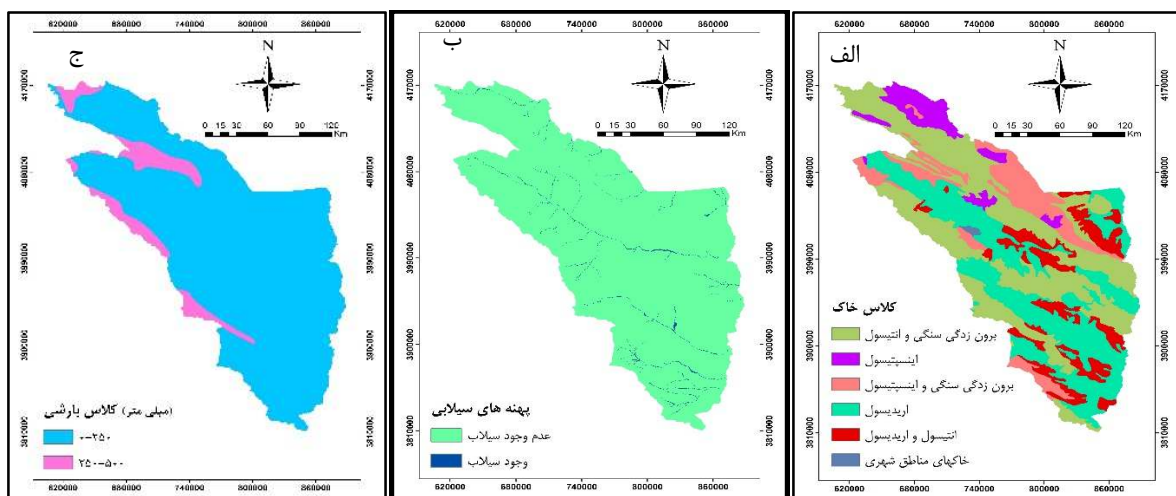
شکل ۳- لایه‌های شیب (الف)، جهت (ب) و شاخص قدرت جریان (ج)



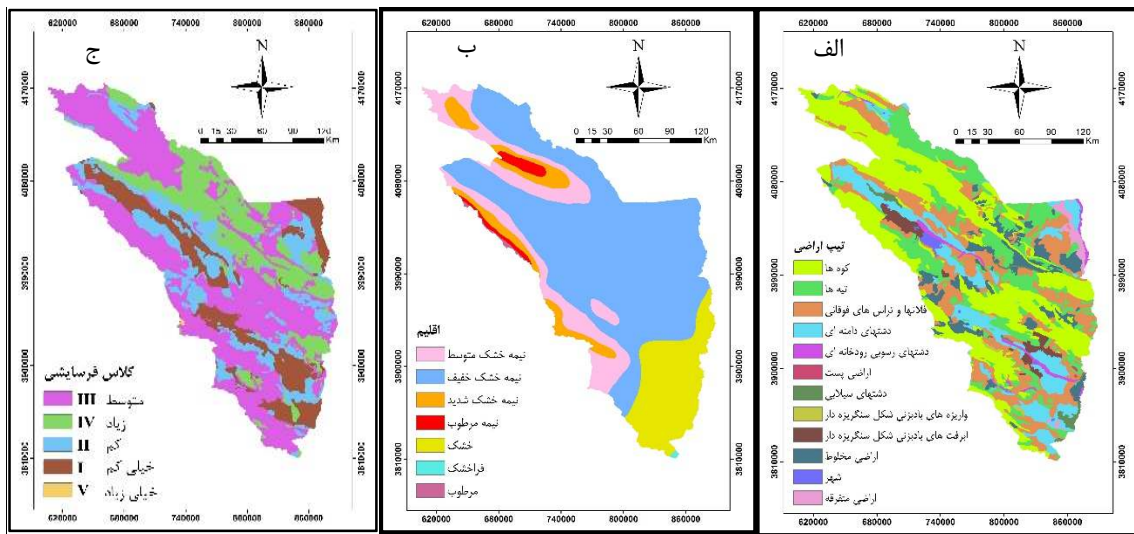
شکل ۴- لایه‌های ارتفاع (الف)، فاصله از آبراهه (ب) و شاخص ناهمواری زمین (ج)



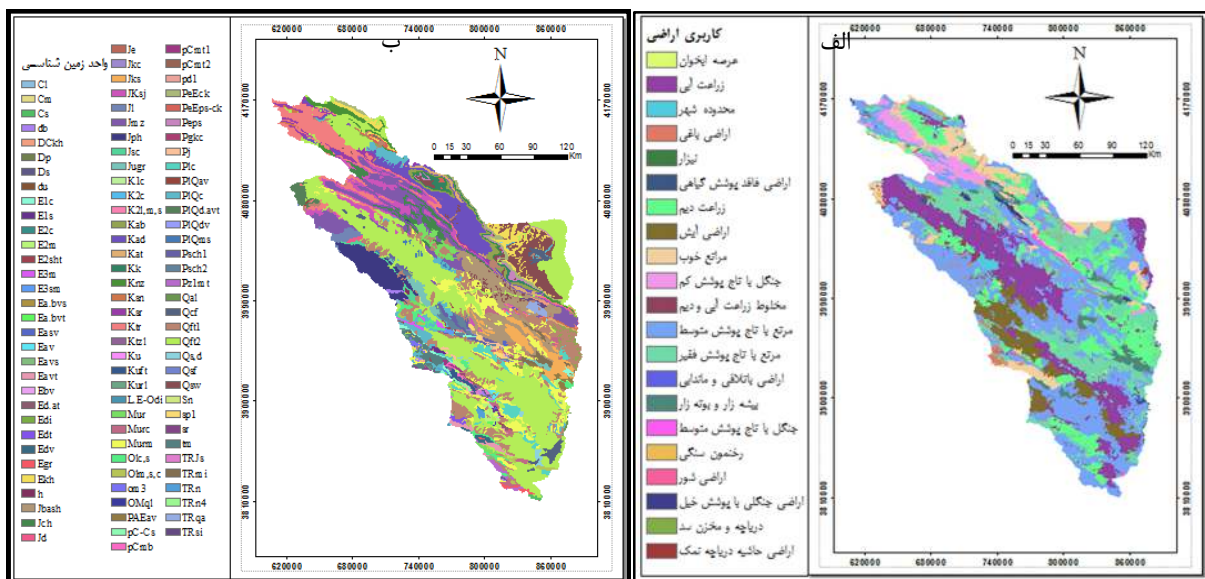
شکل ۵- لایه‌های فاصله از جاده (الف)، فاصله از تاق‌دیس (ب) و فاصله از ناودیس (ج)



شکل ۶- لایه‌های خاک‌شناسی (الف)، پهنه‌های سیلابی (ب) و کلاس بارش (ج)



شکل ۷- لایه‌های ژئومورفولوژی (الف)، اقلیم (ب) و کلاس فرسایش (ج)



شکل ۸- لایه‌های کاربری اراضی (الف) و زمین‌شناسی (ب)

بهترین عملکرد را در طبقه‌بندی وجود یا عدم وجود گسل دارند. شکل ۹ نتایج ارزیابی الگوریتم‌ها با منحنی ROC را نشان می‌دهد. این منحنی یک نمودار گرافیکی است که توان تشخیص یک سیستم طبقه‌بندی دوحالتی را نمایش می‌دهد.

ارزیابی دقت طبقه‌بندی الگوریتم‌ها با کاربرد داده‌های آزمون، نتایج ارزیابی الگوریتم‌ها با منحنی ROC نشان داد که الگوریتم‌های جنگل تصادفی و ماشین‌بردار پشتیبان با سطح زیرمنحنی ۸۸ و ۸۶ درصد،

جدول ۱- طبقات دارای بالاترین ارزش در لایه‌های مختلف مورد بررسی

وزن طبقه		مساحت (هکتار)	طبقه	متغیر
تراکم سطح	ارزش اطلاعات			
۷۱/۱	۲/۴۳	۹۴۷/۰۱	اراضی باتلاقی و ماندابی	کاربری اراضی
۴۳/۱	۱/۹۸	۵۹/۶۰	۹۵۹ <	شاخص ناهمواری زمین
۵/۱	۰/۵۵	۱۴۴۱۳/۳۸	۱۵۰۰-۲۰۰۰	شاخص قدرت جریان
۴/۱	۱/۴۱	۵۸۵۰۹۸/۱۲	شمال	جهت شیب
۰/۹	۰/۱۲	۱۳۷۲۲۰۳/۰۲	۵-۱۵	درجه شیب
۸/۱	۰/۷۷	۵۳۶۰۵/۳۰	۲۵۰۰-۳۰۰۰	ارتفاع (متر)
۲/۲	۰/۲۸	۱۹۷۰۰۶۲/۳۵	متوسط (III)	کلاس فرسایش
۳/۱	۰/۳۷	۱۶۷۲۷۹۱/۰۵	برون‌زدگی سنگی و اتنی‌سول	کلاس خاک
۳۳/۱	۱/۷۶	۳۸۶۴/۳۷	فراخشک	اقلیم
۱۱۳/۱	۲/۸۶	۲۷۶/۸۱	Qsf (شوره‌زار)	واحد زمین‌شناسی
۴۸/۱	۲/۰۸	۲۶۹۱/۳۱	اراضی پست	تیب ژئومورفولوژی
۰/۲	۰/۰۳	۴۰۲۹۸۷۸/۲۲	۸۰۰ <	فاصله از جاده (متر)
۰/۳	۰/۰۴	۲۹۴۸۴۰۴/۲۸	۷۰۰ <	فاصله از آبراهه (متر)
۶۲/۱	۲/۳۰	۴۲۳۵۲۶/۴۳	۸۰۰ <	فاصله از طاق‌دیس (متر)
.	.	۴۱۸۱۳/۲۲	۶۰۰-۸۰۰	فاصله از ناودیس (متر)
۱/۲	۰/۱۶	۳۸۶۱۸۰/۴۴	۲۵۰-۵۰۰	کلاس بارش (میلی‌متر)
.	.	۴۷۸۶۳/۲۵	وجود سیلاب	بهنه‌های سیلابی

حقیقی در داده‌ها می‌باشد. FPR نیز تحت عنوان احتمال هشدار غلط یا رخداد یا نرخ ورودی‌های نامرتب نامیده می‌شود که برابر با نسبت خطای مثبت به شرایط منفی یا تعداد موارد منفی حقیقی در داده‌ها می‌باشد. سطح زیر منحنی طبق رابطه ۴ به‌دست می‌آید (Fawcett, 2006).

این نمودار با پلات کردن نرخ مثبت صحیح (TPR: True Positive Rate) در مقابل نرخ خطای مثبت (FPR: False Positive Rate) به‌دست می‌آید که TPR در یادگیری ماشین تحت عنوان حساسیت یا احتمال تشخیص نامیده می‌شود و برابر با نسبت مثبت صحیح به شرایط مثبت یا تعداد موارد مثبت

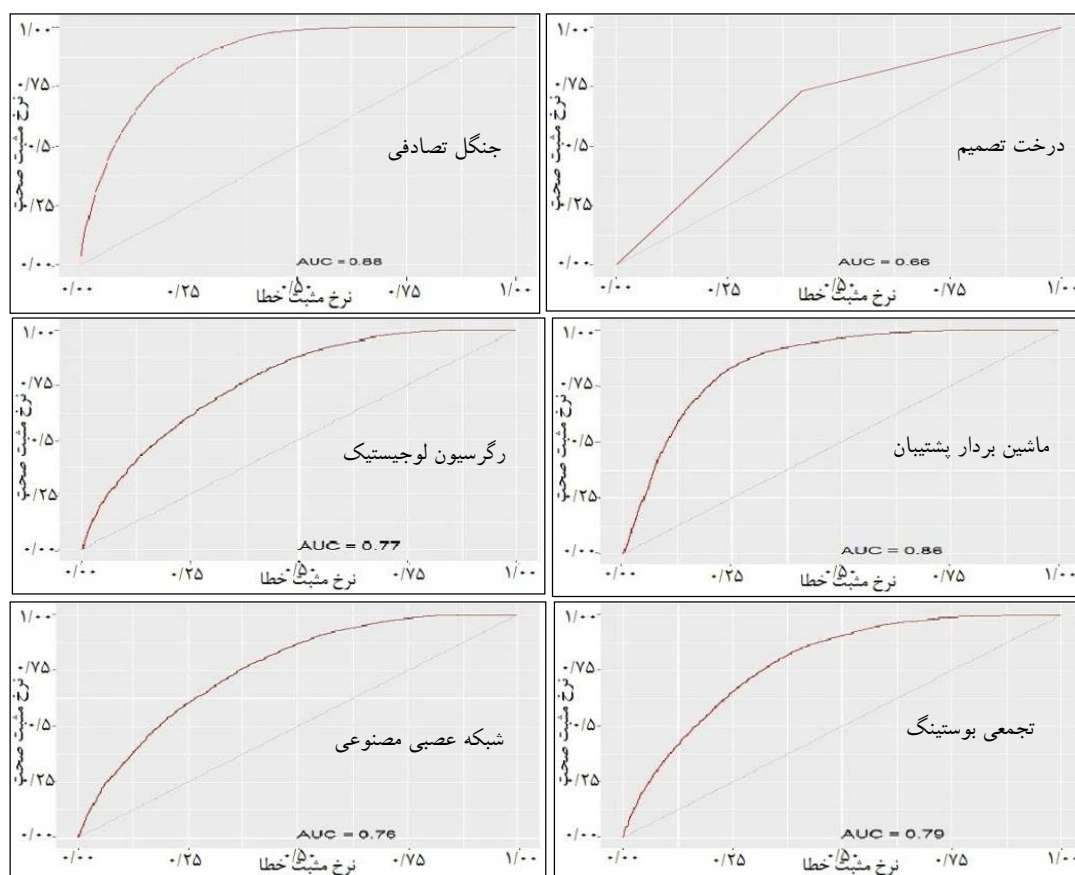
$$AUC = \int_{x=0}^1 TPR(FPR^{-1}(x))dx \quad \left\{ \begin{array}{l} TPR(T): T \rightarrow y(x) \\ FPR(T): T \rightarrow x \end{array} \right. \quad \text{رابطه ۴}$$

متغیرهای ورودی بر اساس شاخص جینی (Gini Index) و شاخص صحت (Accuracy Index) نشان می‌دهد. در هر دو

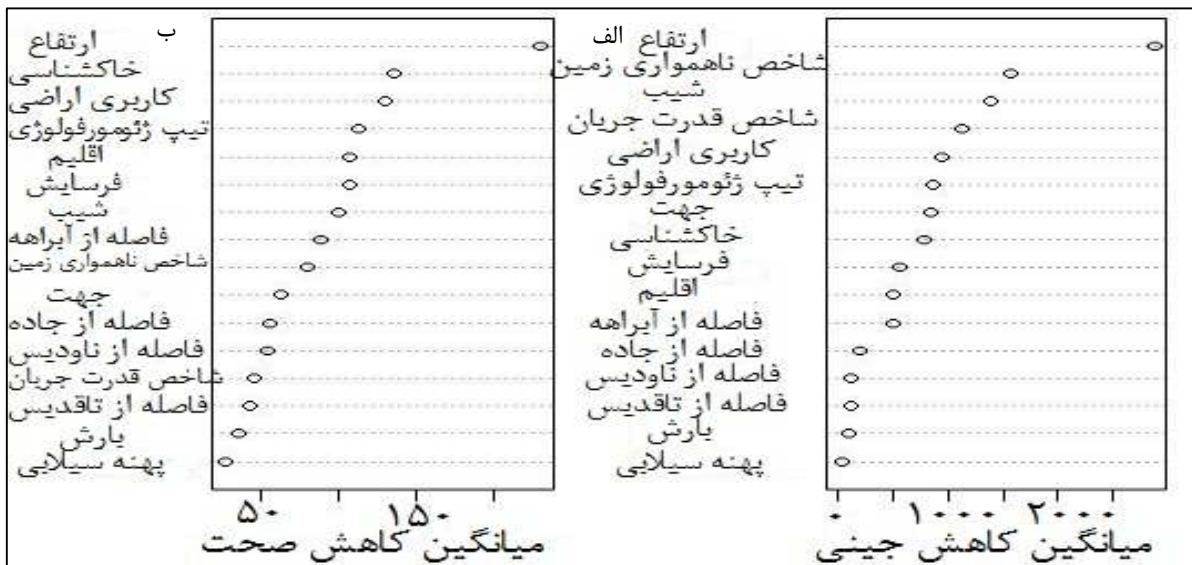
با توجه به انتخاب الگوریتم جنگل تصادفی به‌عنوان برترین الگوریتم، شکل ۱۰ نتایج این الگوریتم را برای اهمیت

در درختان مختلف، مجموع‌گیری و در انتهای محاسبه نرمال می‌گردد. میانگین کاهش صحت نیز هر چه با کنار گذاشتن یک متغیر خاص، کاهش بیشتری پیدا کند، فرض می‌شود که آن متغیر اهمیت بالاتری دارد. از این رو متغیرهای با میانگین کاهش صحت بالاتر، برای طبقه‌بندی داده‌ها اهمیت بالاتری دارند. این ضریب برابر با نسبت میانگین کاهش در صحت درختان جنگل تصادفی به انحراف معیار آن می‌باشد که به صورت درصد یا فراوانی مشاهدات بیان می‌گردد (Hur et al., 2017). سپس با استفاده از نقشه پراکندگی گسل‌ها (۳۰ درصد)، اقدام به ارزیابی و مقایسه روش‌های پهنه‌بندی خطر شد. برای انجام این امر، نقشه پراکنش با نقشه‌های پهنه‌بندی خطر گسل در سیستم اطلاعات جغرافیایی قطع داده شدند که نتایج در جدول ۲ و شکل ۱۲ ارائه شده است.

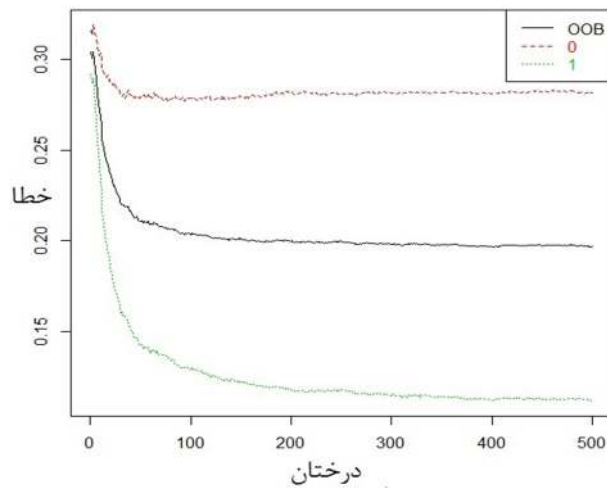
شاخص، متغیر ارتفاع به‌عنوان مهمترین متغیر شناخته‌شد. سپس متغیر خاک‌شناسی برای شاخص صحت و شاخص ناهمواری زمین برای شاخص جینی به ترتیب به‌عنوان مهمترین پارامترهای مؤثر در وقوع گسل شناخته شدند. شکل ۱۱ نیز میزان تخمین نمونه‌های خارج از سید (OOB) را با افزایش تعداد درختان الگوریتم نشان می‌دهد. با توجه به دقت بالاتر الگوریتم جنگل تصادفی نسبت به سایر الگوریتم‌ها، نقشه حساسیت وقوع گسل با استفاده از ضرایب جینی در این الگوریتم تهیه شد. میانگین کاهش جینی ضریبی است که نشان می‌دهد چگونه هر متغیر به همگنی گره و برگ‌ها در جنگل تصادفی نهایی کمک می‌کند. هر بار یک متغیر خاص برای تفکیک گره‌ها استفاده شده و ضریب جینی برای گره‌های پایین‌دستی محاسبه می‌گردد. ضریب جینی مقیاسی از (همگنی) تا ۱ (ناهمگنی) دارد و تغییرات جینی برای هر متغیر



شکل ۹- منحنی ROC برای پیش‌بینی وقوع گسل با الگوریتم‌های مختلف



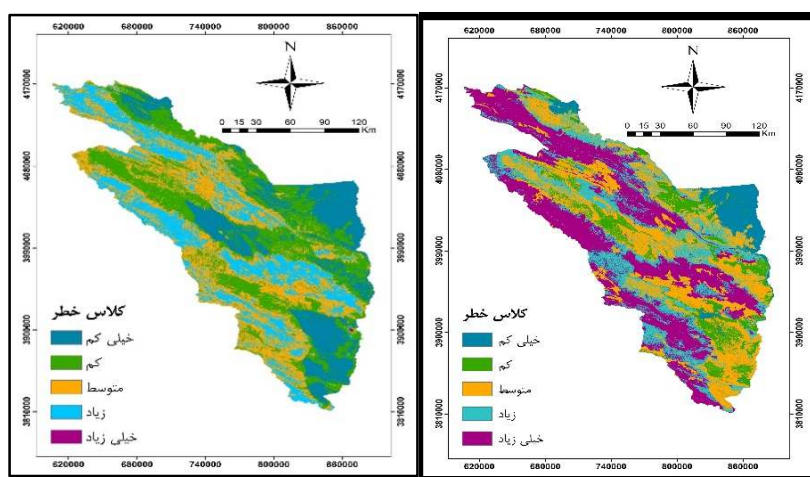
شکل ۱۰- ارزش متغیرها در الگوریتم جنگل تصادفی با شاخص‌های جینی (الف) و صحت (ب)



شکل ۱۱- تغییرات میزان خطا با افزایش تعداد درختان الگوریتم

جدول ۲- درصد طبقات در پهنه‌بندی حساسیت وقوع گسل در روش ارزش اطلاعات و تراکم سطح

کلاس خطر	ارزش اطلاعات			تراکم سطح		
	وسعت (هکتار)	طبقه (درصد)	گسل (درصد)	وسعت (هکتار)	طبقه (درصد)	گسل (درصد)
خیلی کم	۲۸۱۴۸۶	۶/۳۸	۰/۲۷	۸۸۲۸۴۱	۲۰/۰۱	۵/۷۶
کم	۴۶۹۸۷۸	۱۰/۶۵	۳/۴۸	۱۳۲۲۲۷۶	۲۹/۹۷	۱۸/۱۱
متوسط	۱۲۲۳۸۸۹	۲۷/۷۴	۲۲/۱۸	۱۲۲۵۲۱۲	۲۷/۷۷	۳۴/۹۰
زیاد	۱۰۱۳۸۷۸	۲۲/۹۸	۲۱/۹۱	۹۸۰۷۸۸	۲۲/۲۳	۴/۹۷
خیلی زیاد	۱۴۲۲۸۷۰	۳۲/۲۵	۵۲/۱۶	۸۸۲	۰/۰۲	۰/۲۶



شکل ۱۲- پهنه‌بندی خطر گسل به روش تراکم سطح (الف) و ارزش اطلاعات (ب)

برای روش‌های ارزش اطلاعات و تراکم سطح واقع شدند، اما ناودیس هیچ تأثیری در وقوع گسل منطقه ندارد. طبق نتایج حاصل از کاربری اراضی، بیشترین وزن متعلق به اراضی باتلاقی و ماندابی می‌باشد. همچنین اقلیم فراخشک بیشترین وزن را داراست و از لحاظ تیپ اراضی، بیشترین وزن‌ها به اراضی پست اختصاص داده شده است. بررسی واحدهای سنگ‌شناسی نشان داد که بیشترین وزن برای روش تراکم سطح مربوط به شوره‌زار (Qsf) به میزان ۱۱۳/۱ و روش ارزش اطلاعات مربوط به توف ریولیتیک به ردوکیتیک (Edt) به میزان ۳۳/۱ می‌باشد. در نهایت کلاس فرسایش متوسط و کلاس خاک دارای برون‌زدگی سنگی و انتی‌سول بیشترین وزن را از نظر وقوع گسل به خود اختصاص دادند.

بحث

با توجه به نتایج بررسی رابطه متغیرها با وقوع گسل در رگرسیون لجستیک ($p < 0.05$)، مشخص شد که وقوع گسل با عواملی همانند شیب، جهت (شمال‌شرق)، شاخص ناهمواری زمین، فاصله از ناودیس (۲۰۰-۴۰۰ و ۴۰۰-۶۰۰ متری)، فاصله از طاق‌دیس (۶۰۰-۸۰۰ متری)، فاصله از آبراهه (۱۰۰-۳۰۰، ۳۰۰-۵۰۰، ۵۰۰-۷۰۰ و > 700 متری)،

با توجه به وزن‌های به‌دست آمده در هر طبقه بر اساس لایه طبقات ارتفاعی در روش ارزش اطلاعات و تراکم سطح می‌توان گفت در هر دو روش، کلاس ارتفاعی ۲۵۰۰ تا ۳۰۰۰ متر با مقادیر به ترتیب ۰/۷۷ و ۸/۱ بیشترین وزن را به خود اختصاص داده است. نتایج بررسی عوامل شیب و جهت دامنه نشان داد که شیب پنج تا ۱۵ درجه با مقادیر ۰/۱۲ و ۰/۹ برای ارزش اطلاعات و تراکم سطح و جهت شمال با مقادیر ۱/۴۱ و ۴/۱ در دو روش اخیر تأثیر بیشتری نسبت به کلاس‌های دیگر شیب و جهت در وقوع گسل دارند. تراکم گسل‌ها در طبقات با مقادیر بالای شاخص‌های قدرت جریان (با وزن ۰/۵۵ و ۵/۱ در ارزش اطلاعات و تراکم سطح) و ناهمواری زمین بیشتر از سایر طبقات بوده است. بررسی عامل بارش در منطقه نشان داد که بیشترین محل تشکیل گسل‌ها در کلاس ۲۵۰ تا ۵۰۰ میلی‌متر بارندگی با مقادیر ۰/۱۶ و ۱/۲ برای ارزش اطلاعات و تراکم سطح می‌باشد. بررسی نتایج فاصله از جاده و آبراهه، فاصله از طاق‌دیس و فاصله از ناودیس نشان داد که بیشترین وزن گسل‌ها در فواصل بیشتر از ۷۰۰ متری آبراهه (مقادیر ۰/۰۴ و ۰/۳ در ارزش اطلاعات و تراکم سطح) و ۸۰۰ متری جاده (مقادیر ۰/۰۳ و ۰/۲ در ارزش اطلاعات و تراکم سطح) و طاق‌دیس با مقادیر ۲/۳ و ۶۲/۱

متوسط و زیاد، تیپ دشت‌های سیلابی، اراضی پست، آبرفت‌های بادبزی سنگریزه‌دار و دشت‌های رودخانه‌ای و کاربری آیش، اراضی فاقد پوشش، مراتع خوب و شهری و واحدهای زمین‌شناسی trsqai .pj, pcmb, le.odi و trsi و خاک‌های انتیسول/اریدیسول، اینسپتیسول و انتیسول/برون‌زدسنگی در سطح اعتماد ۰/۹۹۹ معنادار است. همچنین با متغیرهای وجود پهنه‌های سیلابی، تیپ واریزه‌های بادبزی شکل سنگریزه‌دار، دشت‌های دامنه‌ای، واحدهای زمین‌شناسی ea.bvt, e3sm, e3m, elc, dckh, pc.cs, olm.sc, jsc, jch, edt و psch2, psch1, trmi و خاک مناطق شهری با اعتماد ۰/۹۹ رابطه معناداری دارد. در نهایت در سطح اعتماد ۰/۹۵ با متغیرهای مناطق مسطح، جهت شمال‌شرق و جنوب‌غرب، شاخص قدرت جریان، فاصله از طاق‌دیس بالای ۸۰۰ متر، کاربری زراعت دیم و زراعت ترکیبی آبی و دیم، واحدهای زمین‌شناسی e2m, jugr, jmjz, jks, jbash, edv, ed.at, eavt, eav, e2sht, olc.s, murc, mur, ktzl, ksr, ksn, k2l.ms, klc و رابطه معناداری نشان داد. Lee (۲۰۰۴)، Ayalew و همکاران (۲۰۰۵) و Lee و Sambath (۲۰۰۶) مدل رگرسیون لجستیک را برای مطالعه زمین‌لغزش مناسب معرفی نموده‌اند، در حالی‌که در بین شش الگوریتم مورد بررسی در این پژوهش، این مدل از نظر دقت طبق نتایج ارائه شده در شکل ۹، در جایگاه چهارم قرار گرفت. زیرا الگوریتم جنگل تصادفی به‌عنوان روشی جدید معرفی و اخیراً در مطالعات زمین‌لغزش استفاده شده است. ولی نسبت به روش‌های قبلی کارایی مناسب‌تری را از خود نشان داده‌است (Andre & Norman, 2010).

الگوریتم برتر جنگل تصادفی، اهمیت متغیرهای ورودی را بر اساس شاخص‌های جینی و صحت ارائه نمود. در هر دو شاخص، متغیر ارتفاع به‌عنوان مهمترین ورودی به‌دست آمد. سپس متغیر خاک‌شناسی برای شاخص صحت و متغیر ناهمواری زمین طبق شاخص جینی به‌ترتیب به‌عنوان مهمترین پارامترهای مؤثر در پهنه‌بندی وقوع گسل شناخته شدند. در حالی‌که Komac (۲۰۰۶)، Parmar و همکاران

کلاس بارشی ۲۵۰-۵۰۰ میلی‌متر، اقلیم (نیمه‌خشک شدید، نیمه‌خشک متوسط، نیمه‌خشک خشک، مرطوب و نیمه‌مرطوب)، کلاس فرسایشی خیلی شدید (V)، قابلیت اراضی (تیپ آبرفت‌های بادبزی شکل سنگریزه‌دار، تیپ دشت‌های رسوبی رودخانه‌ای و تیپ اراضی متفرقه)، کاربری اراضی (اراضی باغی، زراعت دیم، اراضی آیش، جنگل با تاج‌پوشش کم و متوسط، دریاچه و مخزن سد، مراتع با تاج‌پوشش فقیر، مراتع خوب، رخنمون سنگی، اراضی شور، اراضی جنگلی با پوشش خیلی کم، اراضی حاشیه دریاچه، محدوده شهری و بیشه‌زار و بوته‌زار)، واحدهای زمین‌شناسی (شامل واحدهای cm: سنگ آهک تیره با شیل زیرین، dp: کوارتز آرنیت با گچ، ea.bvc: رسوبات بازالت و آندزیتی، easv: آندزیت آتشفشانی، eavs: رسوبات آندزیتی، ebv: بازالت آتشفشانی، edi: دیوریت، egr: گرانیت، ekh: ماسه‌سنگ و شیل، h: سنگ‌های دگرگونی، jd: سنگ آهک و شیل، ji: سنگ آهک روشن، kad: سنگ آهک خاکستری، ku: سنگ‌های نامتعارف کرتاسه، kurl: سنگ آهک و چرت، om3: سنگ آهک، چرت و شیل با گدازه، PAEav: آتشفشانی آندزیتی، pd1: سنگ الترابازیک، peEck: سنگ آهک و مارن گچی، peEpc-ck: ماسه‌سنگ، کنگلومرا و آهک و مارن، peps: کنگلومرا، ماسه‌سنگ و خاکستر، pgkc: کنگلومرا و ماسه‌سنگ درشت، plQav: آتشفشان آندزیتی، plqd.avt: توف آندزیت-داکتیتی، plQdv: آتشفشان ریولیتی، qs.d: رسوبات بادی، qsw: باتلاق، sn: شیل، ماسه‌سنگ و سنگ آهک، sr: سرپنتینیت و tm4: سنگ آهک سیاه، شیل و ماسه‌سنگ) و کلاس خاک (رخنمون‌سنگی/اینسپتیسول و مناطق شهری) رابطه منفی دارد و احتمال وقوع گسل با وجود این عوامل کاهش می‌یابد. اما ایجاد گسل در منطقه با سایر عوامل رابطه مثبت دارد و بیانگر این است که احتمال وقوع گسل در این ترکیب از ویژگی‌های محیطی افزایش می‌یابد. این رابطه با متغیرهای ارتفاع، جهت شمال، شمال‌غرب، جنوب‌شرق و غرب، شاخص ناهمواری زمین، اقلیم فراخشک، نیمه‌خشک شدید، متوسط، خشک و نیمه‌مرطوب، کلاس فرسایشی کم،

گسل‌هایی که به سطح زمین نمی‌رسند، گسل‌های نهان یا گسل‌های کور (Blind Faults) هستند که می‌توانند منشأ مهم برخی فعالیت‌های لرزه‌ای باشند. رابطه گسل و زلزله دوطرفه است. وجود گسل‌های زیاد در منطقه موجب بروز زلزله جدید است. زلزله گسل جدیدی را بوجود می‌آورد و در نتیجه تعداد شکستگی‌ها زیادتر شده و قابلیت زلزله‌زایی منطقه افزایش می‌یابد. در این پژوهش نیز تلاش گردید تا مناطق مستعد برای بروز گسل‌های کور شناسایی گردد، زیرا کاهش پیامدهای ناگوار زلزله نیازمند طراحی برنامه‌هایی برای کاهش میزان آسیب‌پذیری شهرها در برابر آن است که لازمه این برنامه‌ها اجرای مطالعاتی از این دست با هدف شناسایی مناطق پرخطر است. با توجه به نتایج پژوهش، خطر بالای منطقه برای وقوع گسل و زلزله‌های متعدد به‌وقوع پیوسته در آن، طراحی و اجرای طرح‌های توسعه‌ای حوضه دقت بالایی را می‌طلبد. چون اصولاً گسل‌ها دارای توان لرزه‌ای نهفته هستند و در صورت رهاسازی انرژی ذخیره‌ای، زلزله‌های متعدد به وجود می‌آورند. طبق نتایج، با توجه به اهمیت عواملی همانند ارتفاع از سطح دریا، ناهمواری زمین و خاک‌شناسی حوضه در شناسایی مناطق مستعد، ویژگی‌های طبیعی و به‌ویژه توپوگرافی اهمیت خود را در وقوع گسل نشان می‌دهند. از این رو برای اجرای یک مدیریت پایدار و کارا در حوضه، در نظر گرفتن نقش این عوامل در هر برنامه توسعه‌ای ضرورت دارد و لازم است قبل از اعمال هر گونه تغییر در حوضه، پیامدهای آن بر پدیده گسل به‌صورت جامع ارزیابی گردد.

سپاسگزاری

این پژوهش به شماره ۳۵ با استفاده از اعتبارات پژوهشی دانشگاه تربیت‌حیدریه انجام شده است، از این رو از مسئولان محترم دانشگاه تشکر و قدردانی می‌گردد.

(۲۰۱۲) و Papadakis و Karimalis (۲۰۱۷) عامل شیب، Ebrahimi و همکاران (۲۰۱۷) عامل فاصله از جاده و Mohammadzadeh و همکاران (۲۰۱۷) عامل کاربری اراضی را مهمترین تشخیص دادند. دلیل این یافته‌ها ناشی از لایه‌های ورودی متفاوت و استفاده از الگوریتم‌های متفاوت می‌باشد. زیرا طبق یافته‌های این پژوهش می‌توان با ورودی‌های متفاوت به نقشه‌های پهنه‌بندی دارای دقت پایین‌تر نیز دست یافت. اما به‌طور کلی متغیرهای توپوگرافی مهمترین عوامل تعیین‌کننده در رخداد گسل و شناسایی مناطق مستعد وقوع گسل بودند. در بررسی نرخ خطا در تعداد متفاوت درختان این الگوریتم با بررسی میزان تخمین نمونه‌های خارج از سبد (OOB) طبق شکل شماره ۱۱، میزان خطای الگوریتم، از حدود ۲۰۰ درخت به بعد دارای یک روال ثابت است. همچنین کمترین مقدار خطای تخمین نمونه‌های خارج از سبد که توسط این روش در مدل‌سازی استفاده نشده و در اعتبارسنجی مستقل به کار می‌روند، ۱۹/۶۹٪ بود که با تعداد ۵۰۰ درخت و ۴ متغیر در هر درخت به‌دست آمد.

باتوجه به نتایج ارزیابی، در روش ارزش اطلاعات حدود ۷۴ درصد گسل‌ها در کلاس خطر زیاد و خیلی زیاد واقع شدند اما در روش تراکم سطح حدود ۵ درصد در این طبقات قرار گرفتند. نتایج به‌دست آمده از روش ارزش اطلاعات بر مبنای استفاده از الگوریتم جنگل تصادفی، با نتایج Ayalew و همکاران (۲۰۰۵) از نظر دقت پهنه‌بندی خطر مطابقت دارد. همچنین نتایج بیانگر این است که روش ارزش اطلاعات نسبت به روش تراکم سطح برای تعیین پهنه‌های حساس به وقوع گسل دارای دقت بیشتری می‌باشد که همسو با نتایج Naderi و همکاران (۲۰۱۰) و Pourhashemi و همکاران (۲۰۱۴) است. همچنین برخلاف روش تراکم سطح، در روش ارزش اطلاعات پهنه با خطر خیلی‌زیاد مساحت بیشتری را به‌خود اختصاص داده است. پژوهش Shadfar و همکاران (۲۰۱۱) روش تراکم سطح را در بررسی خطر زمین‌لغزش دارای دقت مناسب‌تری معرفی نمود.

منابع مورد استفاده

- Momeni, A., 2015. 40 years of pedological studies in soil and water research institute. Agricultural Research, Education and Extension Organization, Tehran, 60p.
- Jedari-Eyvazi, J. and Mahmoudi, F., 2001. Dynamic Geomorphology. Payam-Noor University Publication, Tehran, 326p.
- Komac, M., 2006. A landslide susceptibility model using the analytical hierarchy process method and multivariate statistics in perialpine Slovenia. *Geomorphology*, 74: 17-28.
- Lee, S., 2004, Application of likelihood ratio and logistic regression models to landslide susceptibility mapping using GIS. *Journal of Environmental Management*, 34: 223-232.
- Lee, S. and Sambath, T., 2006. Landslide susceptibility mapping in the Damrei Romel area, Cambodia using frequency ratio and logistic regression models. *Journal of Environmental Geology*, 50(6): 847-855.
- Luca, F., Conforti, M. and Robustelli, G., 2011. Comparison of GIS-based gully susceptibility mapping using bivariate and multivariate statistics: Northern Calabria, South Italy. *Journal of Geomorphology*, 134(3): 297-308.
- Mahdavi, R., Alievazi, A., Gholami, H. and Kamali, A., 2017. Identifying the sediment source zones using maximum likelihood, minimum distance and parallelepiped algorithms (Case Study: South Roudbar, Kerman). *Iranian Journal of Range and Desert Research*, 24 (3): 610-622.
- Ministry of Energy., 2015. Iran Water Statistical Yearbook 2011-2012. Water and Waste Water Macro Planning Bureau, Tehran, 283p.
- Mohammadzadeh, K., Bahmani, S. and Fathi, M. H., 2017. Logistic regression assessment in the investigation of the landslide potential (case study: from Nasirabad to Sattar khan dam). *Hydrogeomorphology*, 3(11): 127-148.
- Naderi, F. and Karimi, H., 2011. Efficiency assessment of two Information Value and Gopta-Joshi methods in landslide hazard mapping in the Talkhab watershed of Ilam. *Journal of Watershed Management Research (Pajouhesh & Sazandegi)*, 92: 95-103.
- Naderi, F., Naseri, B., Karimi, H. and Habibi Bibalani, G. H., 2010. Efficiency evaluation of different landslide susceptibility mapping methods (Case study: Zangvan watershed, Ilam province). First international conference of soil and roots engineering relationship (LANDCON1005). Iran, Ardebil Province, 24-26 May: 1-8.
- Papadakis, M. and Karimalis, A., 2017. Producing a
- Anbalagan, R., 1992. Landslide hazard development and zonation mapping in mountainous Terrain. *Journal of Engineering Geology*, 32: 269-277.
- Andre, S. and Norman, K., 2010. Combining random forests and object-oriented analysis for landslide mapping from very high resolution imagery. *Procedia Environmental Sciences*, 3: 123-129.
- Ayalew, L., Ymagishi, H., Marui, H. and Kanno, T., 2005. GIS-based susceptibility mapping with comparisons of result from methods and verifications. *Journal of Engineering Geology*, 81: 432-445.
- Ebrahimi, P., Eslah, M. and Azarakhshi, M., 2017. Landslide hazard zonation using SMCE method and AHP technic (case study: Hafshejan watershed, Chaharmahal-O-Bakhtiari). *Journal of Range and Watershed Management*, 70(1): 1-17.
- Esfandiyari-Darabadi, F. and Beheshti Javid, E., 2016. Landslides susceptibility zoning using bayes' theorem-ANP hybrid model (case study: Heyran defile). *Journal of Hydrogeomorphology*, 2(8): 93-111.
- Fathizad, H., Safari, A., Bazgir, M. and Khosravi, G. H., 2017. Evaluation of SVM with Kernel method (linear, polynomial, and radial basis) and neural network for land use classification. *Iranian Journal of Range and Desert Research*, 23 (4): 729-743.
- Fawcett, T., 2006. An introduction to ROC analysis, *Pattern Recognition Letters*, 27: 861–874
- Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P., 1996. From data mining to knowledge discovery in databases. *AI magazine*, 17(3): 37-54.
- Ghazanfari, M., Alizadeh, S. and Teimourpour, B., 2008. Data Mining and knowledge discovery. University of Science & Technology Publication, Tehran, 403p.
- Gohardoust, A., Sadoddin, A., Ownegh, M. and Najafinejad, A., 2017. Identification of hazard areas by using land use planning (Case Study: Chehelchai Minodasht Watershed-Golestan Province). *Iranian Journal of Range and Desert Research*, 24 (3): 524-536.
- Hur, J. H., Ihm, S. Y. and Park, Y. H., 2017. A Variable Impacts Measurement in random forest for mobile cloud computing. *Wireless Communications and Mobile Computing*, 2017: 1-13.
- Ilanlou, M., Moghimi, E. and Servati, M. R., 2009. Mass movement hazard zonation with analyzing hierarchy process (AHP) method (case study: Karaj dam basin). *Journal of Physical Geography*, 2(5): 85-95.
- Jamshidi, M., Eftekhari, K., Navidi, M. N. and

- Riley, S. J., De-Gloria, S. D. and Elliot, R., 1999. A terrain ruggedness that quantifies topographic heterogeneity. *Intermountain Journal of Science*, 5(1-4): 23-27.
- Sarp, G., 2014. Evolution of neotectonic activity of East Anatolian Fault System (EAFS) in Bingöl Pull-aport basin based on fractal dimension and morphometric. *Journal of Asian Earth Sciences*, 88: 168-177.
- Shadfar, S., Yamani, M. and Namaki, M., 2011. Landslide hazard zonation using information value, area density and LNRF models in Chalkrood catchment. *Journal of Watershed Engineering and Management*, 3(1): 40-47.
- Wilson, J. P. and Gallant, J. C., 2000. Digital terrain analysis. *Terrain analysis: Principles and applications*, 6(12): 1-27.
- landslide susceptibility map through the use of analytic hierarchical process in Finikas watershed, North Peloponnese, Greece. *American Journal of Geographic Information System*, 6(1): 14-22.
- Parmar, M. K., Malik, A. and Godiyal. M., 2012. Landslide hazard zonation using remote sensing and GIS: a case study of Giri valley. District Sirmaur Himachal Pradesh. *International Journal of Environmental Sciences*, 1: 26-39.
- Pourghasemi, H. R., Moradi, H. R., Fatemi Aghda, S. M., Gokceoglu, C. and Pradhan, B., 2014. GIS-based landslide susceptibility mapping with probabilistic likelihood ratio and spatial multi-criteria evaluation models (North of Tehran, Iran). *Arabian Journal of Geosciences*, 7(5): 1857-1878.
- Pourhashemi, S., Amirahmadi, A. and Akbari, E., 2014. Among bivariate method selection for landslide hazard zonation in GIS (case study: Baghi basin). *Arid Regions Geographic Studies*, 4(15): 71-89.

Evaluation of data mining and bivariate statistical methods in risk zoning of fault occurrence (Case study: Qara-Qum watershed)

M. Bashiri^{1*}, S.M. Kavousi-Davoudi² and A. Afzali³

1*-Corresponding author, Assistant Professor, Department of Nature Engineering and Medicinal Plants, Faculty of Agriculture and Natural Resources, University of Torbat Heydarieh, Razavi-Khorasan, Iran, Email: m.bashiri@torbath.ac.ir

2- M.Sc. in Watershed Management, Samane-Ab-Koomesh Company, Razavi-Khorasan, Iran

3- M.Sc. in Desert Area Management Department, International Desert Research Center (IDRC), University of Tehran, Iran

Received:12/29/2018

Accepted:07/01/2019

Abstract

Fault is one of the main processes in tectonics which has relationship with phenomena such as earthquake. Therefore, awareness of the effect of environmental factors on fault occurrence and recognition of high risk areas is very important, that these goals in the present study have been investigated. The research area is Qara-Qum where after determining its faults, altitude, slope, direction, climate, land use, geomorphology, geology, erosion, precipitation, flood zones, pedology, stream power, topographic ruggedness index, distance from road, waterway, anticline and syncline were extracted. Then, data mining algorithms including of decision tree, random forest, cumulative, backing machine, logistic regression and neural network in R software are used to identify the value of variables and bivariate statistical methods including of information value and area density for identification of the values for each variable class fitted in fault occurrence. The accuracy of classification algorithms with ROC curve showed that based on input variables, random forest and support vector machine algorithms with 88% and 86% area under a curve had the best performance in classifying fault occurrence, respectively. Finally, according to the Gini coefficients in random forest algorithm, the zoning maps obtained by combining this algorithm were prepared and validated by bivariate statistical methods. According to this algorithm, the height, pedology and topographic ruggedness index variables, identified as the most important parameters in the fault occurrence respectively. Based on the zoning maps evaluation, information value and area density methods, around 52 and 35 percent of faults placed in very high risk class respectively. Therefore, the information value method was more accurate in identifying fault-sensitive zones. According to the results, data mining methods were introduced as a useful tool in fault risk management. It is also necessary to pay attention to environmental variables, especially topography, during the basin management and land use change stages.

Keywords: Information value, classification Algorithm, area density, ROC curve.