

پیش‌بینی مکانی حساسیت زمین لغزش با استفاده از الگوریتم‌های پیشرفته یادگیری ماشین (مطالعه موردی: شهرستان سروآباد، استان کردستان)

هیمن راست‌خدیو^۱، بهارک معتمدوزیری^{۲*}، سیداکبر جوادی^۲ و حسن احمدی^۳

(۱) دانشجوی دکتری، گروه مهندسی طبیعت، دانشکده منابع طبیعی و محیط زیست، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران.

(۲) دانشیار گروه مهندسی طبیعت، دانشکده منابع طبیعی و محیط زیست، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران.

*رایانامه نویسنده مسئول مکاتبات: bm vaziri@gmail.com

(۳) استاد گروه احیا مناطق خشک و کوهستانی، دانشکده منابع طبیعی، دانشگاه تهران، کرج، ایران.

<https://doi.org/10.71916/jrn.2024.10317>

تاریخ پذیرش: ۱۴۰۲/۱۲/۲۳

تاریخ دریافت: ۱۴۰۲/۱۰/۰۳

چکیده

وقوع رخداد زمین لغزش در مناطق کوهستانی ممکن است به زیرساخت‌ها از جمله جاده‌ها آسیب جدی وارد کند، همچنین ممکن است به مرگ‌ومیر انسان‌ها منجر شود. هدف از انجام این پژوهش، پیش‌بینی مکانی خطر زمین لغزش با استفاده از الگوریتم‌های پیشرفته داده‌کاوی در شهرستان سروآباد (استان کردستان) بود. در این مطالعه، پتانسیل‌یابی خطر زمین لغزش با استفاده از دو الگوریتم پیشرفته داده‌کاوی شامل جنگل تصادفی (RF) و درخت تصمیم (DT) انجام شد. ابتدا فایل نقطه‌ای ۱۶۶ زمین لغزش رخ داده در شهرستان سروآباد به‌عنوان نقشه موجودی زمین لغزش در نظر گرفته شد. به‌منظور تهیه مدل و اعتبارسنجی آن، نقاط زمین لغزش به دو بخش داده‌های آموزشی (۷۰ درصد) و اعتبارسنجی (۳۰ درصد) تقسیم شدند. در مجموع ۱۶ پارامتر شامل شیب، جهت جغرافیایی، ارتفاع از سطح دریا، فاصله از آبراهه، فاصله از جاده، تراکم رودخانه، فاصله از گسل، تراکم گسل، تراکم جاده، بارندگی، کاربری و پوشش اراضی، شاخص NDVI، لیتولوژی، زمین لرزه، شاخص توان آبراهه (SPI) و شاخص رطوبت توپوگرافی (TWI) به‌منظور پهنه‌بندی خطر زمین لغزش استفاده شدند. در نهایت، عملکرد مدل‌ها با استفاده از منحنی مشخصه عملکرد سیستم (ROC) مورد بررسی قرار گرفت. نتایج تحلیل منحنی ویژگی عملکرد نسبی نشان داد مدل‌های درخت تصمیم و جنگل تصادفی به‌ترتیب دارای مقدار AUC برابر ۰/۹۴۲ و ۰/۹۵۱ می‌باشند. بنابراین مدل جنگل تصادفی نسبت به درخت تصمیم دارای بالاترین مقدار AUC بوده و بهترین مدل برای پیش‌بینی خطر زمین لغزش در آینده در منطقه مورد مطالعه می‌باشد. نقشه‌های پتانسیل وقوع زمین لغزش، ابزارهای کارآمدی است، به‌طوری‌که می‌توان آنها را جهت مدیریت زیست‌محیطی، برنامه‌ریزی کاربری زمین و توسعه زیرساخت‌ها مورد استفاده قرار داد.

واژه‌های کلیدی: الگوریتم جنگل تصادفی، الگوریتم درخت تصمیم، استان کردستان، زمین لغزش، مخاطرات طبیعی.

مقدمه

ایجاد این رویداد خطرناک نقش دارند (Basu & Pal, 2019). Froude و Petley (۲۰۱۸) با جمع‌آوری زمین لغزش از سال ۲۰۰۴ تا ۲۰۱۶ یک پایگاه داده جهانی زمین لغزش ایجاد کردند. در طول این ۷ سال در مجموع ۴۸۶۲ زمین لغزش مرگبار در سطح جهان ثبت شد که منجر به مرگ ۵۵،۹۹۷ نفر شد. تهیه نقشه حساسیت زمین لغزش به‌عنوان یکی از اقدامات رایج جهت پیشگیری و کاهش اثرات زمین لغزش در نظر گرفته

زمین لغزش به‌عنوان یکی از بلایای طبیعی که در مناطق کوهستانی سراسر جهان گسترده است، تهدید بزرگی برای زندگی، اموال و محیط طبیعی انسان‌ها به‌شمار می‌رود (Sun et al., 2020). بارندگی، زمین لرزه و سایر رویدادهای طبیعی و همچنین جنگل‌زدایی، محیط‌های پیچیده زمین‌شناسی و زمین‌ساختی، کشاورزی و سایر فعالیت‌های انسانی، همگی در

می‌شود (Xiao et al., 2020). از این رو در دهه گذشته، تحقیقات و تحلیل‌های فراوانی در مورد روش‌های تهیه نقشه حساسیت زمین‌لغزش در بسیاری از مناطق در سراسر جهان انجام شد (Wang et al., 2020; Shirzadi et al., 2018). علاوه بر استفاده از داده‌های آرشو شده، تکنیک‌های متنوعی جهت پهنه‌بندی و پتانسیل‌یابی خطر زمین‌لغزش استفاده شد که می‌توان آنها را در چهار دسته طبقه‌بندی کرد (Reichenbach et al., 2018): (۱) بررسی چشمی و تفسیر عکس‌های هوایی و ماهواره‌ای؛ (۲) مشاهدات میدانی و نقشه‌برداری؛ (۳) تفسیر اتوماتیک و نیمه‌اتوماتیک مدل‌های رقومی ارتفاع؛ و (۴) پهنه‌بندی اتوماتیک و نیمه‌اتوماتیک با استفاده از سنجش‌ازدور. در تکنیک‌های پهنه‌بندی خودکار و نیمه‌خودکار از مدل‌های آماری جهت بررسی و تعریف همبستگی آماری بین مجموعه‌ای از متغیرهای وابسته و مستقل استفاده می‌شود. سپس با توجه به همبستگی آماری بین این متغیرها، نقشه پتانسیل وقوع زمین‌لغزش تهیه می‌شود (Sassa et al., 2022). Reichenbach و همکاران (۲۰۱۸) این رویکردها را به (الف) آمار کلاسیک (به‌عنوان مثال، رگرسیون خطی و لجستیک)؛ (ب) آمار مبتنی بر شاخص (به‌عنوان مثال، شواهد وزنی و تجزیه و تحلیل اکتشافی)؛ (ج) یادگیری ماشین (به‌عنوان مثال، منطق فازی، ماشین‌های بردار پشتیبان و جنگل تصادفی)؛ (د) شبکه‌های عصبی مصنوعی؛ و (ه) تجزیه و تحلیل تصمیم‌گیری چندمعیاره طبقه‌بندی کردند. هر یک از این روش‌ها، مزایا و محدودیت‌های خود را داشته و محققان سطوح مختلفی از صحت و دقت را برای این رویکردها گزارش کرده‌اند (Dou et al., 2021; Reichenbach et al., 2018). علاوه بر کارایی روش پهنه‌بندی، انتخاب صحیح عوامل تاثیرگذار بر وقوع زمین‌لغزش نیز بر عملکرد نقشه پتانسیل زمین‌لغزش حاصل، تاثیر به‌سزایی دارد (Jebur et al., 2014). Reichenbach و همکاران (۲۰۱۸) ۵۶۵ مقاله تحقیقاتی در زمینه پتانسیل‌یابی خطر زمین‌لغزش را از سال ۱۹۸۳ تا ۲۰۱۶ مورد بررسی قرار دادند و ۵۹۶ عامل مورد استفاده برای ارزیابی حساسیت زمین‌لغزش را برشمردند. آنها همچنین دریافتند تعداد فاکتورهای مورد استفاده در هر مدل از ۲ تا ۲۲ با میانگین ۹ فاکتور متغیر است. عوامل تاثیرگذار بر وقوع زمین‌لغزش شامل عوامل شرطی بوده که شرایط حرکت

(مانند شیب) را برای زمین‌لغزش‌ها فراهم می‌کنند. شرایط توپوگرافی و ژئومورفیک (ارتفاع، شیب، جهت جغرافیایی، موقعیت شیب، انحناء شیب، میکروتوپوگرافی، شاخص رطوبت توپوگرافیک، شاخص ناهمواری‌های زمین و شاخص قدرت جریان)، وضعیت زمین‌شناسی (سنگ‌شناسی و فاصله از گسل‌ها) و عوامل مرتبط با محیط (پوشش زمین، میانگین بارندگی سالانه و فاصله از رودخانه‌ها) به‌طور گسترده در پتانسیل‌یابی خطر زمین‌لغزش مورد مطالعه قرار گرفته‌اند (Zhu et al., 2018; Othman et al., 2018; Yu & Gao, 2020). Merghadi و همکاران (۲۰۲۰) اظهار داشتند متداول‌ترین الگوریتم‌های مورد استفاده در زمینه پتانسیل‌یابی خطر زمین-لغزش شامل رگرسیون لجستیک، شبکه عصبی مصنوعی، ماشین‌بردار پشتیبانی، درخت تصمیم و جنگل تصادفی است. در مطالعات انجام شده کارایی مدل‌های یادگیری ماشین جنگل تصادفی و درخت تصمیم در زمینه پیش‌بینی مکانی زمین‌لغزش مورد تایید قرار گرفته است (Ado et al., 2022; Prasad et al., 2021).

کشور ایران با توجه به وضعیت ناهمواری‌ها و توپوگرافی، فعالیت‌های زمین‌ساختی و لرزه‌خیزی، شرایط متنوع زمین-شناسی و اقلیمی، با خطر زمین‌لغزش‌های متعدد در مناطق مختلف روبه‌رو است. از طرف دیگر، شهرستان سروآباد واقع در استان کردستان یکی از مناطقی است که در آن حرکات توده‌های بسیار زیادی نسبت به سایر مناطق کشور اتفاق می‌افتد. بر اساس داده‌های ثبت شده در بانک اطلاعاتی زمین‌لغزش‌های کشور، استان کردستان سومین استان از نظر فراوانی وقوع زمین‌لغزش پس از مازندران و گلستان محسوب می‌شود (زمانی و ریاحی، ۱۳۹۳). از طرف دیگر شهرستان سروآباد نیز دارای توپوگرافی خشن و کوهستانی با اقلیم متنوع است که مستعد رخداد زمین‌لغزش می‌باشد.

هیچ قانون خاصی جهت تعیین اینکه چه فاکتورهایی در پتانسیل‌یابی خطر زمین‌لغزش استفاده شود و چه رویکردهایی برای پتانسیل‌یابی وقوع زمین‌لغزش استفاده شود، وجود ندارد. پتانسیل‌یابی و پهنه‌بندی حساسیت زمین‌لغزش به‌عنوان روشی موثر برای پیش‌بینی زمین‌لغزش مورد توجه قرار گرفته است. عملکرد مدل‌های پتانسیل‌یابی زمین‌لغزش توسط بسیاری از عوامل محدود شده، بنابراین بهبود دقت نقشه‌های پتانسیل

پیش‌بینی مکانی حساسیت زمین‌لغزش با استفاده از الگوریتم‌های پیشرفته یادگیری ماشین/۸۹

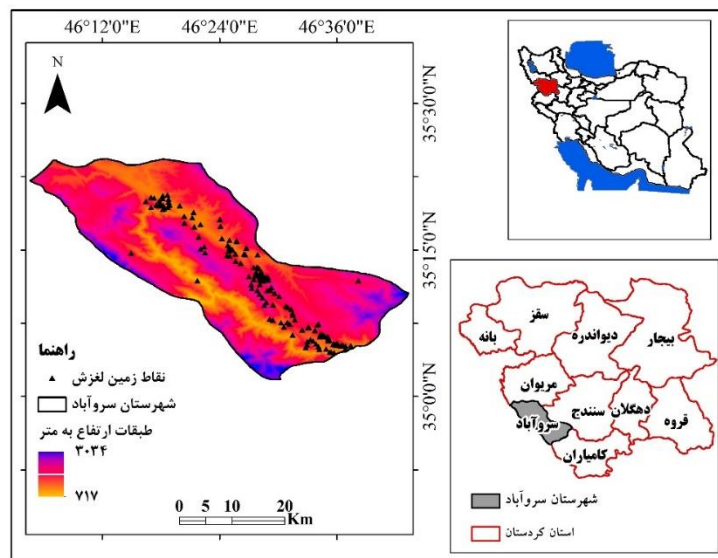
جغرافیایی واقع شده است. شهرستان سروآباد از شمال به شهرستان مریوان، از شرق به شهرستان سنندج و از جنوب به شهرستان کامیاران محدود می‌شود. مرکز این شهرستان، شهر سروآباد است که تقریباً در ۸۰ کیلومتری غرب سنندج قرار دارد. شهرستان سروآباد متشکل از دو شهر سروآباد و اورامان می‌باشد. شهرستان سروآباد با میکروکلیم‌های متفاوت شامل مناطق کوهستانی و سردسیر تا مناطقی با آب و هوایی تقریباً معتدل است و حتی در بعضی از نقاط این شهرستان، درختان مختص مناطق نیمه‌گرمسیری مشاهده می‌شود (سایت رسمی فرمانداری شهرستان سروآباد، ۱۳۹۰). موقعیت منطقه مورد مطالعه در ایران و استان کردستان در شکل (۱) ارایه شده است.

زمین‌لغزش اهمیت زیادی برای پیشگیری از زمین‌لغزش دارد. بنابراین روش‌ها و مدل‌های فراوانی جهت پتانسیل‌یابی مخاطرات وجود داشته و با توجه به مزایای دو مدل جنگل تصادفی و درخت تصمیم، این دو مدل برای پتانسیل‌یابی خطر زمین‌لغزش در شهرستان سروآباد استفاده شده است.

مواد و روش‌ها

منطقه مورد مطالعه

محدوده این مطالعه شهرستان سروآباد بود که در استان کردستان و در غرب ایران با وسعت $104,847/8$ هکتار و در موقعیت جغرافیایی $38^{\circ} 43' 46''$ تا $38^{\circ} 46' 46''$ طول جغرافیایی شرقی و $35^{\circ} 15' 15''$ و $35^{\circ} 24' 42''$ عرض



شکل ۱. موقعیت شهرستان سروآباد در استان کردستان و ایران

(۷۰ درصد) شامل ۱۱۶ نقطه و داده‌های اعتبارسنجی (۳۰

درصد) شامل ۵۰ نقطه تقسیم می‌شوند.

عوامل موثر در وقوع زمین‌لغزش

در این پژوهش با بررسی مطالعات قبلی انجام شده (Wang *et al.*, 2020; Zhou *et al.*, 2021; Dou *et al.*, 2021) و شرایط منطقه، فاکتورهای شیب، جهت جغرافیایی، ارتفاع، فاصله از آبراه، فاصله از جاده، تراکم رودخانه، فاصله از گسل، تراکم گسل، تراکم جاده، کاربری اراضی، شاخص پوشش گیاهی نرمال شده (NDVI)، لیتولوژی، بارندگی، بزرگی زمین لرزه، شاخص توان آبراه^۱ (SPI) و شاخص رطوبت

تهیه نقشه پراکنش زمین‌لغزش‌ها

با تهیه نقشه پراکنش وقوع زمین‌لغزش‌ها می‌توان رابطه کلی بین عوامل ایجادکننده و زمین‌لغزش‌های گذشته را به صورت آماری بررسی کرد که این امر می‌تواند مبنایی برای ارزیابی پتانسیل‌یابی خطر زمین‌لغزش در منطقه باشد (Hamza & Raghuvanshi, 2017). فایل نقطه‌ای ۱۶۶ زمین‌لغزش رخ داده در شهرستان سروآباد که توسط اداره کل منابع طبیعی استان کردستان و بر اساس تفسیر تصاویر ماهواره‌ای و بازدیدهای میدانی تهیه شده بود، به‌عنوان نقشه زمین‌لغزش‌های واقعی در منطقه در نظر گرفته شد. به‌منظور آموزش مدل و اعتبارسنجی مدل تهیه شده، نقاط زمین‌لغزش به دو بخش داده‌های آموزشی

¹ Stream Power Index

توپوگرافی^۱ (TWI) برای پهنه‌بندی پتانسیل خطر زمین‌لغزش مدنظر قرار گرفتند.

ابتدا مدل رقومی ارتفاعی (DEM) با قدرت تفکیک ۱۲/۵×۱۲/۵ متر از ماهواره ALOS PALSAR و از سایت Alaska Satellite Facility تهیه شد. در ادامه نقشه طبقات ارتفاع با استفاده از مدل رقومی ارتفاع با پیکسل سایز ۱۲/۵×۱۲/۵ متر در محیط نرم‌افزارهای Arc GIS 10.3 با شش طبقه تهیه شد. نقشه شیب با استفاده از مدل رقومی (با قدرت تفکیک ۱۲/۵ متر) و نرم‌افزار ArcGIS تهیه و به شش کلاس طبقه‌بندی شدند. همچنین جهت جغرافیایی نیز با استفاده از مدل رقومی (با قدرت تفکیک ۱۲/۵ متر) و نرم‌افزار ArcGIS تهیه و به ۹ کلاس طبقه‌بندی شدند. شاخص تفاوت نرمال شده گیاهی، با وضوح ۱۰ متر (معادله ۱) با استفاده از تصاویر سنتینل ۲ (تاریخ ۱۴۰۱/۲/۲۲) که از سایت scihub.copernicus دانلود شده تهیه گشت. در رابطه (۱) R، مادون قرمز و NIR، مادون قرمز نزدیک است.

$$NDVI = \frac{[NIR-R]}{[NIR+R]} \quad (۱) \text{ رابطه}$$

نقشه شبکه جاده با استفاده از نقشه توپوگرافی رقومی ۱:۵۰۰۰۰ سازمان نقشه‌برداری کشور در محیط ArcGIS تهیه گردید. نقشه فاصله از جاده با اعمال تابع Euclidean distance در محیط ArcGIS تهیه و این فاکتور به ۵ طبقه تقسیم‌بندی شد. نقشه تراکم جاده با اعمال تابع Line density در محیط ArcGIS تهیه و این فاکتور به ۵ طبقه تقسیم‌بندی شد.

نقشه شبکه زهکشی با استفاده از نقشه توپوگرافی رقومی ۱:۵۰۰۰۰ سازمان نقشه‌برداری کشور در محیط ArcGIS تهیه گردید. نقشه فاصله از شبکه زهکشی با اعمال تابع Euclidean distance در محیط ArcGIS تهیه و این فاکتور به ۵ طبقه تقسیم‌بندی شد. نقشه تراکم آبراهه‌ها با اعمال تابع Line density در محیط ArcGIS تهیه و این فاکتور به ۵ طبقه تقسیم‌بندی شد.

نقشه فاصله از گسل با استفاده از نقشه توپوگرافی رقومی ۱:۵۰۰۰۰ و اعمال توابع Distance در محیط ArcGIS تهیه گردید و این فاکتور به ۵ طبقه تقسیم‌بندی شد. نقشه تراکم گسل‌ها با استفاده از نقشه توپوگرافی رقومی ۱:۵۰۰۰۰ سازمان

نقشه‌برداری کشور در محیط ArcGIS تهیه و در نهایت این فاکتور به ۵ طبقه تقسیم‌بندی شد.

در این پژوهش، داده‌های سنجنده OLI ماهواره لندست ۸ مربوط به تاریخ ۲۵ ژوئن سال ۲۰۲۱ برای تهیه نقشه کاربری اراضی شهرستان سروآباد استفاده شد. همچنین از نقشه توپوگرافی ۱:۲۵۰۰۰ و تصاویر گوگل ارث منطقه جهت تصحیح هندسی و انتخاب نقاط تعلیمی بهره گرفته شد. کلیه مراحل پردازش در نرم‌افزارهای ENVI 5.3 و ArcGIS انجام گرفت. صحت طبقه‌بندی لایه کاربری تهیه شده با شاخص کاپا بررسی شد که ۰/۸۶ به دست آمد. لایه کاربری اراضی نیز بر اساس کاربری‌های مختلف به کلاس‌های مختلف تقسیم‌بندی شد. به منظور استخراج لایه لیتولوژی از نقشه زمین‌شناسی ۱:۱۰۰۰۰۰ استفاده گردید.

شاخص رطوبت توپوگرافی (TWI)، شاخص ترکیب پستی و بلندی بوده که نسبت بین شیب‌ها را در حوضه به نمایش می‌گذارد. شاخص خیسی، شاخصی از توزیع مکانی رطوبت خاک در طول چشم‌انداز زمین است که به وسیله معادله (۳) قابل محاسبه می‌باشد (Nefeslioglu et al., 2010). در رابطه (۲) As، مساحت بالادست شیب که به یک نقطه زهکش می‌شود (m^2) و β زاویه شیب در آن نقطه (رادیان) است.

$$TWI = \ln\left(\frac{As}{tg\beta}\right) \quad (۲) \text{ رابطه}$$

شاخص توان آبراهه (SPI)، برای کمی کردن اثر توپوگرافی بر روی فرآیندهای هیدرولوژیکی مورد استفاده قرار می‌گیرد (Crosby, 2006). مقدار این پارامتر از رابطه (۳) محاسبه می‌شود (Devkota et al., 2013). در این رابطه As مساحت بالادست شیب است که به یک نقطه زهکش می‌شود (m^2) و β زاویه شیب در آن نقطه (رادیان) است.

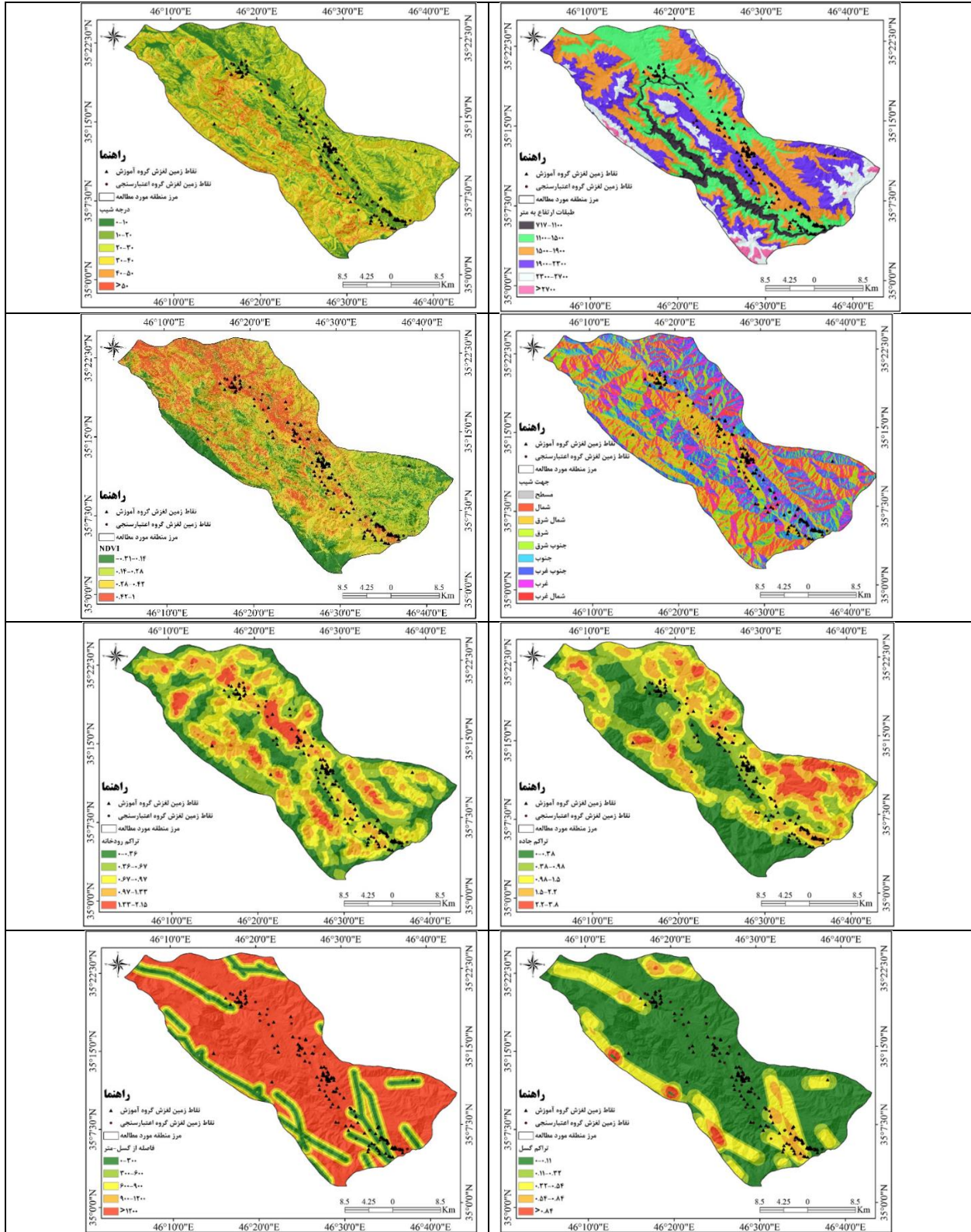
$$SPI = As \times tg\beta \quad (۳) \text{ رابطه}$$

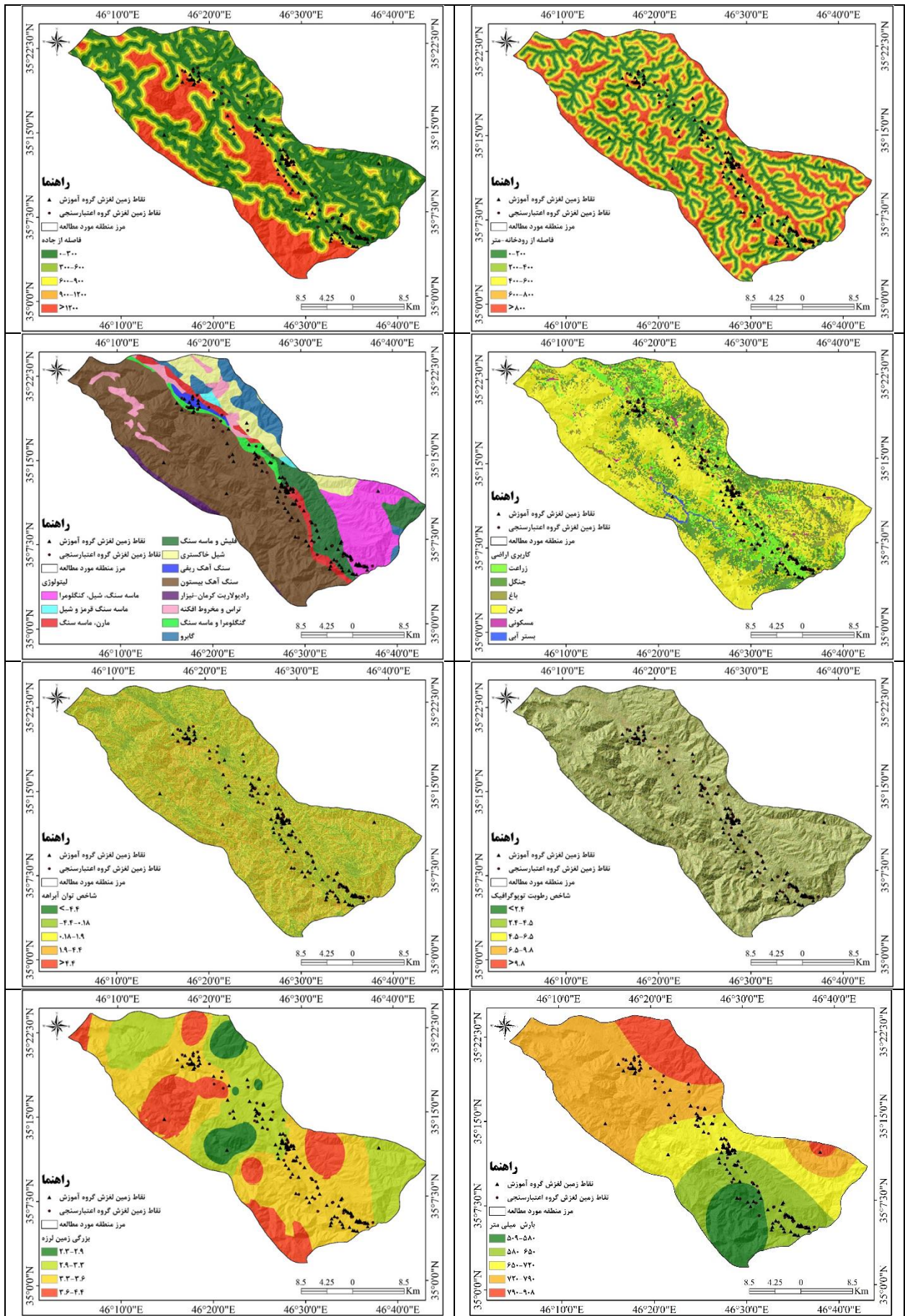
برای تهیه لایه بارندگی از داده‌های ۲۰ ساله (۱۳۸۰-۱۴۰۰) ۵ ایستگاه باران‌سنجی داخل و خارج محدوده مطالعاتی استفاده شد و به صورت سالانه میانگین گرفته شد. سپس نقشه بارندگی با استفاده از روش درون‌یابی کریجینگ تهیه شد. با توجه به پارامترهای ریشه متوسط مربع خطاها (۶۸/۷۴)، میانگین مطلق خطا (۵۳/۳۴) و میانگین خطای انحراف (۰/۰۳-)، مدل دایره‌ای کریجینگ معمولی بهترین مدل برای تهیه نقشه توزیع مکانی بارندگی بود. در نهایت نقشه بارندگی به ۵ طبقه تقسیم‌بندی

پیش‌بینی مکانی حساسیت زمین‌لغزش با استفاده از الگوریتم‌های پیشرفته یادگیری ماشین ۹۱/

استفاده از روش درون‌یابی در Arc GIS نقشه پهنه‌بندی زمین‌لرزه تهیه و در نهایت نقشه بزرگی زمین‌لرزه به چهار طبقه تقسیم شد.

شد. این داده‌ها از اداره کل هواشناسی تهیه شد. تعداد زمین‌لرزه‌های ثبت شده در منطقه مورد مطالعه ۱۹ واقعه دارای بزرگی $2/3-4/4$ ریشتر می‌باشد. داده‌های زمین‌لرزه از وب سایت پژوهشگاه بین‌المللی زلزله‌شناسی تهیه گشت. سپس با





شکل ۲. نقشه فاکتورهای موثر در وقوع زمین لغزش

روش جنگل تصادفی

الگوریتم جنگل تصادفی برای اولین بار توسط Breiman (۱۹۹۹) مطرح و توسعه یافت. این الگوریتم یکی از کارآمدترین تکنیک‌هایی است که برای ارزیابی مسایل مرتبط با متغیرهای هدف یا طبقه‌بندی الگوها مورد استفاده قرار می‌گیرد. جنگل تصادفی، فضای ورودی را به مجموعه‌ای از مناطق جداگانه تقسیم کرده و به هر بخش از جواب، ارزشی را اختصاص می‌دهد (Trigila et al., 2015). در حالت ساده، این پاسخ می‌تواند بر اساس متوسط ارزش رگرسیون هدف مرتبط با الگوها در هر منطقه تعیین شود. هنگام استفاده از جنگل تصادفی برای ساختن یک درخت، طبقات مجزای الگوی موجود با هدف جایگزینی هر الگوی منتخب، دوباره انتخاب می‌شود (Stumpf & Kerle, 2011). اندازه این نمونه‌ها با تعداد کل مدل‌های موجود برابر خواهد بود. به‌منظور تعیین اولویت هر یک از پارامترهای موثر از دو فاکتور میانگین کاهش دقت و میانگین کاهش دقت در مقایسه با شاخص اهمیت جینی در تعیین اهمیت عوامل موثر، پایدارتر می‌باشد (Nicodemus, 2011). در نهایت، با وارد کردن داده‌های مربوط به عوامل موثر و نقشه پراکنش زمین‌لغزش‌ها (با فرمت CSV) به نرم‌افزار وکا (Weka)، مدل‌سازی انجام شد (Cheng et al., 2021; Chen et al., 2017) و نقش عوامل موثر در وقوع زمین‌لغزش تعیین شد. در نهایت وزن‌های به‌دست آمده برای عوامل موثر در نرم‌افزار Weka به محیط ArcGIS منتقل شد و نقشه نهایی زمین‌لغزش با ۵ کلاس تهیه گردید. تعداد درختان برابر با ۱۰۰ و متغیر تقسیم تصادفی برابر با ۱ منجر به بالاترین دقت با حداقل زمان برای دستیابی به نتایج شد.

الگوریتم درخت تصمیم

درخت تصمیم یک روش یادگیری نظارت شده ناپارامتری بوده که برای طبقه‌بندی و پیش‌بینی استفاده می‌شود (Nefeslioglu et al., 2010). به‌طور کلی درخت تصمیم به دو دسته تقسیم می‌شود: درخت‌های طبقه‌بندی و درخت‌های رگرسیون. درخت‌های طبقه‌بندی برای پیش‌بینی متغیرهای گسسته استفاده می‌شوند، در حالی‌که درخت‌های رگرسیون برای متغیرهای پیوسته استفاده می‌شوند. ایده اصلی در اینجا، ایجاد مدلی است که با یادگیری چندین قانون تصمیم‌گیری

استنباط شده از کل داده، ارزش یک عامل وابسته را پیش‌بینی کند. این مجموعه قوانین درخت تصمیم نامیده می‌شوند. برای این منظور، کل جمعیت یا نمونه بر اساس مهم‌ترین تقسیم‌کننده‌ها در متغیرهای ورودی، به دو یا چند مجموعه همگن تقسیم می‌شوند. مزیت اصلی درخت تصادفی این است که به تبدیل متغیرها نیاز نبوده زیرا ساختار مدل با یا بدون تبدیل یکسان باقی خواهد ماند، بنابراین در زمان صرفه‌جویی می‌شود. علاوه بر این، توانایی مدل‌سازی روابط پیچیده میان متغیرها را داشته و ساخت و تفسیر آن آسان می‌باشد (Tien Bui et al., 2012). در الگوریتم درخت تصمیم به‌منظور مدل‌سازی از نرم‌افزار Weka استفاده شد. لازم به ذکر است که کلیه پارامترها با استفاده از تابع Resample و در محیط ArcGIS هم مقیاس (۱۲/۵*۱۲/۵ متر) شدند.

ارزیابی عملکرد و کارایی الگوریتم‌ها

ارزیابی عملکرد الگوریتم‌های پیش‌بینی مکانی زمین‌لغزش توسط معیارهای درصد مساحت زیر منحنی ROC^۱، دقت^۲، صحت^۳ و شاخص کاپا^۴ که در محیط نرم‌افزار وکا محاسبه شده بود، بررسی شد. بر این اساس، در مدلی که مقادیر معیارهای اشاره شده به سمت یک تمایل داشته باشند، مدل مناسب‌تر شناخته می‌شود (Shirzadi et al., 2018). همچنین با استفاده از داده‌های اعتبارسنجی و در محیط نرم‌افزار SPSS نسخه ۲۴، منحنی مشخصه عملکرد (ROC) محاسبه شد. درصد مساحت زیر این منحنی (AUC) یک شاخص کمی برای تعیین عملکرد کلی مدل‌ها شناخته شد. هرچه مقدار این شاخص بزرگ‌تر باشد، عملکرد مدل بهتر می‌باشد. محدوده این شاخص از ۰/۵ (مدل با عملکرد ضعیف) تا ۱ (عملکرد دقیق مدل) متغیر می‌باشد (Shirzadi et al., 2018).

نتایج

در این پژوهش، از تکنیک IGR^۵ و شاخص میانگین شایستگی^۶ جهت ارزیابی قدرت پیش‌بینی عوامل موثر بر وقوع زمین‌لغزش استفاده شد. نتایج این روش در جدول (۱) آرایه شده است. نتایج نشان داد از میان ۱۶ عامل موثر در نظر گرفته شده در ابتدا، تنها ۱۳ عامل شامل شیب، ارتفاع، فاصله از آبراهه، فاصله از جاده، تراکم رودخانه، فاصله از گسل، تراکم جاده، کاربری اراضی، NDVI، لیتولوژی، بارندگی، بزرگی زمین‌لرزه و شاخص رطوبت توپوگرافی در وقوع زمین‌لغزش نقش

1 Area Under the ROC Curve

2 Precision

3 Recall or Accuracy

4 Kappa

5 Information Gain Ratio

6 Average Merit

داشتند. همچنین، نتایج نشان داد سه عامل جهت جغرافیایی، تراکم گسل و SPI به دلیل مقدار میانگین شایستگی برابر صفر از مدل سازی نهایی حذف شدند. همچنین متغیرهای تراکم جاده، ارتفاع و تراکم آبراهه نسب به سایر متغیرها بیشترین تاثیر را بر وقوع زمین لغزش داشتند.

جدول ۱. انتخاب مهم ترین عوامل موثر بر وقوع زمین لغزش با استفاده از شاخص IGR

AM± St.D	عوامل موثر	AM± St.D	عوامل موثر
۰/۱۳۸±۰/۰۱۱	فاصله از آبراهه	۰/۳۳۵±۰/۰۱۵	تراکم جاده
۰/۰۹۸±۰/۰۰۸	NDVI	۰/۲۷۸±۰/۰۱۸	ارتفاع
۰/۰۹۶±۰/۰۰۲	لیتولوژی	۰/۲۱۹±۰/۰۱۳	تراکم آبراهه
۰/۰۵۸±۰/۰۰۸	کاربری اراضی	۰/۲۱۵±۰/۰۴۲	فاصله از جاده
۰/۰۰۶±۰/۰۱۹	فاصله از گسل	۰/۱۹±۰/۰۱۱	زمین لرزه
۰/۰۰۷±۰/۰۲۱	TWI	۰/۱۶۶±۰/۰۳۲	بارندگی
		۰/۱۴۱±۰/۰۶۳	شیب

پیکسل های زمین لغزش را به درستی به عنوان مناطق تحت سیطره زمین لغزش طبقه بندی کند و نسبت به مدل درخت تصمیم قدرت پیش بینی بالاتری داشته است. شاخص شفافیت مدل های جنگل تصادفی و درخت تصمیم به ترتیب ۰/۹۵۷ و ۰/۹۸۱ نشانگر آن است که ۹۸/۱ درصد از پیکسل های بدون رخداد زمین لغزش به درستی به عنوان پیکسل های بدون رخداد زمین لغزش طبقه بندی شده است.

پس از آموزش مدل های جنگل تصادفی و درخت تصمیم، عملکرد آنها در زمینه پتانسیل یابی وقوع زمین لغزش با استفاده از معیارهای آماری (جدول ۲) سنجیده شد. بنابراین از نظر نمونه های آموزشی، مدل جنگل تصادفی (۰/۹۵۶) نسبت به مدل درخت تصمیم (۰/۹۰۷) دارای صحت بالاتری بود. مقدار شاخص حساسیت در مدل های جنگل تصادفی و درخت تصمیم به ترتیب ۰/۹۵۷ و ۰/۹۴۳ بود که این امر بدان معنی است که مدل جنگل تصادفی قادر است ۹۵/۷ درصد از

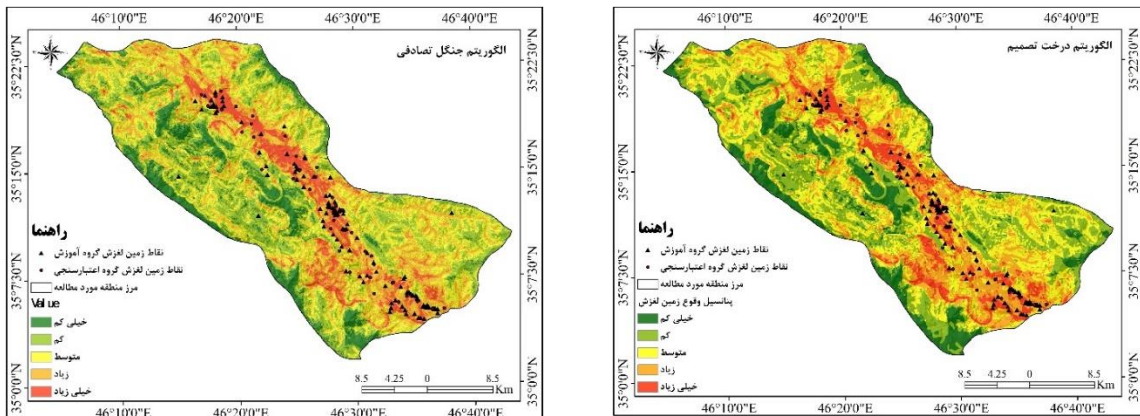
جدول ۲. عملکرد الگوریتم های پیش بینی مکانی زمین لغزش با استفاده از داده های آموزشی

جنگل تصادفی	درخت تصمیم	TRAIN
۲۳	۲۹	True positive
۲۱	۳۰	True negative
۱	۱	False positive
۱	۵	False negative
۰/۹۵۷	۰/۹۴۳	Sensitivity (%)
۰/۹۵۷	۰/۹۸۱	Specificity (%)
۰/۹۵۶	۰/۹۰۷	Accuracy (%)
۰/۹۱۲۹	۰/۸۱۶	Kappa
۰/۲۴۲۱	۰/۲۹۲	RMSE
۰/۹۹۴	۰/۹۳۱	AUC

زمین لغزش داشته است، کلاس های با حساسیت زیاد و خیلی زیاد (رنگ قرمز در نقشه ها) در مناطقی با بیشترین تراکم جاده متمرکز شدند. مساحت طبقات پتانسیل وقوع زمین لغزش دو مدل دارای مقادیر نزدیک به هم بودند. همچنین نتایج نشان داد در مدل های جنگل تصادفی و درخت تصمیم به ترتیب ۳۱/۷۵ و ۳۲/۰۵ درصد از سطح منطقه در طبقات پتانسیل وقوع

در این پژوهش، نقشه های پیش بینی مکانی زمین لغزش بر اساس الگوریتم های جنگل تصادفی و درخت تصمیم تهیه شدند (شکل ۳). مساحت و درصد مساحت طبقات پتانسیل وقوع زمین لغزش بر اساس دو مدل جنگل تصادفی و درخت تصمیم در جدول (۳) نشان داده شده است. از آن جایی که تراکم جاده نسبت به سایر فاکتورها بیشترین تاثیر را بر وقوع

زمین‌لغزش زیاد و خیلی زیاد واقع شده است.



شکل ۳. نقشه پتانسیل وقوع زمین‌لغزش با استفاده از مدل‌های ترکیبی درخت تصمیم و جنگل تصادفی

جدول ۳. مساحت و درصد مساحت طبقات پتانسیل وقوع زمین‌لغزش بر اساس دو مدل جنگل تصادفی و درخت تصمیم

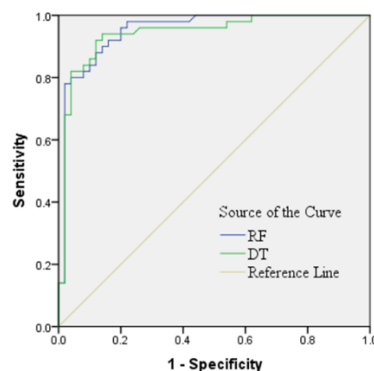
مدل درخت تصمیم		مدل جنگل تصادفی		طبقات پتانسیل وقوع زمین‌لغزش	
درصد رخداد	درصد مساحت	درصد رخداد	درصد مساحت	مساحت طبقات (هکتار)	زمین‌لغزش
۰	۱۱/۹۶	۰	۱۰/۶۵	۱۱۱۱۳/۵	خیلی کم
۴	۲۵/۷۹	۲	۲۶/۶۸	۲۷۸۳۶/۲	کم
۱۲	۳۰/۱۹	۱۴	۳۰/۹۲	۳۲۲۵۳/۹	متوسط
۴۰	۲۲/۳۸	۴۸	۲۲/۴۹	۲۳۴۵۶/۶	زیاد
۴۴	۹/۶۷	۳۶	۹/۲۶	۹۶۵۹/۲	خیلی زیاد

مدل جنگل تصادفی و درخت تصمیم مورد مطالعه، بیشترین صحت به مدل جنگل تصادفی (۰/۹۵۱) اختصاص داشت. بنابراین در زمینه پتانسیل‌یابی وقوع زمین‌لغزش، مدل جنگل تصادفی نسبت به مدل درخت تصمیم دارای عملکرد بهتری می‌باشد.

به منظور اعتبارسنجی نقشه‌های پتانسیل وقوع زمین‌لغزش از منحنی ROC استفاده شد. مقدار AUC برای مدل‌های مورد ارزیابی با توجه به داده‌های اعتبارسنجی در جدول (۴) ارائه شده است. همچنین منحنی ROC مدل‌های مورد ارزیابی با توجه به داده‌های اعتبارسنجی در شکل (۴) ارائه شد. از میان

جدول ۴. مقدار AUC مدل‌های پیش‌بینی پتانسیل وقوع زمین‌لغزش

داده‌های اعتبارسنجی	مدل پیش‌بینی	ردیف
۰/۹۵۱	جنگل تصادفی	۱
۰/۹۴۲	درخت تصمیم	۲



شکل ۴. منحنی ROC مدل‌های استفاده شده با توجه به داده‌های اعتبارسنجی

بحث و نتیجه گیری

تکنیک‌های یادگیری ماشین توانایی فراوانی برای پردازش داده‌های بزرگ دارند (Colkesen *et al.*, 2016). از جمله تکنیک‌های یادگیری ماشین می‌توان به رگرسیون لجستیک، ماشین‌های بردار پشتیبان (Hong *et al.*, 2017) و جنگل تصادفی (Sevgen *et al.*, 2019) اشاره کرد. در این مطالعه به منظور پتانسیل‌یابی وقوع زمین‌لغزش در شهرستان سروآباد از دو مدل جنگل تصادفی و درخت تصمیم استفاده شد. همچنین به منظور تعیین وزن و اهمیت طبقات هر پارامتر از روش نسبت فراوانی استفاده شد.

بیشترین میزان رخداد زمین‌لغزش در جهت‌های جنوب‌شرقی و جنوب‌غربی و شیب‌های کمتر از ۲۰ درجه مشاهده شد. Wang و Li (۲۰۱۷) و Nohani و همکاران (۲۰۱۹) گزارش کردند که احتمال وقوع زمین‌لغزش با افزایش شیب تا حدودی افزایش و سپس کاهش می‌یابد. همچنین نتایج مدل نسبت فراوانی نشان داد بیشترین رخداد زمین‌لغزش در کاربری‌های جنگل و کشاورزی اتفاق افتاده است. Akinci و همکاران (۲۰۲۰) در این خصوص به نتایج مشابهی دست یافتند و اظهار داشتند مناطق تحت کاربری کشاورزی مستعد وقوع زمین‌لغزش هستند. در مطالعه‌ای که توسط Bulut و Dağ (۲۰۱۲) در منطقه چایلی در استان ریزه در ترکیه انجام شد، بیشتر زمین‌لغزش‌ها در منطقه مورد مطالعه در دامنه‌هایی با جنبه شمالی-شمال شرقی رخ داده است.

بیشترین رخداد زمین‌لغزش در طبقات ارتفاع ۱۱۰۰-۱۹۰۰ متر مشاهده شد که این امر می‌تواند به این دلیل باشد که تمرکز فعالیت‌های انسان‌ساخت از جمله احداث جاده در این طبقه ارتفاعی باشد. همچنین Zhou و همکاران (۲۰۲۱) اظهار داشتند فراوانی وقوع زمین‌لغزش‌ها با افزایش ارتفاع کاهش می‌یابد. Kavzoglu و همکاران (۲۰۱۴) گزارش کردند تأثیر ارتفاع بر روی حساسیت زمین‌لغزش موضوعی بحث‌برانگیز است و هنوز توسط محققان روشن نشده است.

همچنین بیشترین رخداد زمین‌لغزش در فاصله کمتر از ۶۰۰ متر از رودخانه رخ می‌دهد. Akinci و همکاران (۲۰۲۰) اظهار داشتند با افزایش فاصله از شبکه‌های زهکشی و جاده‌ها، احتمال زمین‌لغزش کاهش می‌یابد. مناطق با میزان بارش متوسط (کمتر از ۷۲۰ میلی‌متر)، لیتولوژی سنگ آهک ریفی و شیل

خاکستری، با فاصله از جاده کمتر از ۱۲۰۰ متر و مناطق با کمترین و بیشترین میزان NDVI با شاخص توان آبراهه متوسط و با میزان شاخص رطوبت توپوگرافیک بالاتر دارای بالاترین رخداد زمین‌لغزش بودند. فاصله از جاده‌ها عامل انسانی مهمی است که نقش موثری در ایجاد زمین‌لغزش دارد. در مناطق تپه ماهوری جاده‌ها یکی از عوامل اصلی وقوع زمین‌لغزش بوده که بخشی از فعالیت‌های توسعه انسانی را نشان می‌دهد (Sujatha *et al.*, 2012). همچنین Akinci و همکاران (۲۰۲۰) گزارش کردند احتمال وقوع زمین‌لغزش با افزایش فاصله از گسل‌ها کاهش می‌یابد.

نتایج این مطالعه نشان داد متغیرهای تراکم جاده، ارتفاع و تراکم آبراهه نسب به سایر متغیرها بیشترین تأثیر را بر وقوع زمین‌لغزش داشتند. نتایج مطالعه Akinci و همکاران (۲۰۲۰) نیز نشان داد مهم‌ترین عوامل در وقوع زمین‌لغزش در منطقه مورد مطالعه شیب، ارتفاع، سنگ‌شناسی، فاصله از گسل‌ها و پوشش زمین بوده که با نتایج مطالعه کنونی مغایرت دارد. این تفاوت می‌تواند به سبب تفاوت در نوع زمین‌لغزش‌های اتفاق افتاده، تفاوت در وضعیت اقلیمی، لیتولوژی و میزان دخالت انسان در طبیعت منطقه باشد. نتایج مطالعه Cheng و همکاران (۲۰۲۱) نشان داد کاربری زمین، فاصله تا زمین‌لغزش‌های قبلی، فاصله تا جاده‌ها، تغذیه سالانه آب زیرزمینی، تراکم گسل، جهت جغرافیایی و میزان شیب تأثیرگذارترین عوامل ایجاد کننده زمین‌لغزش در چین هستند.

در این مطالعه مدل جنگل تصادفی نسبت به درخت تصمیم دارای عملکرد بهتری در زمینه پتانسیل‌یابی وقوع زمین‌لغزش بود. نتایج مطالعه Dou و همکاران (۲۰۱۹) نیز به نتایج مشابهی دست یافت. Wang و همکاران (۲۰۲۰) از مدل جنگل تصادفی و نسبت فراوانی برای ارزیابی پتانسیل وقوع زمین‌لغزش در چین استفاده کردند و به این نتیجه رسیدند ارتفاع، میانگین بارندگی سالانه، شیب، سنگ‌شناسی، فاصله از جاده‌ها و فاصله از رودخانه‌ها عوامل اصلی ایجادکننده زمین‌لغزش هستند. Arumugam و همکاران (۲۰۲۳) در مطالعه‌ای که انجام دادند نقشه پتانسیل خطر زمین‌لغزش را در هند تهیه کردند. نتایج مطالعه آنها نشان داد مناطق شمال‌شرق، جنوب و جنوب‌شرق منطقه مورد مطالعه در برابر زمین‌لغزش آسیب‌پذیرتر هستند. Sevgen و همکاران (۲۰۱۹) عملکرد

روستایی شهرستان سروآباد. نشریه تحقیقات کاربردی علوم جغرافیایی، ۱۴(۳۵): ۱۱۷-۱۳۸.

سایت رسمی فرمانداری شهرستان سروآباد. (۱۳۹۰) موقعیت جغرافیایی شهرستان سروآباد. قابل دسترس در سایت: <https://sarvabad.ostan-kd.ir/>

Achour, Y. and Pourghasemi, H.R. (2020) How do machine learning techniques help in increasing accuracy of landslide susceptibility maps. *Geoscience Frontiers*, 11(3): 871-883.

Ado, M., Amitab, K., Maji, A.K., Jasińska, E., Gono, R., Leonowicz, Z. and Jasiński, M. (2022) Landslide susceptibility mapping using machine learning: A literature survey. *Remote Sensing*, 14(13); 25-35.

Akinci, H., Kilicoglu, C. and Dogan, S. (2020) Random forest-based landslide susceptibility mapping in coastal regions of Artvin, Turkey. *ISPRS International Journal of Geo-Information*, 9(9): 553-553.

Arumugam, T., Kinattinkara, S., Velusamy, S., Shanmugamoorthy, M. and Murugan, S. (2023) GIS based landslide susceptibility mapping and assessment using weighted overlay method in Wayanad: A part of Western Ghats, Kerala. *Urban Climate*, 49(1): 101508.

Basu, T. and Pal, S. (2019) RS-GIS based morphometrical and geological multi-criteria approach to the landslide susceptibility mapping in Gish River Basin, West Bengal, India. *Advances in Space Research*, 63(3): 1253-1269.

Breiman, L. (1999) Random forests—Random features. Technical Report 567, Statistics Department, University of California, Berkeley: Berkeley, CA, USA, 29p.

Chen, W., Xie, X., Wang, J., Pradhan, B., Hong, H., Bui, D.T., Duan, Z. and Ma, J. (2017) A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *Catena*, 151(1): 147-160.

Cheng, Y.S., Yu, T.T. and Son, N.T. (2021) Random forests for landslide prediction in tsengwen river watershed, central taiwan. *Remote Sensing*, 13(2): 199-208.

Colkesen, I., Sahin, E.K. and Kavzoglu, T. (2016) Susceptibility mapping of shallow landslides using kernel-based Gaussian process, support vector machines and logistic regression. *Journal of African Earth Sciences*, 118(2016): 53-64.

Crosby, D.A. (2006) The effect of DEM resolution on the computation of hydrologically

مدل‌های رگرسیون لجستیک، جنگل تصادفی و شبکه عصبی مصنوعی را در پتانسیل‌یابی خطر زمین‌لغزش مقایسه کردند و نتایج نشان داد مدل جنگل تصادفی از قابلیت‌های پیش‌بینی بهتری نسبت به دو مدل دیگر برخوردار هستند.

Merghadi و همکاران (۲۰۲۰) گزارش دادند الگوریتم‌های مبتنی بر درخت در مقایسه با سایر الگوریتم‌های یادگیری ماشین نتایج بسیار خوبی ارائه می‌دهد و مدل جنگل تصادفی عملکرد مناسبی را برای تهیه نقشه پتانسیل وقوع زمین‌لغزش در پی دارد. به‌طور مشابه، مطالعات متعدد گزارش کردند که مدل جنگل تصادفی نتایج دقیق‌تری را در مقایسه با سایر مدل‌های یادگیری ماشین ارائه می‌دهد (Chen *et al.*, 2017; Achour & Pourghasem, 2020).

در نهایت در این مطالعه، عملکرد دو مدل جنگل تصادفی و درخت تصمیم در زمینه پتانسیل‌یابی وقوع زمین‌لغزش در شهرستان سروآباد مورد بررسی قرار گرفت. همچنین به‌منظور پتانسیل‌یابی وقوع زمین‌لغزش در منطقه مورد مطالعه، ۱۶ پارامتر در نظر گرفته شد که ۱۳ پارامتر شامل شیب، ارتفاع، فاصله از آبراهه، فاصله از جاده، تراکم رودخانه، فاصله از گسل، تراکم جاده، کاربری اراضی، NDVI، لیتولوژی، بارندگی، بزرگی زمین‌لرزه و شاخص رطوبت توپوگرافی در وقوع زمین‌لغزش نقش داشتند. نتایج حاصل از مقادیر AUC نشان داد هر دو مدل دارای عملکرد مناسبی هستند، اما الگوریتم جنگل تصادفی نسبت به الگوریتم درخت تصمیم دارای عملکرد مناسب‌تری در زمینه پتانسیل‌یابی وقوع زمین‌لغزش می‌باشد. با توجه به داده‌های اعتبارسنجی، تعداد زمین‌لغزش‌های به وقوع پیوسته در طبقات خطر زیاد و خیلی‌زیاد در مدل‌های جنگل تصادفی و درخت تصمیم به ترتیب ۸۴ درصد بود. نقشه‌های پتانسیل وقوع زمین‌لغزش، ابزارهای کارآمدی است، به‌طوری‌که می‌توان آنها را برای مدیریت زیست‌محیطی، برنامه‌ریزی کاربری زمین و توسعه زیرساخت‌ها مورد استفاده قرار داد. بنابراین به‌منظور پیشگیری از وقوع زمین‌لغزش در مناطق پرخطر، بایستی جنگل‌کاری با درختان جنگلی ریشه‌دار انجام شود.

منابع

زمانی، ل. و ریاحی، و. (۱۳۹۳) مدیریت بحران و شناخت پهنه‌های خطر و امن ناشی از زمین‌لغزش در نواحی

- Merghadi, A., Yunus, A.P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D.T., Avtar, R. and Abderrahmane, B. (2020) Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. *Earth-Science Reviews*, 207(1): 103-225.
- Nefeslioglu, H.A., Sezer, E., Gokceoglu, C., Bozkir, A.S. and Duman, T.Y. (2010) Assessment of landslide susceptibility by decision trees in the metropolitan area of Istanbul, Turkey. *Mathematical Problems in Engineering*, 2010(1): 1-15.
- Nicodemus, K.K. (2011) On the stability and ranking of predictors from random forest variable importance measures. *Briefings in Bioinformatics*, 12(4): 369-373.
- Nohani, E., Moharrami, M., Sharafi, S., Khosravi, K., Pradhan, B., Pham, B.T., Lee, S and M. Melesse, A. (2019) Landslide susceptibility mapping using different GIS-based bivariate models. *Water*, 11(7): 1402-1402.
- Othman, A.A., Gloaguen, R., Andreani, L. and Rahnema, M. (2018) Improving landslide susceptibility mapping using morphometric features in the Mawat area, Kurdistan Region, NE Iraq: Comparison of different statistical models. *Geomorphology*, 319(1): 147-160.
- Prasad, P., Loveson, V.J., Das, S. and Chandra, P. (2021) Artificial intelligence approaches for spatial prediction of landslides in mountainous regions of western India. *Environmental Earth Sciences*, 80(1): 1-20.
- Reichenbach, P., Rossi, M., Malamud, B.D., Mihir, M. and Guzzetti, F. (2018) A review of statistically-based landslide susceptibility models. *Earth-Science Reviews*, 180(1): 60-91.
- Sassa, K., Mikoš, M., Sassa, S., Bobrowsky, P.T., Takara, K. and Dang, K. (Eds.) (2022) Understanding and reducing landslide disaster risk: Volume 1 Sendai Landslide Partnerships and Kyoto Landslide Commitment, Springer Nature, 1(1): 65-81.
- Sevgen, E., Kocaman, S., Nefeslioglu, H.A. and Gokceoglu, C. (2019) A novel performance assessment approach using photogrammetric techniques for landslide susceptibility mapping with logistic regression, ANN and random forest. *Sensors*, 19(18): 3940.
- Shirzadi, A., Soliamani, K., Habibnejhad, M., Kaviani, A., Chapi, K., Shahabi, H., Chen, W., Khosravi, K., Thai Pham, B., Pradhan, B. and Ahmad, A. (2018) Novel GIS based machine learning algorithms for shallow landslide susceptibility mapping. *Sensors*, 18(11): 3777.
- Stumpf, A. and Kerle, N. (2011) Object-oriented mapping of landslides using Random Forests. significant topographic attributes. M.A. Thesis, University of South Florida.
- Dağ, S. and Bulut, F. (2012) An example for preparation of GIS-based landslide susceptibility maps: Çayeli (Rize, NE Türkiye). *Jeoloji Mühendisliği Dergisi*, 36(1): 35-62.
- Devkota, K.C., Regmi, A.D., Pourghasemi, H.R., Yoshida, K., Pradhan, B., Ryu, I.C., Dhital, M. and Althuwaynee, O.F. (2013) Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in GIS and their comparison at Mugling–Narayanghat road section in Nepal Himalaya. *Natural Hazards*, 65(1): 135-165.
- Dou, J., Yunus, A.P., Bui, D.T., Merghadi, A., Sahana, M., Zhu, Z., Chen, C.W., Khosravi, K. and Pham, B.T. (2019) Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan. *Science of the total environment*, 662(1): 332-346.
- Dou, J., Yunus, A.P., Merghadi, A., Wang, X.K. and Yamagishi, H. (2021) A comparative study of deep learning and conventional neural network for evaluating landslide susceptibility using landslide initiation zones. *Understanding and reducing landslide disaster risk: Volume 2 from Mapping to Hazard and Risk Zonation* (5th): 215-223.
- Froude, M.J. and Petley, D.N. (2018) Global fatal landslide occurrence from 2004 to 2016. *Natural Hazards and Earth System Sciences*, 18(8): 2161-2181.
- Hamza, T. and Raghuvanshi, T.K. (2017) GIS based landslide hazard evaluation and zonation: A case from Jeldu District, Central Ethiopia. *Journal of King Saud University-Science*, 29(2): 151-165.
- Hong, H., Liu, J., Zhu, A.X., Shahabi, H., Pham, B.T., Chen, W., Pradhan, B. and Bui, D.T. (2017) A novel hybrid integration model using support vector machines and random subspace for weather-triggered landslide susceptibility assessment in the Wuning area (China). *Environmental Earth Sciences*, 76(19): 1-19.
- Jebur, M.N., Pradhan, B. and Tehrany, M.S. (2014) Optimization of landslide conditioning factors using very high-resolution airborne laser scanning (LiDAR) data at catchment scale. *Remote Sensing of Environment*, 152(1): 150-165.
- Kavzoglu, T., Sahin, E.K. and Colkesen, I. (2014) Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression. *Landslides*, 11(3): 425-439.

- Wang, Y., Sun, D., Wen, H., Zhang, H. and Zhang, F. (2020) Comparison of random forest model and frequency ratio model for landslide susceptibility mapping (LSM) in Yunyang County (Chongqing, China). *International Journal of Environmental Research and Public Health*, 17(12): 42-51.
- Xiao, T., Segoni, S., Chen, L., Yin, K. and Casagli, N. (2020) A step beyond landslide susceptibility maps: A simple method to investigate and explain the different outcomes obtained by different approaches. *Landslides*, 17(3): 627-640.
- Yu, X. and Gao, H. (2020) A landslide susceptibility map based on spatial scale segmentation: A case study at Zigui-Badong in the Three Gorges Reservoir Area, China. *PLoS One*, 15(3): e0229818.
- Zhou, X., Wen, H., Zhang, Y., Xu, J. and Zhang, W. (2021) Landslide susceptibility mapping using hybrid random forest with GeoDetector and RFE for factor optimization. *Geoscience Frontiers*, 12(5): 101211.
- Zhu, A.X., Miao, Y., Yang, L., Bai, S., Liu, J. and Hong, H. (2018) Comparison of the presence-only method and presence-absence method in landslide susceptibility mapping. *Catena*, 171(1): 222-233.
- Remote Sensing of Environment, 115(10): 2564-2577.
- Sujatha, E.R., Rajamanickam, G.V. and Kumaravel, P. (2012) Landslide susceptibility analysis using probabilistic certainty factor approach: a case study on Tevankarai stream watershed, India. *Journal of earth system science*, 121(5): 1337-1350.
- Sun, D., Wen, H., Wang, D. and Xu, J. (2020) A random forest model of landslide susceptibility mapping based on hyperparameter optimization using Bayes algorithm. *Geomorphology*, 362(1): 107201.
- Tien Bui, D., Pradhan, B., Lofman, O. and Revhaug, I. (2012) Landslide susceptibility assessment in Vietnam using support vector machines, decision tree, and Naive Bayes Models. *Mathematical Problems in Engineering*, 2012(6): 1-26.
- Trigila, A., Iadanza, C., Esposito, C. and Scarascia-Mugnozza, G. (2015) Comparison of Logistic Regression and Random Forests techniques for shallow landslide susceptibility assessment in Giampileri (NE Sicily, Italy). *Geomorphology*, 249(1): 119-136.
- Wang, Q. and Li, W. (2017) A GIS-based comparative evaluation of analytical hierarchy process and frequency ratio models for landslide susceptibility mapping. *Physical Geography*, 38(4): 318-337.

Landslide susceptibility mapping using advanced machine learning algorithms (Case study: Sarovabad city, Kurdistan province)

Hemen Rastkhadiv¹, Baharak Motamedvaziri^{2*}, Seied Akbar Javadi² and Hasan Ahmadi³

- 1) Ph.D. Student, Department of Nature Engineering, Faculty of Natural Resources and Environment, Science and Research Branch, Islamic Azad University, Tehran, Iran.
- 2) Associate Professor, Department of Nature Engineering, Faculty of Natural Resources and Environment, Science and Research Branch, Islamic Azad University, Tehran, Iran.
*Corresponding Author Email address: bmvaziri@gmail.com
- 3) Full Professor, Department of Reclamation of Arid and Mountainous Regions, University of Tehran, Karaj, Iran.

Date of Submission: 2023/12/24

Date of Acceptance: 2024/03/13

Abstract

The occurrence of landslides in mountainous regions can cause severe damage to road infrastructure and can also lead to the death of people. Therefore, the aim of this study is to map landslide susceptibility using advanced machine learning algorithms in the city of Sarovabad (Kurdistan province). In this study, landslide susceptibility was determined using two advanced data mining algorithms including Random Forest (RF) and Decision Tree (DT). Initially, the point file of 166 landslides occurred in Sarovabad city was considered as a map of landslide inventory. The landslide points are divided into training data (70%) and validation data (30%). A total of 16 parameters including slope, aspect, elevation, river proximity, road proximity, river density, fault proximity, fault density, road density, precipitation, land use, NDVI, lithology, earthquake, stream power index (SPI) and topographic wetness index (TWI) were used for landslide susceptibility mapping. Finally, the performance of the models was evaluated using the ROC curve. The results of the ROC curve show that the decision tree and random forest models have AUC values of 0.942 and 0.951, respectively. Therefore, the random forest model has the highest AUC value compared to the decision tree and was the best model for predicting the risk of landslides in the future in the study area. Landslide potential maps are efficient tools that can be used for environmental management, land use planning and infrastructure development.

Keywords: Decision tree algorithm, Kurdistan province, Landslide, Natural hazards, Random forest algorithm.