

استفاده از الگوریتم انتخاب ویژگی به منظور پیش‌بینی زمین‌لغزش در جنوب غرب ایران

مرضیه مکرم*: گروه مرتع و آبخیزداری، دانشکده‌ی کشاورزی و منابع طبیعی داراب، دانشگاه شیراز

سعید نگهبان: گروه جغرافیا، دانشکده‌ی اقتصاد، مدیریت و علوم اجتماعی، دانشگاه شیراز

تاریخچه مقاله (تاریخ دریافت: ۱۳۹۵/۰۹/۰۳ تاریخ پذیرش: ۱۳۹۶/۰۲/۲۵)

چکیده

حرکات توده‌ای دامنه‌ها یکی از مهمترین مخاطراتی است که همه ساله حیات انسان‌ها را تهدید می‌کند و پرداختن به این موضوع از اهمیت زیادی برخوردار است. یکی از مهمترین حرکات توده‌ای دامنه‌ای، زمین‌لغزش است که تهیه‌ی نقشه‌های حساسیت به آن به منظور مدیریت مخاطرات، هدف مهمی در علوم طبیعی به حساب می‌آید. با توجه به اهمیت موضوع، هدف از این مطالعه استفاده از الگوریتم انتخاب ویژگی به منظور تعیین مؤثرترین پارامترها در پیش‌بینی زمین‌لغزش در جنوب غرب ایران است. برای این منظور از روش‌های بهترین انتخاب، رتبه‌بندی و گام به گام استفاده شد. نتایج حاصل از روش بهترین انتخاب نشان داد که مؤثرترین داده‌ها برای زمین‌لغزش SAR، ارتفاع، شاخص TWI و فاصله از رودخانه است. همچنین نتایج حاصل از روش گام به گام، مشابه با روش بهترین انتخاب به شمار می‌رود؛ در حالی که مؤثرترین داده‌ها برای روش رتبه‌بندی، شاخص TWI، ارتفاع، فاصله از جاده و فاصله از رودخانه است. داده‌های انتخاب شده توسط روش رتبه‌بندی با کمترین میزان خطا و حداکثر ضریب همبستگی (۸۷/۵)، با طبقه‌بندی LMT به عنوان مؤثرترین داده‌ها برای تعیین زمین‌لغزش به شمار می‌رود. همچنین با توجه به اینکه هدف از این مطالعه بررسی زمین‌لغزش به طور مکانی است؛ بنابراین بعد از تعیین پارامترهای مؤثر در زمین‌لغزش، نقشه‌های پهنه‌بندی آن با استفاده از مؤثرترین داده‌ها در محیط GIS تهیه شد. نتایج ناشی از تهیه‌ی زمین‌لغزش با استفاده از داده‌های انتخاب شده، توسط الگوریتم انتخاب ویژگی نشان داد که بخش‌هایی از مناطق شمالی و جنوب شرق منطقه مورد مطالعه در معرض خطر لغزش بیشتری قرار دارند.

واژگان کلیدی: الگوریتم انتخاب ویژگی، زمین‌لغزش، جنوب غرب ایران.

۱- مقدمه

امروزه انسان با تغییر کاربری اراضی و گسترش شبکه‌های ارتباطی، حساسیت وقوع زمین‌لغزش‌ها را در سطح کره‌ی زمین به ویژه در نواحی کوهستانی تشدید کرده است (Gorsevski et al, 2006). در قرن بیستم، قاره‌ی آسیا بیشترین حادثه وقوع زمین‌لغزش - که ۲۲۰ مورد گزارش شده - را به خود اختصاص داده است. قاره‌ی آمریکا بیشترین تعداد کشته و مجروحان را داشته (بیش از ۲۵۰۰ نفر) و اروپا بالاترین میزان خسارات را تجربه کرده است (Karami, 2013).

پدیده‌ی زمین‌لغزش یکی از پدیده‌های مهم و قابل توجه در مسائل زیست‌محیطی، آبخیزداری و منابع طبیعی است. اهمیت زمین‌لغزش‌ها را می‌توان از دیدگاه‌های گوناگونی مورد بحث و بررسی قرار داد. مهم‌ترین دلیل با اهمیت بودن

این پدیده، خسارت‌های جانی و مالی ناشی از به وقوع پیوستن آنها است (Rajab zadeh, 2014).

بررسی روابط دینامیکی بین فاکتورهای مؤثر در وقوع زمین‌لغزش، در موفقیت این ارزیابی از اهمیت بالایی برخوردار است. برای تعیین روابط بین انواع فاکتورهای محیطی و وقوع زمین‌لغزش، تحقیقات مختلفی صورت گرفته است (Anbalagan, 1992 & Ayalew and Yamagishi, 2005). برخی از این تحقیقات، روابط منسجمی را بین توزیع فاکتورهای زمین‌شناسی و ژئومورفولوژیک و زمین‌لغزش مشاهده کردند. با این وجود برای تحلیل نتایج این تحقیقات و پیش‌بینی احتمال وقوع زمین‌لغزش‌ها، از ابزارهای محاسبات آماری متداول استفاده شد. در حقیقت، پیش‌بینی احتمال وقوع زمین‌لغزش نیازمند مدل‌های کمی در ترکیب رخدادهای گذشته‌است و با استفاده از داده‌های میدانی یا آزمایشگاهی به مدل‌سازی می‌پردازد (Melchiorre et al, 2008). درمان‌کونی و گیردان به پیش‌بینی زمین‌لغزش در شمال ایتالیا پرداختند که در این مطالعه، حد آستانه‌ی زمین‌لغزش مشخص شد. همچنین در این نتایج مشخص شد که به کمک این حد آستانه می‌توان زمین‌لغزش را مدیریت کرد و خطرات احتمالی را کاهش داد (Manconi and Giordan, 2015). ژانگ و همکاران^۱ در مطالعه‌ای به پهنه‌بندی خطر زمین‌لغزش براساس مدل توزیع چند وزنی در مناطق ساحلی جنوب شرق چین پرداختند. براساس نتایج پژوهش، این مدل روش مناسبی برای پهنه‌بندی خطر زمین‌لغزش است (Zhang et al, 2012).

یکی دیگر از کارهای انجام شده در مورد زمین‌لغزش توسط وان صورت گرفت. او برای تهیه‌ی نقشه‌ی حساسیت زمین‌لغزش در تایوان، با استفاده از سیستم پشتیبانی تصمیم‌گیری فضایی به استخراج عوامل اصلی زمین‌لغزش پرداخت (Wan, 2009). علاوه بر موارد ذکر شده گورسوسکی و همکاران^۲، وسترن و ترلین^۳، سلرسی و همکاران^۴ و آیالو و یاماگیشی^۵ نقشه‌های حساسیت به زمین‌لغزش را تهیه کردند (Ayalew and Yamagishi, 2005 & Gorsevski et al, 2000).

فیض‌نیا و همکاران، عوامل مؤثر بر وقوع زمین‌لغزش‌ها و پهنه‌بندی این خطر را در حوزه‌ی آبخیز شیرین‌رود-سدتن بررسی کردند. پس از بررسی‌های میدانی و مرور مطالعات انجام شده در مناطق مشابه با حوزه‌ی آبخیز شیرین‌رود، عامل ارتفاع از سطح دریا، شیب، جهت، بارندگی، کاربری زمین‌ها، زمین‌شناسی، فاصله از جاده، گسل و شبکه‌ی هیدروگرافی به عنوان عوامل مؤثر اولیه در وقوع زمین‌لغزش منطقه تشخیص داده شدند. نتایج نشان داد مناطقی که در آنها تغییرات زیاد کاربری زمین‌ها صورت گرفته‌است، جزء مناطق پر خطر محسوب می‌شود و مناطق اطراف دهانه‌ی خروجی حوزه نیز در وقوع زمین‌لغزش پتانسیل بالایی دارد (Fayznia et al, 2003).

بای و میررحیمی، نقشه‌ی پهنه‌بندی زمین‌لغزش را با استفاده از روش تحلیل سلسله مراتبی انجام دادند. به منظور بررسی حرکات توده‌های حوضه‌ی آبریز رودخانه‌ی مادرسو، از عوامل شیب، جهت شیب، لیتولوژی، کاربری ارضی، باران، فاصله از رودخانه، گسل و جاده برای پهنه‌بندی خطر لغزش استفاده شد. پس از تجزیه و تحلیل داده‌ها توسط نرم-

¹ Zhang et al

² Gorsevski et al

³ Westen and Terlien

⁴ Clerici

⁵ Ayalew and Yamagishi

افزار ArcGIS، تأثیر هر کدام از لایه‌های اطلاعاتی در وقوع لغزش مشخص شد و پهنه‌بندی صورت گرفت. نتایج حاصل از مقایسه‌ی این پهنه‌بندی و پراکندگی موجود از زمین‌لغزش‌ها نشان دادند که روش سلسله مراتبی، روش مناسبی برای پهنه‌بندی خطر لغزش حوضه‌ی آبریز مادرسو است (Bay & Hajimir rahimi, 2009).

بهنیاfer و همکاران، با استفاده از روش‌های AHP و احتمال به پهنه‌بندی خطر زمین‌لغزش و ناپایداری دامنه‌ای در حوضه‌ی آبریز رودخانه‌ی کنگ، دامنه‌های شمالی بینالود پرداختند. نتایج این پژوهش به صورت نقشه‌های نواحی مستعد لغزش ارائه شد که می‌تواند برای برنامه‌ریزان محلی و منطقه‌ای قابل استفاده باشد. در بین معیارهای تأثیرگذار بر وقوع زمین‌لغزش‌ها و دیگر حرکات توده‌ای دامنه‌ای مشابه، متغیرهای سازند زمین‌شناسی، کاربری زمین، توپوگرافی و شیب و تراکم زهکشی بیشترین وزن را داشتند (Behniafar et al, 2010).

متولی و همکاران با استفاده از روش رگرسیون لجستیک، وقوع زمین‌لغزش را در حوضه‌ی آبخیز واز بررسی کردند. نتایج نشان داد که لیتولوژی، فاصله از جاده و شیب نقش مهمی در وقوع زمین‌لغزش دارند. با استفاده از نقشه‌ی پیش‌بینی احتمال وقوع زمین‌لغزش، مشخص شد که بیش از ۵۰ درصد منطقه دارای حساسیت متوسط هستند (Motavalli, et al, 2010).

روستایی و علیزاده با استفاده از روش آنالگان به پهنه‌بندی خطر زمین‌لغزش در حوضه‌ی صوفی‌چای (مراغه) پرداختند. در این تحقیق برای هر یک از پارامترهای تأثیرگذار در زمین‌لغزش از قبیل شیب، جهت شیب و ... نقشه تهیه و در آخر از تلفیق نقشه‌های عامل، نقشه‌ی پهنه‌بندی نهایی آماده شد (Roosatei & Alizadeh, 2013).

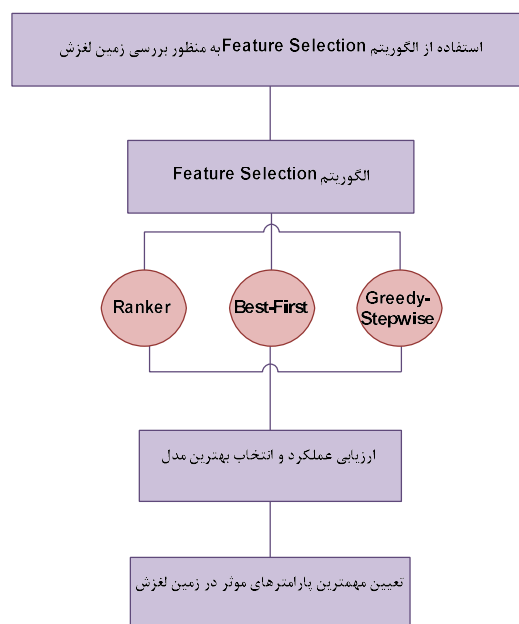
مقیمي و همکاران در مقاله‌ای با استفاده از فرآیند تحلیل شبکه‌ای (ANP)، به ارزیابی و پهنه‌بندی خطر زمین‌لغزش در شهر رودبار پرداختند. نتایج نشان داد ۷۱ درصد منطقه دارای خطر متوسط زمین‌لغزش هستند، همچنین در این فرآیند عامل شیب و حساسیت لیتولوژی مهمترین سهم را برعهده داشته‌اند (Moghimi et al, 2013).

روستایی و احمدزاده با استفاده از تکنیک سنجش از دور و GIS، پهنه‌بندی خطر زمین‌لغزش را در جاده‌ی تبریز بررسی کردند. نتایج ارزیابی نشان داد که ساختار خاص زمین‌شناسی، شرایط اقلیم محلی، تراکم نهشته‌های کواترنری و شیب از عوامل اصلی بروز زمین‌لغزش در منطقه‌ی مورد مطالعه به شمار می‌رود (Roostaei & Ahmadzadeh, 2013).

رجبی و فیض‌اله‌پور در تحقیقی تحت عنوان «پهنه‌بندی زمین‌لغزش‌های حوضه‌ی رودخانه‌ی گیوی‌چای با استفاده از مدل پرسپترون چند لایه از نوع پیش‌خور پس انتشار (BP)» برای ارزیابی شبکه عصبی ایجاد شده، داده‌های ۴۱ زمین‌لغزش رخ داده در کنار ۸ لایه متشکل از لایه‌های شیب، جهت شیب، ارتفاع، لیتولوژی، فاصله از گسل، شبکه‌ی هیدروگرافی، کاربری اراضی و پراکنش زمین‌لغزش به پهنه‌بندی زمین‌لغزش پرداخته‌اند. نتایج حاصل نشان داد که ساختار زمین‌شناسی شکل گرفته از آهک‌های کرتاسه و آندزیت‌های پرفیری همچنین دسترسی به منابع رطوبتی بالا باعث شده که ارتفاعات شرقی کوه بوغروداغ و آلاداغ در محدوده‌ی کوه‌های تالش، قابلیت بالایی در رخداد زمین‌لغزش داشته باشد (Rajabi & Fayzollah poor, 2014).

رسایی و همکاران، به بررسی زمین لغزش با استفاده از مدل رگرسیون چند متغیره در محیط GIS پرداختند. نتایج نشان داد که با تعیین مهمترین پارامترهای مؤثر در زمین لغزش، می توان به راحتی و با سرعت مکان های وقوع این پدیده را پیدا کرد (Rasaei et al, 2015).

با توجه به اینکه پارامترهای متعددی در زمین لغزش تأثیرگذار هستند، انتخاب داده های مؤثرتر به منظور کاهش وقت و هزینه مهم به نظر می رسد. مسأله ی انتخاب ویژگی، یکی از مسائلی است که در مبحث یادگیری ماشین و شناسائی آماری الگو مطرح است. این مسأله در بسیاری از کاربردها مانند طبقه بندی اهمیت به سزائی دارد؛ زیرا در این کاربردها تعداد زیادی ویژگی وجود دارد که بسیاری از آنها یا بدون استفاده هستند، یا اینکه بار اطلاعاتی چندانی ندارند. حذف نکردن این ویژگی ها مشکلی از لحاظ اطلاعاتی ایجاد نمی کند، ولی بار محاسباتی را برای کاربرد مورد نظر بالا می برد؛ علاوه بر این، باعث می شود که اطلاعات نامفید زیادی را به همراه داده های مفید ذخیره کنیم. بنابراین با توجه به اهمیت موضوع در این مطالعه، هدف از تعیین نقشه ی حساسیت زمین لغزش جنوب غرب ایران با استفاده از الگوریتم انتخاب ویژگی و انتخاب ویژگی^۱ است. مراحل الگوریتم انتخاب ویژگی به منظور انتخاب بهترین پارامترهای مؤثر در زمین لغزش در شکل ۱ نشان داده شده است.



شکل ۱: روند الگوریتم انتخاب ویژگی به منظور انتخاب بهترین پارامترهای مؤثر در زمین لغزش

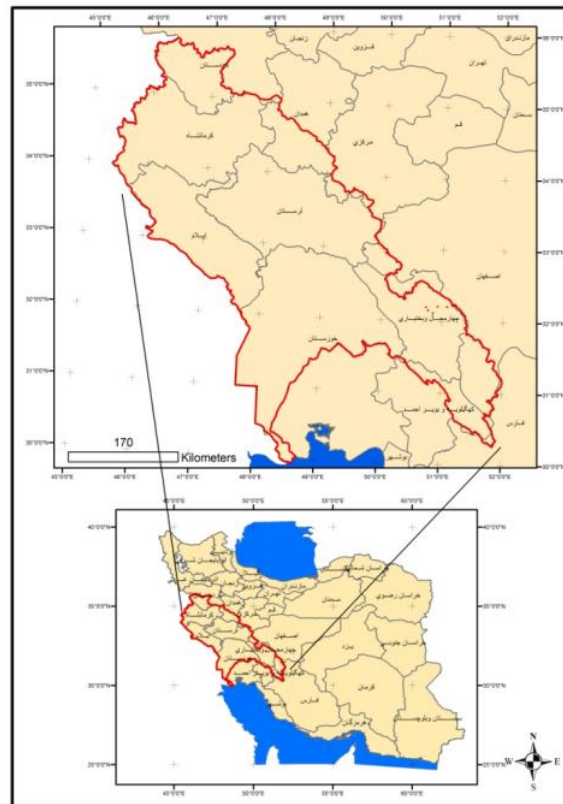
۲- مواد و روش

۲-۱- منطقه مورد مطالعه

منطقه ی مورد مطالعه در جنوب غربی ایران (بخش هایی از خوزستان، خرم آباد، ایلام، کرمانشاه، سنندج و همدان) و در موقعیت جغرافیایی ۲۹ درجه و ۵۶ دقیقه تا ۳۵ درجه و ۴۶ دقیقه شرقی و ۴۵ درجه و ۲۴ دقیقه تا ۵۲ درجه و ۱ دقیقه

^۱ Feature selection

شمالی واقع شده است (شکل ۲). مساحت منطقه‌ی مورد مطالعه ۱۵۴۲۷۲/۴۸ کیلومتر مربع است. بیشترین و کمترین ارتفاع در این منطقه به ترتیب ۳۰ و ۴۴۱۵ متر است.



شکل ۲: موقعیت منطقه‌ی مورد مطالعه

۲-۲- الگوریتم Feature selection

روش‌های مختلف انتخاب ویژگی، تلاش می‌کنند تا از میان 2^T زیر مجموعه‌ی کاندید، بهترین زیرمجموعه را پیدا کنند. در تمام این روش‌ها بر اساس کاربرد و نوع تعریف، زیر مجموعه‌ای به عنوان جواب انتخاب می‌شود که بتواند مقدار یک تابع ارزیابی را بهینه کند. با وجود اینکه هر روشی سعی می‌کند که بتواند بهترین ویژگی‌ها را انتخاب کند، اما با توجه به وسعت جواب‌های ممکن و اینکه این مجموعه‌های جواب به صورت توانی با T افزایش پیدا می‌کنند، پیدا کردن جواب بهینه‌ی مشکل و در T های متوسط و بزرگ بسیار پر هزینه است.

برای اینکه بتوانیم تقسیم‌بندی درستی از روش‌های مختلف انتخاب ویژگی داشته باشیم، به این صورت عمل می‌کنیم که فرآیند انتخاب ویژگی را در تمامی روش‌ها به این بخش‌ها تقسیم می‌کنیم:

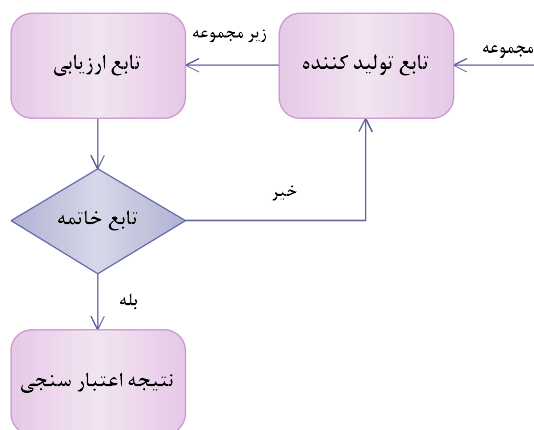
۱. تابع تولیدکننده^۱: این تابع زیر مجموعه‌های کاندید را برای روش مورد نظر پیدا می‌کند.
۲. تابع ارزیابی^۲: زیرمجموعه‌ی مورد نظر را براساس روش داده شده، ارزیابی و یک عدد به عنوان میزان خوبی روش باز می‌گرداند. روش‌های مختلف سعی در یافتن زیرمجموعه‌ای دارند که این مقدار را بهینه کند.

¹ Generation procedure

² Evaluation function

۳. شرط خاتمه: برای تصمیم‌گیری در مورد زمان توقف الگوریتم.

به کمک تابع تعیین اعتبار^۱ مشخص می‌شود که آیا زیرمجموعه‌ی انتخاب شده معتبر است یا خیر. روند تعیین اعتبار به کمک این تابع در شکل ۳ نشان داده شده‌است:



شکل ۳: مراحل الگوریتم انتخاب ویژگی و ویژگی

درخت‌های تصمیم روشی برای نمایش یک سری از قوانین هستند که به یک رده یا مقدار منتهی می‌شوند. درخت‌های تصمیمی که برای پیش‌بینی متغیرهای دسته‌ای استفاده می‌شوند، درخت‌های طبقه‌بندی نام دارند. الگوریتم‌های درخت تصمیم که در این پژوهش مورد استفاده قرار گرفته‌اند عبارتند از:

الف- الگوریتم J48

این الگوریتم سعی می‌کند تا به صورت بازگشتی مجموعه‌ی داده‌ها را با استفاده از اطلاعات نرمال به زیر مجموعه‌هایی تقسیم کند که از ریشه به سمت برگ حرکت می‌کند و در نهایت به کلاس موردنظر می‌رسد. تولید درخت تصمیم اولیه از مجموعه داده‌ها، مهمترین بخش این الگوریتم محسوب می‌شود.

ب- الگوریتم Naive Bayes

باین الگوریتم برای طبقه‌بندی داده‌ها از احتمال وقوع یا عدم وقوع پدیده‌ها استفاده می‌کند. در واقع، براساس ویژگی‌های ذاتی احتمال با آموزش و تست داده‌ها در نهایت داده‌ها را به زیر مجموعه‌هایی طبقه‌بندی می‌کند.

ج- الگوریتم LMT

این الگوریتم یک مدل طبقه‌بندی ترکیبی از پیش‌بینی لجستیک و یادگیری درخت تصمیم‌گیری لجستیک (LMT^۲) است. رگرسیون لجستیک می‌تواند یک مورد خاص از مدل خطی عمومی و رگرسیون خطی باشد.

^۱ Validation procedure

^۲ Logistic model tree

۳-۲- ارزیابی عملکرد

روش‌های مختلفی به منظور بررسی روش‌های انتخاب داده و انتخاب بهترین روش وجود دارد که از معروفترین آنها می‌توان به AMS^1 و $ARAE^2$ اشاره کرد که در این مطالعه از آنها استفاده شده است (Naseriparsa et al, 2014). در روش AMS به منظور بررسی دقت مدل از معادله‌ی زیر استفاده شد. در این روش از نسبت بین میانگین تعداد کلاس‌های نادرست پیش‌بینی شده توسط مدل به تعداد کل داده‌ها استفاده می‌شود (Dash and Liu, 2003).

$$AMS = \frac{\sum_{i=1}^n MS_i}{N} \quad (1)$$

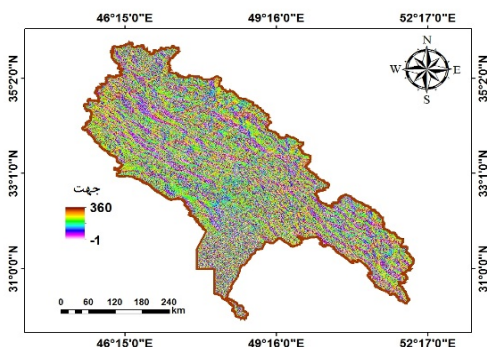
که MS_i تعداد کلاس‌های نادرست پیش‌بینی شده توسط مدل و N تعداد کل داده‌ها است. در روش $ARAE$ از میانگین خطای مطلق نسبی به کل داده‌ها استفاده می‌شود (همان).

$$ARAE = \frac{\sum_{i=1}^n RAE_i}{N} \quad (2)$$

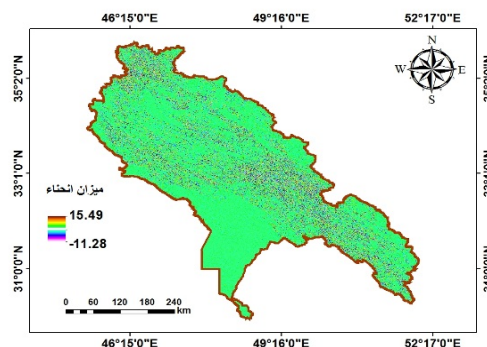
که RAE_i تعداد کلاس‌های نادرست پیش‌بینی شده توسط مدل و N تعداد کل داده‌ها است.

۳- نتایج (یافته‌ها)

به منظور تعیین نقشه‌ی حساسیت زمین‌لغزش در منطقه‌ی مورد مطالعه، پارامترهایی از قبیل جهت شیب، میزان انحناء، مدل رقومی ارتفاع (DEM^3)، پروفیل طولی، پروفیل عرضی، فاصله از رودخانه، جاده، شاخص شاخص SPI^4 ، شاخص SAR^5 و شاخص TWI^6 مطالعه و بررسی شدند (شکل ۴).



جهت شیب



میزان انحناء

¹ Average number of misclassified samples

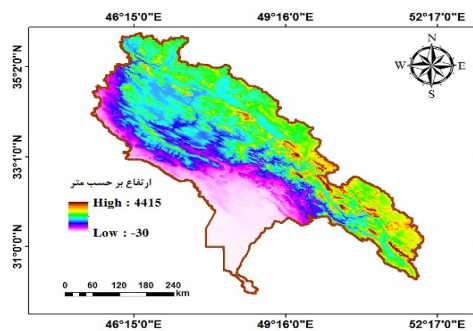
² Average relative absolute error

³ digital elevation model

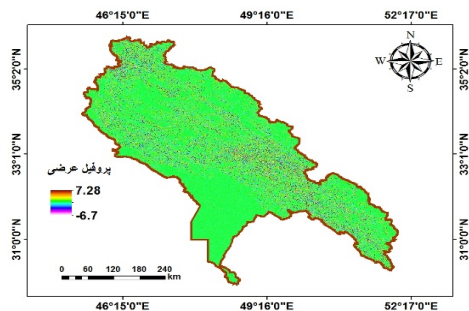
⁴ stream power index

⁵ surface area ratio

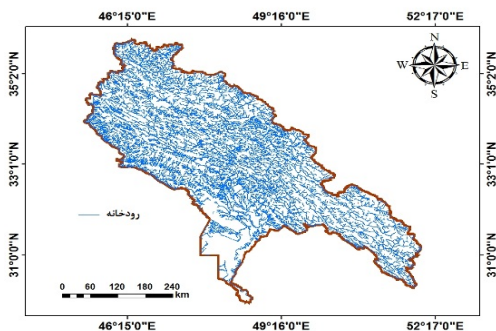
⁶ topographic wetness index



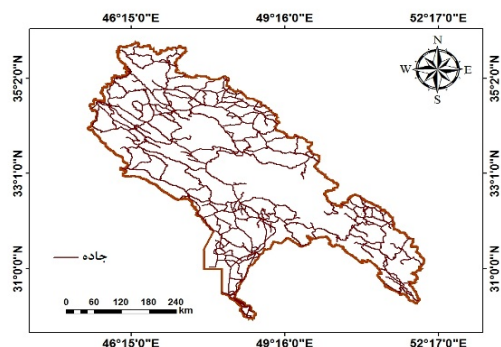
ارتفاع



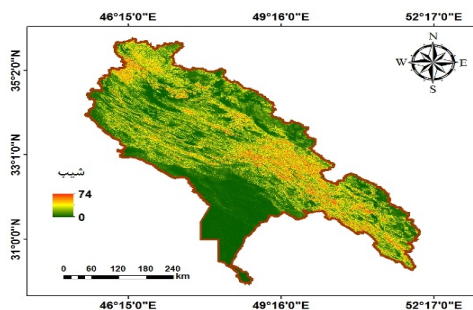
پروفیل عرضی



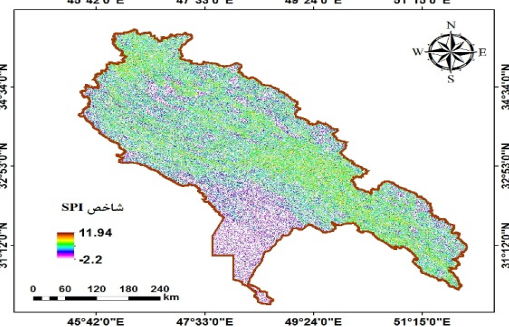
رودخانه



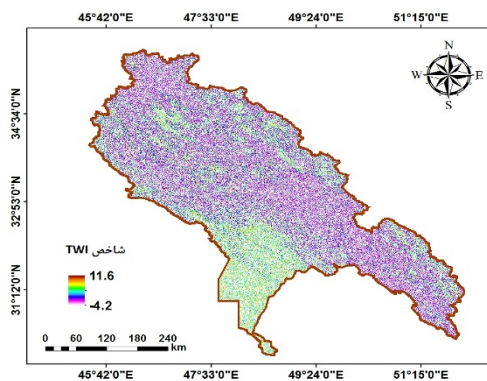
جاده



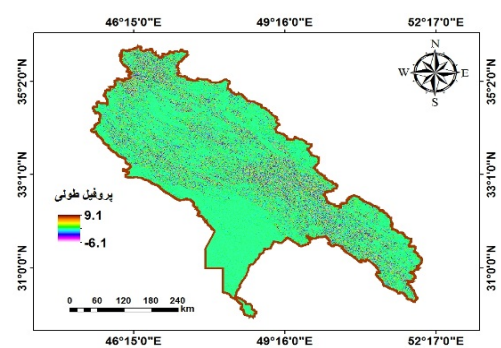
شیب



شاخص SPI



شاخص TWI

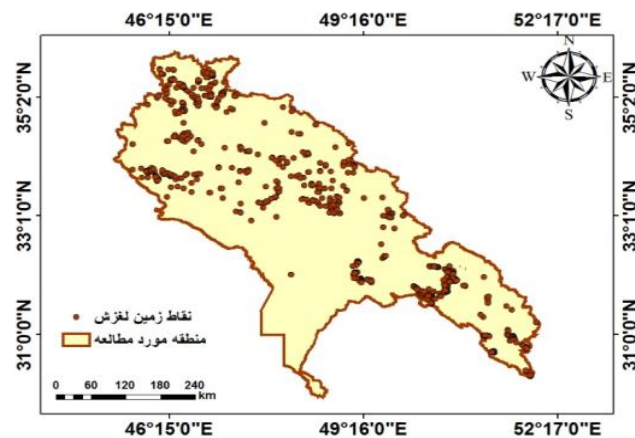


پروفیل طولی

شکل ۴: ویژگی داده‌های مورد بررسی

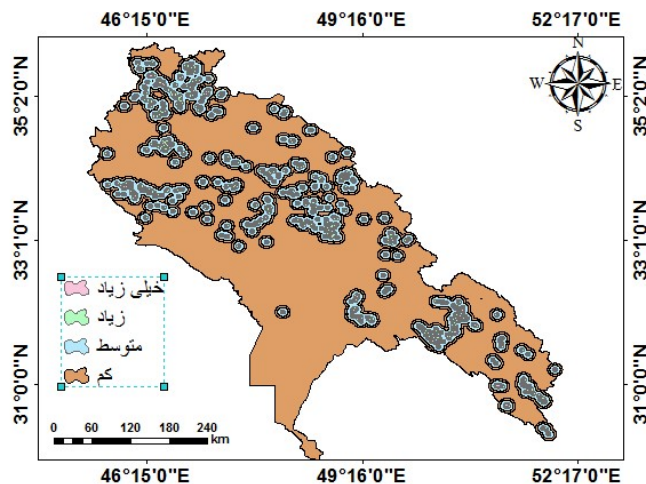
با توجه به شکل ۴ مشخص می‌شود که میزان انحناء بین ۱۱-۱۵ است که در بخش‌های جنوب غرب بیشترین میزان انحناء دیده می‌شود. همچنین بخش‌های جنوب غرب منطقه دارای ارتفاع، شیب، شاخص SPI و TWI بیشتر است که همه‌ی این عوامل در تشدید زمین‌لغزش مؤثر هستند.

همچنین مناطق دارای زمین‌لغزش در شکل ۵ نشان داده شده‌است. با توجه به شکل مشخص می‌شود که به جز بخش‌هایی از مرکز و جنوب و جنوب غرب، اکثر بخش‌های منطقه به زمین‌لغزش حساسیت دارد.



شکل ۵: مناطق دارای زمین‌لغزش

در ادامه با توجه به شکل ۳، مناطق از نظر خطر زمین‌لغزش در ۴ کلاس طبقه‌بندی شد؛ به طوری که مناطق با فاصله‌ی کمتر از ۱۰۰ متر در کلاس یک، بین ۱۰۰ تا ۱۰۰۰ متر در کلاس دو، بین ۱۰۰۰ تا ۷۰۰۰ متر در کلاس سه و بیش از ۷۰۰۰ متر در کلاس ۴ طبقه‌بندی شدند (Zhou and Fang, 2015) (شکل ۶).



شکل ۶: کلاس‌های مختلف خطر زمین‌لغزش

به منظور تعیین رابطه‌ی بین نقاط زمین‌لغزش و میزان انحناء، جهت شیب، پروفیل طولی، پروفیل عرضی، شاخص SPI، شاخص TWI، ارتفاع، فاصله از رودخانه، جاده و انتخاب بهترین و مؤثرترین داده‌ها از بین داده‌های مذکور، از

روش انتخاب ویژگی استفاده شد. در این مطالعه از سه روش Greedy Stepwise و Ranker، Best Search استفاده شد (جدول ۱). نتایج ناشی از روش Bestfirst با استفاده از CFS Subset Eval نشان داد که مؤثرترین داده‌ها برای زمین‌لغزش، SAR، ارتفاع، شاخص TWI و فاصله از رودخانه است. همچنین نتایج حاصل از روش Greedy Stepwise با استفاده از مدل CFS Subset Eval، مشابه با روش Bestfirst است؛ در حالی که مؤثرترین داده‌ها برای روش Ranker با استفاده از مدل‌های Info Gain Attribute Eval، شاخص TWI، ارتفاع، فاصله از جاده و رودخانه است. در روش Ranker با مدل Gain Ratio Attribute Eval، مؤثرترین داده‌ها ارتفاع، شاخص TWI، فاصله از جاده و شیب است. شاخص TWI، ارتفاع، فاصله از جاده و SAR مؤثرترین داده‌ها در مدل Symmetricer Attribute Eval با روش Ranker است؛ در حالی که در مدل ReliefF Attribute Eval، شاخص TWI، ارتفاع، فاصله از جاده و از رودخانه به عنوان مهمترین داده‌ها هستند. همچنین نتایج نشان داد که در مدل Principal Components میزان انحناء، شیب طولی، شیب عرضی و شاخص SPI مهم‌ترین داده‌ها برای تعیین زمین‌لغزش به شمار می‌رود.

جدول ۱: مهم‌ترین ویژگی‌های انتخاب شده توسط روش انتخاب ویژگی به منظور تعیین زمین‌لغزش

نام روش	نام مدل	ویژگی‌های انتخاب شده
Bestfirst	CFSSubsetEva	SAR، ارتفاع، شاخص TWI، فاصله از رودخانه
GreedyStepwise	CFSSubsetEval	SAR، ارتفاع، شاخص TWI، فاصله از رودخانه
Ranker	InfoGainAttributeEval	شاخص TWI، ارتفاع، فاصله از جاده، فاصله از رودخانه
	GainRatioAttributeEval	ارتفاع، شاخص TWI، فاصله از جاده، شیب
	SymmetricerAttributeEval	شاخص TWI، ارتفاع، فاصله از جاده، SAR
	ReliefFAttributeEval	شاخص TWI، ارتفاع، فاصله از جاده، فاصله از رودخانه
	PrincipalComponents	میزان انحناء، شیب طولی، شیب عرضی، شاخص SPI

۴-۲- الگوریتم Naive Bayes

نتایج حاصل از این الگوریتم در جدول ۲ آمده است. با توجه به جدول از حالت 10 Fold برای طبقه‌بندی کلیه روش‌ها استفاده شد. همچنین مدل طبقه‌بندی، مجموعه آموزش کل (Full training set) است. همچنین به منظور انتخاب بهترین روش، اطلاعات مربوط به ضریب همبستگی، میانگین خطای مطلق، ریشه‌ی میانگین مربع خطا، خطای مطلق نسبی، ریشه‌ی مربع خطای نسبی، تعداد طبقات پیش‌بینی شده‌ی درست و غلط برای هر یک از روش‌ها تعیین شد. نتایج نشان داد روش Ranker با مدل Gain-Ratio-Attribute-Eval با بیشترین ضریب همبستگی (۶۷٪/۹۱)، بهترین روش‌ها برای پیش‌بینی میزان زمین‌لغزش است؛ بنابراین ارتفاع، شاخص TWI، فاصله از جاده و شیب به عنوان بهترین و مؤثرترین پارامترها برای پیش‌بینی میزان زمین‌لغزش محسوب شد.

جدول ۲: ارزیابی طبقات با استفاده از Naive Bayes

نام روش	نام مدل ارزیابی ویژگی	ضریب همبستگی	میانگین خطای مطلق	ریشه میانگین مربع خطا	خطای مطلق نسبی	ریشه مربع خطای نسبی	تعداد طبقات پیش‌بینی شده	
							درست	غلط
Best-First	CFSSubsetEva	۵۴/۱۶	۰/۲۲	۰/۳۵	۷۰/۹۰	۸۷/۷۴	۲۶۰	۲۲۰
Greedy-Stepwise	CFS-Subset-Eval	۵۴/۱۶	۰/۲۲	۰/۳۵	۷۰/۹۰	۸۷/۷۴	۲۶۰	۲۲۰
Ranker	Info-Gain-Attribute-Eval	۵۲/۰۸	۰/۲۲۰	۰/۳۵۹	۶۸/۸۹	۸۹/۸۳	۲۵۰	۲۳۰
	Gain-Ratio-Attribute-Eval	۶۷/۹۱	۰/۱۷۳	۰/۲۹۱	۵۴/۳۰	۷۲/۹۵	۳۲۶	۱۵۴
	Symmetrizer-Attribute-Eval	۵۴/۷۹	۰/۲۱۱	۰/۳۴۹	۶۶/۲۹	۸۷/۴۶	۲۶۳	۲۱۷
	Relief-FAttribute-Eval	۵۲/۰۸	۰/۲۲۰	۰/۳۵۱	۶۸/۸۹	۸۹/۸۳	۲۵۰	۲۳۰
	Principal-Components	۲۵/۸۳	۰/۳۱	۰/۴۰	۹۷/۰۲	۱۰۰/۲۲	۱۲۴	۳۵۶

۵-۲- الگوریتم J48

این الگوریتم برای تولید درخت تصمیم الگوریتم C4.5 است که به صورت هرس و غیر هرس کار می‌کند. درخت C4.5 سعی می‌کند تا به صورت بازگشتی، مجموعه داده‌ها را با استفاده از اطلاعات نرمال به زیر مجموعه‌هایی تقسیم کند که از ریشه به سمت برگ حرکت می‌کند و در نهایت به کلاس مورد نظر می‌رسد. نتایج حاصل از این الگوریتم در جدول ۳ آمده‌است. در این طبقه‌بندی همانند روش Naive Bayes از حالت 10 Fold و مدل طبقه‌بندی مجموعه آموزش کل (Full training set) استفاده شد. همچنین به منظور انتخاب بهترین روش، اطلاعات مربوط به ضریب همبستگی، میانگین خطای مطلق، ریشه میانگین مربع خطا، خطای مطلق نسبی، ریشه مربع خطای نسبی، تعداد طبقات پیش‌بینی شده‌ی درست و غلط برای هر یک از روش‌ها تعیین شد. نتایج نشان داد که روش Ranker با مدل Symmetrizer-Attribute-Eval با بیشترین ضریب همبستگی (۸۹/۳۷)، به عنوان بهترین روش‌ها برای پیش‌بینی میزان زمین‌لغزش انتخاب شد؛ بنابراین شاخص TWI، ارتفاع، فاصله از جاده و SAR بهترین داده‌ها برای پیش‌بینی میزان زمین‌لغزش است.

جدول ۳: ارزیابی طبقات با استفاده از J48

نام روش	نام مدل ارزیابی ویژگی	ضریب همبستگی	میانگین خطای مطلق	ریشه میانگین مربع خطا	خطای مطلق نسبی	ریشه مربع خطای نسبی	تعداد طبقات پیش‌بینی شده غلط درست
Best-First	-Subset-CFS Eval	۶۲/۲۹	۰/۱۷۲	۰/۳۵۳	۵۳/۹۴	۸۸/۴۰	۲۹۹ ۱۸۱
Greedy- Stepwise	CFS-Subset- Eval	۶۲/۲۹	۰/۱۷۲	۰/۳۵۳	۵۳/۹۴	۸۸/۴۰	۲۹۹ ۱۸۱
Ranker	Info-Gain- Attribute-Eval	۶۰	۰/۱۷	۰/۳۵	۵۳/۴۲	۸۸/۸۰	۲۸۸ ۱۹۲
	Gain-Ratio- Attribute-Eval	۸۵/۲۰	۰/۰۹	۰/۲۱۲	۲۸/۳۳	۵۳/۲۲	۴۰۹ ۷۱
	Symmetrizer- Attribute-Eval	۸۹/۳۷	۰/۰۶۵	۰/۱۸۰	۲۰/۳۶	۴۵/۱۲	۴۲۹ ۵۱
	Relief- FAttribute-Eval	۸۴/۷۹	۰/۰۹۴	۰/۲۱۷	۲۹/۴۹	۵۴/۳۰	۴۰۷ ۷۳
	Principal- Components	۷۲/۹۱	۰/۱۴۱	۰/۲۶	۴۴/۱۵	۶۶/۴۵	۳۵۰ ۱۳۰

۶-۲- الگوریتم LMT^۱

نتایج حاصل از این الگوریتم در جدول ۴ آمده‌است. نتایج نشان داد که روش Ranker با مدل Gain-Ratio-Attribute-Eval با بیشترین ضریب همبستگی (۸۷٪/۵)، به عنوان بهترین روش‌ها برای پیش‌بینی میزان زمین‌لغزش انتخاب شدند. بنابراین باز هم پارامترهای ارتفاع، شاخص TWI، فاصله از جاده و شیب به عنوان بهترین داده‌ها برای پیش‌بینی میزان زمین‌لغزش انتخاب شدند.

جدول ۴: ارزیابی طبقات با استفاده از LMT

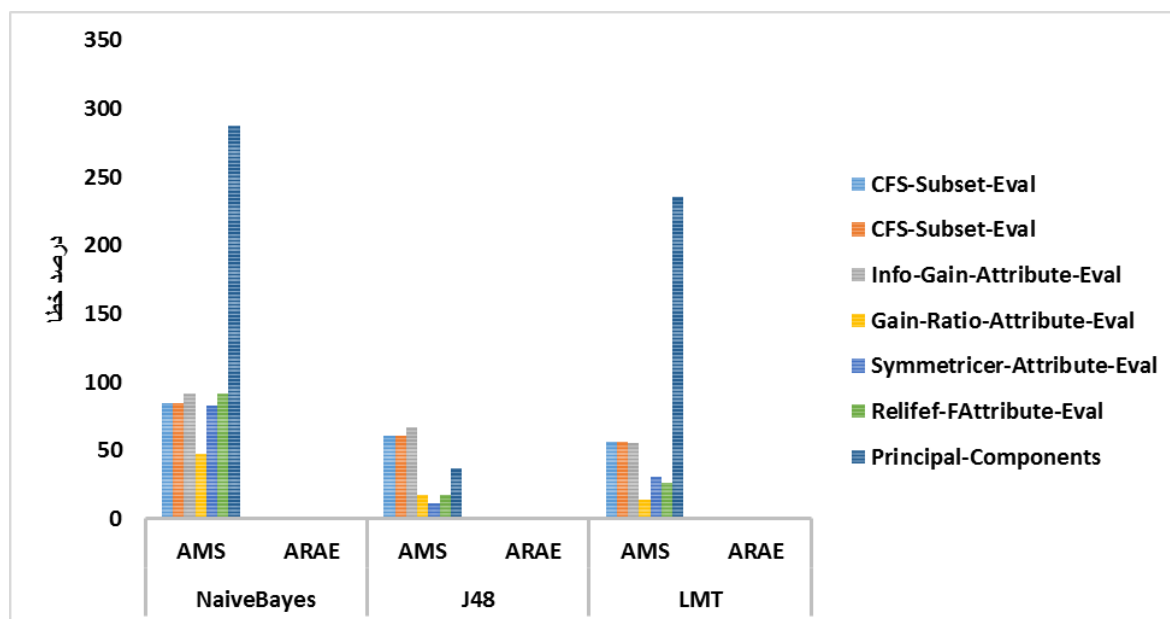
نام روش	نام مدل ارزیابی ویژگی	ضریب همبستگی	میانگین خطای مطلق	ریشه میانگین مربع خطا	خطای مطلق نسبی	ریشه مربع خطای نسبی	تعداد طبقات پیش‌بینی شده غلط درست
Best-First	CFSSubsetEva	۶۳/۹۵	۰/۱۷۳	۰/۳۳	۵۴/۲۰	۸۲/۵۶	۳۰۷ ۱۷۳
Greedy- Stepwise	CFS-Subset- Eval	۶۳/۹۵	۰/۱۷۳	۰/۳۳	۵۴/۲۰	۸۲/۵۶	۳۰۷ ۱۷۳
Ranker	Info-Gain- Attribute-Eval	۶۴/۳۷	۰/۱۶۶	۰/۳۳۳	۵۲/۰۰	۸۳/۳۹	۳۰۹ ۱۷۱
	Gain-Ratio- Attribute-Eval	۸۷/۵	۰/۰۷۳	۰/۱۸۹	۲۲/۹۲	۴۷/۴۰	۴۲۰ ۶۰
	Symmetrizer- Attribute-Eval	۷۶/۴۵	۰/۱۳۲	۰/۲۵۴	۴۱/۳۸	۶۳/۷۷	۳۶۷ ۱۱۳
	Relief- FAttribute- Eval	۷۸/۹۵	۰/۱۲۷	۰/۲۴۸	۳۹/۸۸	۶۸/۲۰	۳۷۹ ۱۰۱
	Principal- Components	۲۹/۷۹	۰/۳۱	۰/۳۹	۹۷/۴۶	۹۸/۶۴	۱۴۳ ۳۳۷

^۱ logistic model tree

در نهایت به منظور اطمینان از انتخاب بهترین روش برای انتخاب مؤثرترین داده‌ها در پیش‌بینی زمین‌لغزش، از دو روش AMS و ARAE برای تعیین میزان خطای هر یک از روش‌ها استفاده شد که نتایج آن در جدول ۵ آمده‌است. با توجه به شکل ۷ و جدول ۵ مشخص می‌شود که بهترین روش برای انتخاب مؤثرترین داده‌ها در پیش‌بینی زمین‌لغزش روش Ranker با ارزیاب Gain-Ratio-Attribute-Eval با کمترین میزان خطاست.

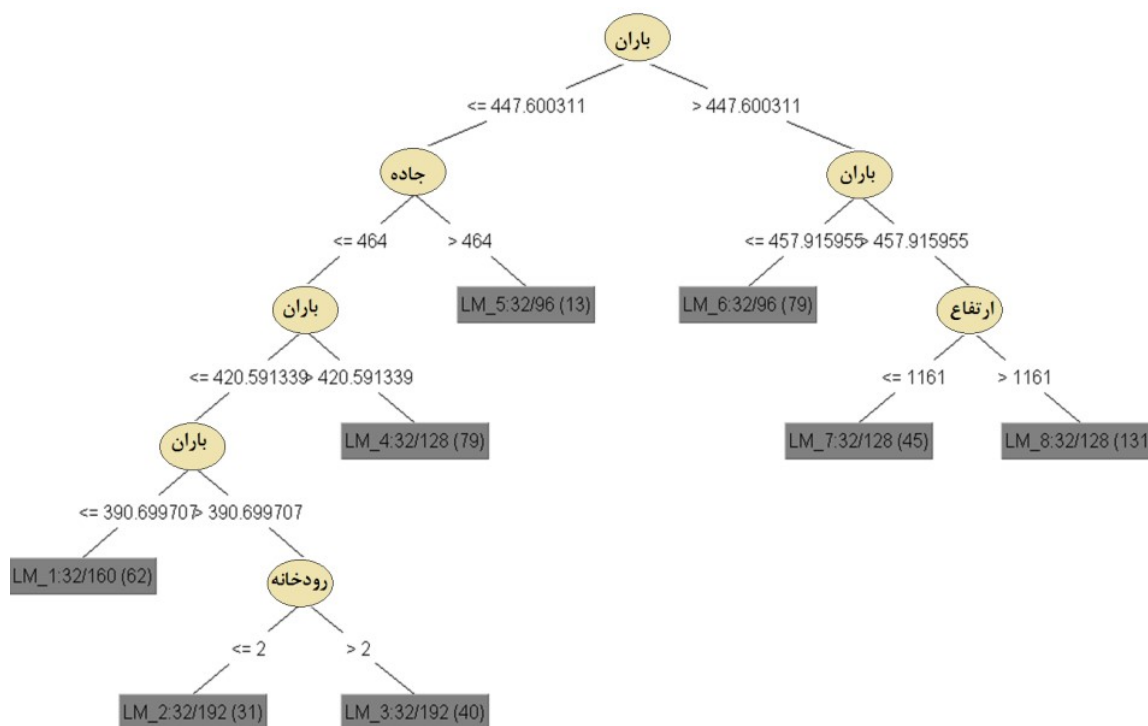
جدول ۵: دقت کلاس‌های مختلف Feature selection

روش ارزیابی						نام مدل ارزیابی ویژگی	روش Method		
Elevation method						Model			
Trees				Bayes		CFSSubsetEva			
Naive Bayes		J48	LMT		CFS-Subset-Eval				
AMS	ARAE	AMS	ARAE	AMS				ARAE	
84.62	0.046	60.54	0.036	56.35			0.036	Info-Gain-Attribute-Eval	
84.62	0.046	60.54	0.036	56.35	0.036	Gain-Ratio-Attribute-Eval			
92.00	0.046	66.67	0.035	55.34	0.035		Symmetrizer-Attribute-Eval		
47.24	0.036	17.36	0.019	14.29	0.015	Relief-FAttribute-Eval			
82.51	0.044	11.89	0.014	30.79	0.028				Principal-Components
92.00	0.046	17.94	0.020	26.65	0.026				
287.10	0.065	37.14	0.029	235.66	0.065				



شکل ۷. میزان خطای هر یک از روش‌ها

با توجه به شکل ۸ مشخص می‌شود که روش طبقه‌بندی LMT با کمترین خطا بهترین روش برای پیش‌بینی میزان زمین‌لغزش در منطقه‌ی مورد مطالعه است. ساختار درختی مربوط به بهترین مدل (LMT) در شکل ۸ نشان داده شده است:



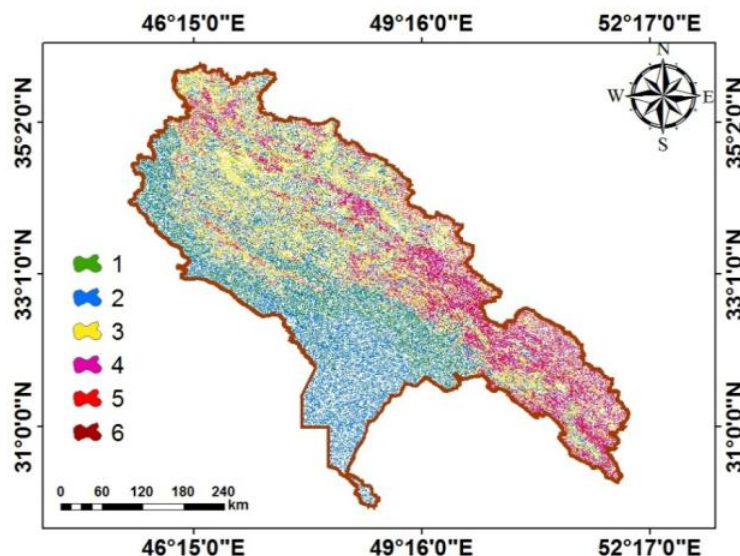
شکل ۸: نمودار شاخه‌ی درختی بهترین مدل به منظور پیش‌بینی مؤثرترین داده‌ها در تعیین زمین‌لغزش

در نهایت از مؤثرترین داده‌های انتخاب شده توسط روش Ranker با مدل GainRatioAttributeEval (ارتفاع، شاخص TWI، فاصله از جاده و شیب برای تعیین میزان حساسیت هر یک از طبقات استفاده شد که نتایج آن در جدول شماره‌ی ۶ ارائه شده است:

جدول ۶: کلاس‌های مختلف حساسیت به زمین‌لغزش مؤثرترین داده‌ها به منظور پیش‌بینی زمین‌لغزش

پارامتر	کلاس ۱ (حساسیت زیاد)	کلاس ۲ (حساسیت کم)	کلاس ۳ (بدون حساسیت)
شیب (درجه)	بیش از ۳۰	بین ۱۵ تا ۳۰	کمتر از ۱۵
ارتفاع (متر)	بیش از ۲۰۰۰	بین ۲۰۰۰ تا ۳۰۰۰	کمتر از ۳۰۰۰
شاخص TWI	بیش از ۱۲	بین ۷ تا ۱۲	کمتر از ۱۲
فاصله از جاده (متر)	بیش از ۴۰۰	بین ۲۰۰ تا ۴۰۰	کمتر از ۴۰۰

به منظور تهیه‌ی نقشه‌ی زمین‌لغزش منطقه استفاده شد. با توجه به جدول ۶ ابتدا داده‌ها در سه کلاس طبقه‌بندی شدند و در نهایت با همپوشانی آنها در محیط احتمال، خطر لغزش افزایش یافت و این طبقه‌بندی صورت گرفت (شکل ۹).



شکل ۹: نقشه‌ی زمین لغزش منطقه‌ی مورد مطالعه با استفاده از داده‌های انتخاب شده توسط الگوریتم Feature selection

با توجه به شکل ۹ مشخص می‌شود این منطقه در ۶ کلاس طبقه‌بندی شده‌اند که با افزایش شماره‌ی کلاس میزان حساسیت افزایش می‌یابد؛ به طوری که مناطق واقع در جنوب شرق، غرب و بخش‌هایی از غرب منطقه‌ی مورد مطالعه دارای بیشترین حساسیت به زمین لغزش هستند. متولی و همکاران (۱۳۸۸) نشان دادند که استفاده از تکنیک‌های جدید مانند LMT به تهیه‌ی نقشه‌ی حساسیت زمین لغزش با تعداد داده‌های کمتر منجر می‌شود. بنابراین می‌توان از الگوریتم انتخاب ویژگی به عنوان روشی کاربردی در مطالعات ژئومورفولوژی، زمین‌شناسی و ... استفاده کرد. رسایی و همکاران (۱۳۹۴) بررسی زمین لغزش را با استفاده از مدل رگرسیون چند متغیره در محیط GIS انجام دادند. نتایج نشان داد که با تعیین مهمترین پارامترهای مؤثر در زمین لغزش، می‌توان به راحتی و با سرعت مکان‌های وقوع زمین لغزش را پیدا کرد.

۴- نتیجه‌گیری

شناسایی عوامل مؤثر در وقوع زمین لغزش‌های موجود در یک منطقه و پیش‌بینی خطر آن یکی از ابزارهای مهم برای دستیابی به راه‌های کنترل این پدیده و انتخاب مؤثرترین گزینه می‌باشد. در این مطالعه از الگوریتم انتخاب ویژگی به منظور تعیین مؤثرترین پارامترها در پیش‌بینی زمین لغزش در جنوب غرب ایران استفاده شد. برای این منظور از روش‌های Ranker، Bestfirst و Greedy Stepwise استفاده شد. نتایج نشان داد که داده‌های انتخاب شده توسط روش Ranker با مدل Gain-Ratio-Attribute-Eval با کمترین میزان خطا و حداکثر ضریب همبستگی (۸۷/۵)، با طبقه‌بندی LMT به عنوان مؤثرترین داده‌ها برای تعیین زمین لغزش است. همچنین نتایج حاصل از تهیه‌ی زمین لغزش با استفاده از داده‌های انتخاب شده توسط الگوریتم انتخاب ویژگی، نشان داد که بخش‌هایی از مناطق شمالی و جنوب شرق منطقه‌ی مورد مطالعه در معرض خطر بیشتری از زمین لغزش قرار دارند. همچنین مشخص شد که در مدل Principal Components میزان انحناء، شیب طولی، شیب عرضی و شاخص SPI مهم‌ترین داده‌ها برای تعیین زمین لغزش محسوب می‌شوند. در این مطالعه از یک سو با استفاده از الگوریتم انتخاب ویژگی، مؤثرترین داده‌های زمین لغزش

مشخص شد که به کاهش وقت و هزینه در جمع‌آوری و اندازه‌گیری انجامید و از سوی دیگر، با استفاده از این داده‌ها نقشه‌ی زمین‌لغزش منطقه به صورت مکانی تهیه شد. مشابه این مطالعه توسط گرای و همکاران^۱ (2010)، Grai et al, (2010)، بهشتی‌راد و همکاران (۱۳۸۸) (Beheshti et al, 2010) و رسایی و همکاران (۱۳۹۴) (Rasaei et al, 2015) انجام شد. این پژوهش اهمیت بررسی پارامترهای مؤثر در زمین‌لغزش را نشان می‌دهد که مزیت و نوآوری این پژوهش نسبت به پژوهش‌های قبلی در تأکید بیشتر بر مدل‌های نرم‌افزاری و دقت بیشتر کار همچنین وسعت منطقه‌ی مورد بررسی است.

References

1. Anbalagan, R., 1992. Landslide hazard evaluation and zonation mapping in mountainous terrain, Engineering geology, 1, 32 (4), 269 - 77.
2. Ayalew, L., & H. Yamagishi., 2005. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan, Geomorphology, 1, 65 (1), 15 - 31.
3. Bay, N., & Hajimir rahimi, M., 2009. Landslide hazard zonation using AHP method, geomechanical congress.Tehran. pp 222-229.
4. Beheshti rad, M., Fayz nia, s., Ahmadi, H., 2010. Survey on the Effectiveness of Critical Landslide Bonding Pattern (CF) (Case Study of Moalem Kalayeh Basin), Natural Geography. V 2. No, 5. Pp: 19-28.
5. Behniafar, A., Ghanbarzadeh, H., Mansouri, M., 2010. Landslide Hazard and Landslide Instability by AHP and Probability (Case Study: Kang River Basin, Binalood North Slopes), Geographic Space, V 9. No, 27. Pp 55-78.
6. Clerici, A.; Perego, S.; Tellini, C.; P. Vescovi, 2002. A procedure for landslide susceptibility zonation by the conditional analysis method, Geomorphology, 1, 48(4), 349 - 364.
7. Dash, M., & H. Liu., 2003. Consistency-based search in feature selection, Artificial intelligence, 1, 151 (1-2), 155 - 76.
8. Fayznia, S., kalarostaqi, A., Ahmadi, H., Safaei, M., 2003. Investigation of Factors Affecting Landslide Rise and Risk of Slipping (Case Study: Shirin Roud Watershed - Tajan Dam). Iranian Natural Resources Journal, vol 57, No 1, pp: 20-35.
9. Gorsevski, P. V.; Gessler, P. E.; Boll, J.; Elliot, W. J.; R. B. Foltz, 2006. Spatially and temporally distributed modeling of landslide susceptibility, Geomorphology, 30, 80(3), 178 - 98.
10. Gorsevski, P. V.; Gessler, P.; & R. B. Foltz, 2000. Spatial prediction of landslide hazard using logistic regression and GIS 4th Int, Conference on Integrating GIS and Environmental Modeling, Alberta, Canada. 9 pp.
11. Grai, P.; Karimi, H.; & M. Tavakoli, 2010. Preparation of landslide susceptibility maps using multivariate regression (case study: Ilam dam basin), Journal of Watershed Management, 1, 1-16.
12. Iwahashi, J.; Watanabe, S.; T. Furuya, 2003. Mean slope-angle frequency distribution and size frequency distribution of landslide masses in Higashikubiki area, Japan, Geomorphology, 1, 50(4), 349 - 64.
13. Karami, F., 2013. Land degradation sensitivity evaluation in semi-arid mountain basins using statistical methods and weighted control model (case study: Saeed Abadchay drainage basin, northwest of Iran), Geographical research, V 3, pp:21- 44.

¹ Grai et al

14. Lee, S., & K. Min., 2001. Statistical analysis of landslide susceptibility at Yongin, Korea. *Environmental geology*, 16, 40(9), 1095 - 113.
15. Manconi, A., & D. Giordan., 2015. Landslide early warning based on failure forecast models: the example of the Mt, de La Saxe rockslide, northern Italy, *Natural Hazards and Earth System Sciences*, 29, 15(7), 1639 - 44.
16. Melchiorre, C.; Matteucci, M.; Azzoni, A.; & A. Zanchi, 2008. Artificial neural networks and cluster analysis in landslide susceptibility zonation, *Geomorphology*, 15, 94(3), 379 - 400.
17. Moghimi, E., Yamani, M., Rahimi, S., 2013. Evaluation and zoning of landslide hazard in Roodbar city using the network analysis process, *Quantitative geomorphology researches*, Vol 4, pp:118 - 103.
18. Mostafai, H.; Onagh, M.; Mesdaghi, M.; & M. Shariat Jafari, 2009. Comparison between efficiency of empirical and statistical model to landslide hazard zonation (Almootrood watershed), *Soil and Water Conservation Research Journal*, 16, 1 - 16.
19. Motavalli, S., Esmaeili, R., Hosseinzadeh, M., 2010. Determination of sensitivity of landslide occurrence using logistic regression in Vaz catchment area (Mazandaran province), *Natural geography*, Vol, 2, No, 5, pp: 73-84.
20. Naseriparsa, M.; Bidgoli, A. M.; & T. 2014. Varace, A hybrid feature selection method to improve performance of a group of classification algorithms. *arXiv preprint arXiv:1403.2372*.
21. Rajab zadeh, H., 2014. Assessment and zoning of the risk of landslide in the catchment area of Godarchay Naghda using GIS, master's thesis of Islamic Azad University, Ahar Branch. Pp: 1-116.
22. Rajabi, M., & Fayzollah poor., 2014. Landslide zonation of Givi-Chai river basin using multi-layered perceptron model of BP, *geography and development*, V, 36, pp: 161-180.
23. Rasaei, A., Khosravi, Kh., Haib nejad, M., Haydari, A., Mashayekhan, A., 2015. Landslide hazard zonation with multivariate regression model in GIS environment (case study: Aq mashhad Basin, Mazandaran province), *Watershed management research journal*, V, 6, No, 12, pp: 215-205.
24. Roosatei, Sh., & Alizadeh, V., 2013. Landslide hazard zonation in the Sufi Chay basin (Maragheh) using the Anaballah method, *geographical space*, V, 12, No, 39, pp: 35-17.
25. Roostaei, Sh., & Ahmadzadeh, H., 2013. Zoning of areas affected by landslide hazard on the Tabriz-Marand road using remote sensing and GIS, *quantitative geomorphology Research*. V, 1. No, 1. Pp: 47-58.
26. Wan, S., 2009. A spatial decision support system for extracting the core factors and thresholds for landslide susceptibility map, *Engineering Geology*, 108(3), 237 - 51.
27. Westen, C. V., & M. J. Terlien., 1996. An approach towards deterministic landslide hazard analysis in GIS. A case study from Manizales (Colombia), *Earth surface processes and landforms*, 21(9), 853 - 68.
28. Zhang, W.; Wang, W.; & Q. Xia, 2012. Landslide Risk Zoning Based on Contribution Rate Weight Stack Method, *Energy Procedia*, 16, 178 - 83.
29. Zhou, S., & L. Fang., 2015. Support vector machine modeling of earthquake-induced landslides susceptibility in central part of Sichuan province, China, *Geoenvironmental Disasters*, 6, 2(1) , 2.

Applying Feature-Selection Algorithm to Predict Landslide in the Southwest of Iran

Marziyeh Mokarram¹: *Assistant professor of Shiraz University*

Saeed Negahban: *Assistant professor of Shiraz University*

Article History (Received: 2016/11/24 Accepted: 2017/05/15)

Extended abstract

1- INTRODUCTION

Nowadays people have an increased sensitivity towards landslides especially in mountainous areas using change in the land use and the expansion of communication networks (Gvrsysky et al., 2006). In the twentieth century, Asia has allocated the highest incident of landslides (220 landslides). Latin America has had the highest number of casualties (more than 2,500) and Europe has experienced the highest loss (Karami, 2012). Landslide is one of the significant phenomena in the environment, watershed management, and natural resources. The importance of landslides can be discussed and analyzed from various perspectives. The most important reason refers to human and financial loss (Rajab Zadeh, 2013). Research on the dynamic relationships between factors in landslides has a high role in the investigation of the respective risk. In fact, much research has been conducted in the realm of determining the relationship between environmental factors and the occurrence of landslide (Anbalagan, 1992, Liu Min, 2001, Ayvashy et al., 2003, Yalv and Yamagyshy, 2005). Some of such research is consistent with the relationship between the distribution of geological and geomorphological factors and landslides observed. However, to analyze the results and predict the likelihood of landslides, there are common tools that are used in statistical calculations.

2- THEORETICAL FRAMEWORK

A landslide, also known as a landslip, is a form of mass wasting that includes a wide range of ground movements, such as rock falls, deep failure of slopes, and shallow debris flows. Slope, fine sediments and ground moisture have important roles in the occurrence of landslides; given that many parameters affect the landslide, a more effective choice to reduce time and costs is important. The subject of feature selection is the one of the issues identified in the machine learning and statistics. The problem in many applications (such as classification) is very important. Because in these applications, there are a large number of features that many of them are unused. In fact, if they are not removed, these features will not create problems, but save a lot of useless and useful information together.

3- METHODOLOGY

This study was carried out in the southwest of Iran (a part of Khuzestan, Khorramabad, Ilam, Kermanshah and Hamedan). It includes an area of about 154272.48 km² and is located at

¹ Corresponding Author: M. Mokarram@shirazu.ac.ir

the longitude of N 29° 56' to 35° 46' and the latitude of E 45° 24' to 52° 1'. The altitude of the study area ranges from the lowest 30 m to the highest 4,415 m.

There are different methods that try to find better subsets among the 2^T subsets. In all of these methods, the selection of the subset is based on the type of application and type of definition that can optimize the value of an evaluation function. In fact, each way tries to make the best attributes of choices, but according to the extent of answers and increasing the answer with T, finding the optimal solution and T medium is costly. Feature selection process has four steps:

Generation function: this function find sub-candidates for the procedure.

Elevation function: it is based on data subset to be evaluated and a number as the method returns. Different methods try to find a subset that optimizes the amount.

Stopping criterion: it is used to decide when to stop the algorithm.

4- RESULTS

The results show that the case study is located in 6 classes increasing sensitivity as the number of class increases. So that the areas located in the South, East, and parts of the West regions are most sensitive to landslide. Motevali et al., (2008) show that using a new method such as LMT can prepare landslide map with low data. So, in the research of geomorphology and geology, feature selection can be used. Rasaie et al (2015) used regression in GIS software to prepare landslide map. The results showed that using effective parameters of landslide can find landslide map easily and quirkily.

5- CONCLUSIONS AND SUGGESTIONS

The results of feature selection method show that the Ranker method with Gain-Ratio-Attribute-Eval, with low error, with highly significant correlation (87.5), and with LMT classification is the best method for the selection of the most effective data to determine landslide. Also, the results indicated that some parts of North and South-East of the study area are located at greater risk of landslide. Also, principal components showed that curvature, profile, plan and SPI were the most important data for determining landslide. In the study, it was attempted to use low data selected by feature selection, and save time and money via determining important data for landslide. Using the data, landslide map was prepared spatially.

Key Words: Feature Selection Algorithm, Landslide, Southwest of Iran.