

ارزیابی حساسیت زمین لغزش با استفاده از مدل‌های داده‌کاوی، مطالعه موردی: حوزه آبخیز چهل‌چای

آیدینگ کرنژادی^۱ و حمیدرضا پورقاسمی^{۲*}

^۱ دانشجوی دکتری علوم و مهندسی آبخیزداری، دانشکده مرتع و آبخیزداری، دانشگاه علوم کشاورزی و منابع طبیعی گرگان و ^۲ استادیار، دانشکده کشاورزی، دانشگاه شیراز

تاریخ پذیرش: ۱۳۹۵/۰۹/۳۰

تاریخ دریافت: ۱۳۹۵/۰۶/۲۵

چکیده

هدف از تحقیق پیش‌رو، پهنه‌بندی حساسیت زمین لغزش در حوزه آبخیز چهل‌چای، واقع در استان گلستان می‌باشد. بدین منظور، از دو مدل داده‌کاوی ماشین بردار پشتیبان (SVM) و درخت رگرسیون تقویت شده (BRT) به لحاظ الگوریتم محاسباتی توانمند در زمینه ارزیابی فرایند زمین لغزش استفاده شد. شواهد زمین لغزش با استفاده از دستگاه موقعیت‌یاب جهانی (GPS)، بازدیدهای میدانی و اطلاعات سازمانی موجود ثبت و در سامانه اطلاعات جغرافیایی (GIS) در قالب نقشه تهیه شد. همچنین، با توجه به مرور منابع گسترده، ۱۲ عامل زمینه‌ساز شامل فاصله از گسل، فاصله از آبراهه، فاصله از جاده، واحدهای سنگ‌شناسی، بافت خاک، کاربری اراضی، درجه شیب، جهت شیب، طبقات ارتفاعی، انحنای طولی و عرضی دامنه و شاخص رطوبت (WI) به‌عنوان عوامل موثر بر وقوع زمین لغزش منطقه مورد مطالعه انتخاب و لایه‌های مذکور در سامانه اطلاعات جغرافیایی تهیه شدند. به‌منظور ارزیابی نتایج مدل‌ها، از مقدار مساحت زیر منحنی تشخیص عملکرد نسبی (ROC) و ۳۰ درصد لغزش‌های استفاده نشده در فرایند مدل‌سازی استفاده شد. نتایج نشان داد که مدل ماشین بردار پشتیبان (SVM) با مساحت زیر منحنی ۰/۸۲ (۸۲ درصد) کارایی بهتری در پهنه‌بندی حساسیت زمین لغزش در حوضه مورد مطالعه داشته و مدل BRT با مقدار معادل ۰/۷۷ (۷۷ درصد) در رتبه بعدی اهمیت قرار گرفت. بر اساس نتایج مدل SVM، حدود ۴۵ درصد از حوزه آبخیز چهل‌چای در پهنه حساسیت زیاد و خیلی زیاد به وقوع زمین لغزش قرار گرفته است.

واژه‌های کلیدی: استان گلستان، درخت رگرسیون تقویت‌شده، سامانه اطلاعات جغرافیایی، ماشین بردار پشتیبان

مقدمه

زمین‌لرزه، سیل و آتشفشان از بلایای طبیعی مهم به شمار می‌رود که هر ساله رخداد آن در مناطق مستعد این پدیده، خسارات و صدمات قابل ملاحظه‌ای به بار می‌آورد. با افزایش جمعیت و اسکان در مناطق مستعد رویداد زمین لغزش، آمارهای جهانی تلفات ناشی از این پدیده روند صعودی دارد. بر اساس تحقیقات

زمین لغزش به حرکت مواد تشکیل دهنده دامنه، شامل صخره‌های طبیعی، خاک، انباشته‌های مصنوعی، قطعات یخی و یا مخلوطی از آن‌ها که به‌وسیله نیروی ثقل به سمت پایین جابه‌جا می‌شوند، اطلاق می‌شود (Varnes, ۱۹۷۶). پدیده زمین لغزش همانند

* مسئول مکاتبات: hr.pourghasemi@shirazu.ac.ir

روش‌های عینی سوق می‌دهند، نظیر مدل‌های عامل اسمی ریسک زمین‌لغزش، ارزش اطلاعاتی، تراکم سطح، وزن واقعه، نسبت فراوانی، آنتروپی شانون، تکنیک‌های فازی و شبکه عصبی مصنوعی، (۳) روش-های ابداعی و مبتنی بر نظر کارشناسی نظیر روش اداره استاندارد هند برای ارزیابی خطر زمین‌لغزش (BIS-LHEF) و (۴) روش‌های قطعی، نظیر عامل ایمنی دامنه که مستقیماً در ارتباط با پایداری دامنه بوده و در مقیاس حوزه آبخیز انجام می‌شوند.

مدل‌های ماشین یادگیری به لحاظ به‌کارگیری الگوریتم‌های تخمین توزیع، طبیعت داده محور و تکرار بالای فرایند مدل‌سازی، توانایی بالایی در شناسایی رفتار وقوع پدیده زمین‌لغزش داشته و در تحقیقات متعددی برتری نسبی خود را نسبت به مدل‌های آماری دو متغیره و چند متغیره اثبات نموده‌اند (Kerle و Stumpf، ۲۰۱۱؛ Vorpahl و همکاران، ۲۰۱۲؛ Lee و همکاران، ۲۰۱۳؛ Trigila و همکاران، ۲۰۱۳). Yamani و همکاران (۲۰۱۲) با به‌کارگیری توابع حلقوی، چند جمله‌ای، شعاعی و خطی در مدل SVM و انتخاب شش عامل زمینه‌ساز، به ارزیابی حساسیت زمین‌لغزش در حوزه آبخیز درکه در شمال شهر تهران پرداختند که در نهایت تابع حلقوی به-لحاظ انطباق بیشتر با واقعیت منطقه به‌عنوان تابع برتر انتخاب شد. Pourghasemi و همکاران (۲۰۱۳) با استفاده از شش تابع خطی، چندجمله‌ای درجه ۲، درجه ۳، درجه ۴، شعاعی (RBF) و حلقوی و انتخاب ۱۴ عامل کنترلی، اقدام به تهیه نقشه حساسیت زمین‌لغزش در سطح استان گلستان نمودند که بر اساس نتایج شاخص نرخ پیش‌بینی^۳، تابع شعاعی به-عنوان بهترین الگوریتم ارزیابی حساسیت زمین‌لغزش در استان گلستان معرفی شد. Ahmadabadi و Rahmati (۲۰۱۵) با استفاده از توابع خطی و چند جمله‌ای در مدل SVM و انتخاب نه عامل کلیدی، حساسیت زمین‌لغزش را در آزادراه خرم‌آباد-پل‌زال پهنه‌بندی نمودند که مبتنی بر یافته‌های مدل، تابع چند جمله‌ای بهترین عملکرد را در مدل‌سازی فرایند

Koehorst و همکاران (۲۰۰۵) مبتنی بر تحلیل داده‌های ۱۰۰ ساله خطرات طبیعی، پس از سیل و زلزله، زمین‌لغزش پر رخدادترین بلای طبیعی در جهان محسوب می‌شود. کشور ایران نیز با توجه به ویژگی‌های زمین‌شناختی، ژئومورفولوژیکی و اقلیمی، همواره تحت تأثیر این پدیده مخرب قرار گرفته و موارد متعددی از خسارت‌های ناشی از این پدیده در کشور گزارش شده است. به‌طوری که از میان ۴۳ تیپ رایج بلایای طبیعی رخ داده در بخش‌های مختلف جهان، حدود ۳۸ تیپ در ایران مشاهده شده، که زمین‌لرزه‌های بزرگ، سیل، خشکسالی، زمین‌لغزش، بیابان‌زایی، جنگل‌زدایی و طوفان از انواع شایع آن به‌شمار رفته و سبب شده تا کشور ایران بر اساس تخمین‌های بین‌المللی در بین ۱۰ کشور بلاخیز جهان قرار گیرد (Ownegh، ۲۰۰۲؛ UNODRR، ۲۰۰۵). استان گلستان به لحاظ آب و هوای معتدل و فراهم بودن شرایط زمین‌شناختی، یکی از مستعدترین استان‌های ایران جهت وقوع زمین‌لغزش به‌شمار می-رود.

Meten و همکاران (۲۰۱۵) اظهار دارند که نقشه‌های حساسیت زمین‌لغزش به مثابه سنگ بنایی برای تصمیم‌گیران برای پیشنهاد اقدامات فنی، نظارتی و یا ترکیبی از این دو می‌باشند. این نقشه‌ها همچنین، ابزار مهمی برای مهندسان، دانشمندان علوم زمین، برنامه‌ریزان و تصمیم‌گیران به‌منظور انتخاب مکان‌های مناسب برای کشاورزی، ساخت و ساز و سایر فعالیت‌های توسعه‌ای محسوب می‌شوند (Ercanoglu و Gokceoglu، ۲۰۰۲). بر اساس طبقه‌بندی Van Westen (۲۰۰۶) روش‌های بررسی زمین‌لغزش‌ها به چهار دسته کلی تقسیم می‌شوند. (۱) روش‌های مبتنی بر لیست‌برداری زمین‌لغزش‌ها که در آن اطلاعات مکانی و زمانی وقوع، تیپ و نرخ زمین‌لغزش‌ها از طریق مطالعات گسترده میدانی و تفسیر تصاویر ماهواره‌ای و عکس‌های هوایی ثبت می‌شوند و زمینه را برای سایر روش‌های پهنه‌بندی حساسیت زمین‌لغزش فراهم می‌نمایند، (۲) روش‌های آماری-احتمالاتی داده محور که وزن‌دهی ذهنی به عوامل را به سمت

² Bureau of Indian Standards' Landslide Hazard Evaluation Factor

³ Prediction rate

¹ United Nations Office for Disaster Risk Reduction

مواد و روش‌ها

در این تحقیق با استفاده از دو مدل SVM با تابع پایه شعاعی و BRT و انتخاب ۱۲ عامل زمینه‌ساز، حساسیت زمین‌لغزش در حوضه بحرانی چهل‌چای ارزیابی و در نهایت بر اساس شاخص AUC مدل برتر معرفی شد. نمودار جریان مراحل انجام تحقیق در شکل ۲ ارائه شده است.

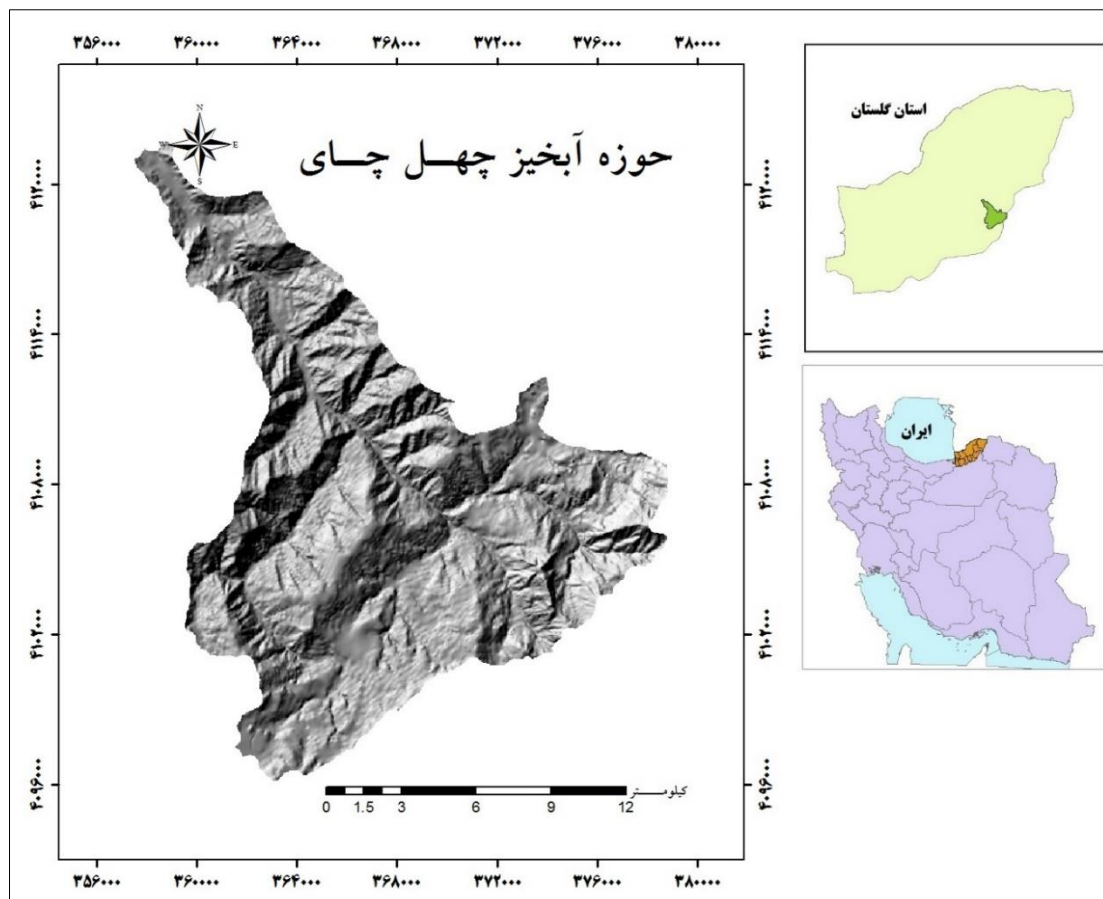
منطقه مورد پژوهش: حوزه آبخیز چهل‌چای از حوضه‌های کوهستانی کشور با وسعت حدود ۲۵ هزار هکتار است که در بین ۲۳° ۵۵' تا ۳۸° ۵۵' طول شرقی و ۵۹° ۳۶' تا ۱۳° ۳۷' عرض شمالی واقع شده است (شکل ۱). این حوضه از لحاظ تقسیمات سیاسی در محدوده شهرستان مینودشت قرار گرفته و یکی از زیرحوضه‌های بزرگ گرگان‌رود است. آبخیز مذکور ۱۳۵ کیلومتر از مرکز استان گلستان یعنی شهر گرگان فاصله دارد و از شمال به جاده ترانزیتی تهران-مشهد و آبخیز اوغان، از جنوب به آبخیز کاشیدار-تیل‌آباد، از غرب به آبخیز نرماب و از شرق به روستای دوزین محدود می‌شود. حداقل و حداکثر ارتفاع منطقه مورد مطالعه به ترتیب ۱۹۰ و ۲۵۲۷ متر از سطح دریا است. آبراهه‌های اصلی حوزه آبخیز چهل‌چای با جهت کلی جنوب به شمال نزولات جوی را جمع‌آوری می‌نمایند. نمودار جریانی مراحل انجام تحقیق در شکل ۲ ارائه شده است.

تهیه نقشه پراکنش زمین‌لغزش‌ها: یکی از مهمترین مراحل ارزیابی حساسیت زمین‌لغزش، شناسایی و تهیه نقشه پراکنش زمین‌لغزش‌های موجود در حوضه مورد مطالعه می‌باشد. در این تحقیق، ثبت و بررسی زمین‌لغزش‌های موجود در حوضه از طریق بازدیدهای میدانی، اطلاعات محلی و نقشه‌های موجود در اداره منابع طبیعی و آبخیزداری استان گلستان (حاصل از تفسیر عکس‌های هوایی ۱:۲۰۰۰۰ و ۱:۵۰۰۰۰) به دست آمده و در نهایت بر اساس روش انتخاب تصادفی به دو دسته داده جهت مدل‌سازی (۷۰ درصد نقاط لغزشی) و ارزیابی نتایج مدل‌ها (۳۰ درصد نقاط لغزشی) تقسیم شد (Pourghasemi و همکاران، ۲۰۱۲a؛ Pourghasemi و همکاران، ۲۰۱۲b).

زمین‌لغزش داشته که نشان از رفتار غیرخطی این پدیده در حوضه مطالعاتی دارد. Naghibi و Pourghasemi (۲۰۱۵) با استفاده از سه مدل ماشین یادگیری شامل درخت رگرسیون تقویت شده BRT، درخت رگرسیون و طبقه‌بندی CART، جنگل تصادفی (RF) و دو مدل آماری دو متغیره شامل تابع شواهد قطعی EBF و چند متغیره مدل خطی تعمیم-یافته GLM انتخاب ۱۴ عامل موثر، پتانسیل آب زیرزمینی را در حوضه بهشت‌آباد در استان چهارمحال بختیاری پهنه‌بندی نمودند که در نهایت بر اساس نتایج ارزیابی مدل‌ها مبتنی بر شاخص‌های نرخ پیش-بینی و نرخ موفقیت^۱، به ترتیب مدل‌های CART، BRT و RF به عنوان مدل‌های برتر در منطقه معرفی شدند. Naghibi و همکاران (۲۰۱۶) در مطالعه‌ای دیگر در زمینه پتانسیل آب زیرزمینی، با انتخاب ۱۳ عامل به ارزیابی کارایی سه مدل BRT، CART و RF در آبخیز کوه‌رنگ در غرب چهارمحال بختیاری پرداختند که طبق نتایج، مدل BRT کارایی بهتری نسبت به دو مدل دیگر داشته است. Pourtaghi و همکاران (۲۰۱۶) با استفاده از سه مدل BRT، GAM و RF و انتخاب ۱۵ عامل موثر، به پهنه‌بندی حساسیت آتش‌سوزی جنگل در شرق استان گلستان پرداختند که بر اساس نتایج حاصل از مساحت زیر منحنی (AUC) مشخصه عملکرد نسبی (ROC)، مدل BRT به عنوان مدل برتر در منطقه مطالعاتی معرفی شده است.

حوضه چهل‌چای به دلیل دارا بودن مقدار و شدت بالای بارندگی و حضور سازندهای مستعد به لغزش، به‌طور بالقوه مستعد وقوع زمین‌لغزش می‌باشد. علاوه بر این، این حرکات در مناطق پر شیب حوضه، تحت تاثیر عوامل انسانی از قبیل جاده‌سازی غیر اصولی و تغییر کاربری، تشدید شده‌اند (Ministry of Jihad-e-Agriculture، ۲۰۰۴). لذا، در این تحقیق، با استفاده از دو مدل SVM (تابع شعاعی) و BRT و انتخاب ۱۲ عامل زمینه‌ساز، حساسیت زمین‌لغزش در حوضه بحرانی چهل‌چای ارزیابی و در نهایت بر اساس شاخص AUC، مدل برتر معرفی شد.

¹ Success rate



شکل ۱- موقعیت حوزه آبخیز چهل چای در ایران و استان گلستان

توپوگرافی و عوامل درجه شیب، جهت شیب، انحنای طولی و عرضی دامنه و شاخص TWI از روی نقشه DEM منطقه استخراج و تهیه شد. نقشه‌های انحنای نشان‌دهنده میزان انحراف سطح از صاف بودن و یا به عبارت دیگر محدب و مقعر بودن دامنه را نشان می‌دهد.

مقادیر مثبت و منفی در انحنای عرضی دامنه^۳ به ترتیب بیانگر تحدب (واگرایی جریان) و تقعر (همگرایی جریان) بوده و مقادیر مثبت و منفی در انحنای طولی دامنه^۴ به ترتیب بیانگر تقعر (کاهش سرعت جریان) و تحدب دامنه (افزایش سرعت جریان) می‌باشند (Jenness ۲۰۱۳). جهت تهیه این دو عامل (انحنای طولی و عرضی) از افزونه DEM Surface Tools در محیط ArcGIS 10.2 و بر اساس روش Thorne و Zevenbergen (۱۹۸۷) استفاده شد. شاخص TWI نیز از رابطه (۱) محاسبه شد (Beven و

تهیه نقشه عوامل کنترل کننده و موثر بر

وقوع زمین لغزش: بر اساس مرور منابع گسترده و

قابلیت دسترسی به داده‌های منطقه مورد مطالعه، ۱۲ عامل زمینه‌ساز شامل فاصله از گسل، فاصله از آبراهه، فاصله از جاده، واحدهای سنگ‌شناسی، بافت خاک، کاربری اراضی، درجه شیب، جهت شیب، ارتفاع از سطح دریا، انحنای طولی و عرضی دامنه و شاخص رطوبت توپوگرافی (TWI^۱) بر اساس رابطه (۱) جهت مدل‌سازی حساسیت زمین لغزش در حوزه آبخیز چهل چای انتخاب و نقشه‌های آن در سامانه اطلاعات جغرافیایی (ArcGIS 10.2) تهیه شد که در ادامه توضیحات ارائه شده است. نقشه مدل رقومی ارتفاع (DEM^۲) با قدرت تفکیک مکانی ۱۰ متری، از روی نقشه توپوگرافی ۱:۲۵۰۰۰ (سازمان نقشه‌برداری کشور) تهیه شد. نقشه‌های آبراهه، جاده از روی نقشه

^۳ Plan curvature

^۴ Profile curvature

^۱ Topographic Wetness Index

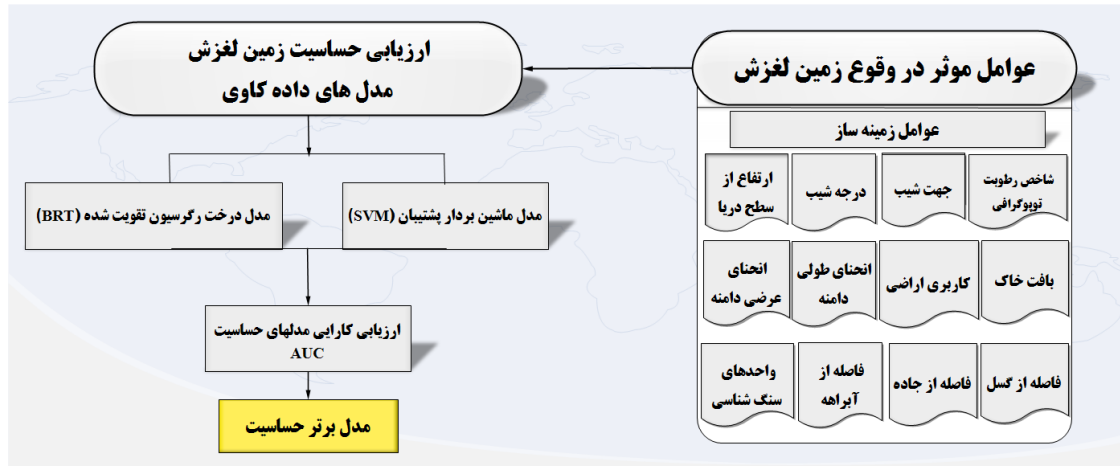
^۲ Digital Elevation Model

(درجه) می‌باشد. مقادیر بالا در این شاخص بیان‌گر مناطقی است که مشارکت بیشتری در هدایت رواناب به خروجی حوضه دارند (Beven و Kirkby، ۱۹۷۹).

(Kirkby، ۱۹۷۹).

$$TWI = \alpha / \tan \beta \quad , \quad \alpha = A / L \quad (1)$$

که در آن، A کل مساحت بالادست، L طول موثر خط تراز (عمود بر جهت جریان) و β زاویه شیب



شکل ۲- نمودار جریان‌ی ارزیابی و پهنه‌بندی حساسیت زمین‌لغزش در حوزه آبخیز چهل‌چای

به‌طور مشترک برای دو عملیات طبقه‌بندی و رگرسیونی به‌کار رود (Cristianini و همکاران، ۲۰۰۰). این روش شامل مجموعه‌ای از توابع طبقه‌بندی است که دارای قابلیت ارزیابی خطا و تعمیم مناسب اطلاعات بوده و با استفاده از اطلاعات موجود در لایه-های عوامل موثر و تکرار بالای مدل‌سازی، پیچیدگی رفتار وقوع پدیده زمین‌لغزش را کاهش می‌دهد. این مدل شامل یک مرحله آموزشی همراه با ورودی و ارزش‌های اهداف خروجی است (Ahmadabadi و Rahmati، ۲۰۱۵). بر اساس نظریه یادگیری آماری، می‌توان کران نرخ خطای ماشین یادگیری را برای داده‌های طبقه‌بندی نشده، به‌عنوان نرخ خطای تعمیم یافته، در نظر گرفت که این کران‌ها به‌عنوان تابعی از مجموعه نرخ خطاهای آموزشی هستند که میزان پیچیدگی طبقه‌بندی کننده‌ها^۱ را نشان می‌دهند (Yamani و همکاران، ۲۰۱۲). لذا، جهت کمینه کردن نرخ خطاهای تعمیم یافته، باید نرخ خطای آموزش و پیچیدگی طبقه‌بندی کاهش یابد که با استفاده از صفحه جدایش^۱ انجام می‌شود که وظیفه افزایش حاشیه بین طبقات را بر عهده دارد (Vapnik، ۲۰۱۳). داده‌های نقطه‌ای نزدیک به این صفحه، بردارهای

نقشه کاربری اراضی با تفسیر تصاویر ماهواره‌ای IRS-P6 (سنجنده LISS-III) با توان تفکیک ۲۳/۵ متر و IRS-P5 (سنجنده Pan) با توان تفکیک ۲/۵ متر سال ۲۰۰۶ با الگوریتم Ehler تهیه شد و با استفاده از تصاویر گوگل‌ارث و بازدیدهای میدانی اصلاح شد (Rahimi و همکاران، ۲۰۱۰). نقشه سنگ‌شناسی و گسل‌های منطقه از نقشه ۱:۱۰۰۰۰۰ سازمان زمین‌شناسی و اکتشافات معدنی کشور تهیه و استخراج شد. نقشه بافت خاک منطقه نیز در مقیاس ۱:۱۰۰۰۰۰ و از موسسه تحقیقات خاک و آب کشور تهیه شد. لایه-های کمی (نظیر درجه شیب و ارتفاع) به‌شکل خام و طبقه‌بندی نشده (مورد نیاز برای مدل‌های ماشینی) و لایه‌های کیفی (لایه‌های کاربری اراضی، سنگ‌شناسی، جهت شیب، بافت خاک) به‌شکل طبقه‌بندی شده وارد مدل‌های ماشینی شدند.

مدل ماشین بردار پشتیبان (SVM): مدل

SVM، یک سیستم طبقه‌بندی نظارت شده به‌شمار می‌رود که بر پایه تئوری یادگیری و نظریه ابعادی آماری بنا شده است. این روش که در چند سال اخیر مورد توجه طیف وسیعی از محققان در رشته‌های مختلف قرار گرفته، ابتدا به‌وسیله Vapnik (۲۰۱۳) ارائه شد. از ویژگی‌های این روش آن است که می‌تواند

¹ Classifier

¹ Separating hyper-plane

کاربر مشخص شود و مقدار صحیح آن تاثیر به‌سزایی در صحت نتایج مدل SVM خواهد داشت.

مدل درخت رگرسیون تقویت شده

(BRT): الگوریتم مدل BRT، نتایج بسیاری از طبقه‌بندی‌کننده‌های ضعیف را برای شکل‌گیری یک طبقه‌بندی‌کننده قوی ترکیب می‌کند. در این روش درختان تصمیم با به‌کار بردن روش CART (درخت رگرسیون و طبقه‌بندی) ساخته می‌شود. روش CART یکی از روش‌های طبقه‌بندی است که نخستین بار به‌وسیله Breiman و همکاران (۱۹۸۴) ابداع و توسعه داده شد. این روش در ابتدا برای شناسایی بازاریابی محصولات مورد استفاده قرار گرفت (Golshani و همکاران، ۲۰۱۴). مدل BRT، در واقع یک روش آماری ناپارامتری است که از درخت رگرسیون برای حل مشکلات مربوط به رگرسیون خطی بهره می‌برد. در کل، تحلیل مربوط به CART از سه راه انجام می‌شود. ابتدا حداکثر درختان ساخته می‌شوند و سپس با مراحل هرس، انتخاب بهترین درختان و فرایند اعتباریابی متقاطع ادامه می‌یابد. عملکرد بهتر الگوریتم BRT، به تنظیم صحیح گزینه‌های مربوط به درختان تقویت شده و پارامترهای توقف شاخه‌بندی درختان وابسته است. در واقع عملیات تقویت^۱ به‌منظور بهبود قدرت پیش‌بینی درخت رگرسیون انجام می‌شود. این عملیات مشابه فرایند میانگین‌گیری مدل‌ها^۲ می‌باشد که طی آن از نتایج مشترک چندین مدل استفاده می‌شود، با این تفاوت که عملیات تقویت، یک فرایند مرحله‌ای و روبه‌جلو می‌باشد، بدین معنا که در هر مرحله تکرار، مدل‌ها به بخشی از مجموعه داده‌های آموزشی برازش داده می‌شوند (Naghibi و همکاران، ۲۰۱۶). لذا، دو پارامتر مهم یعنی پارامتر تقسیم^۳ و پارامتر کاهش^۴ برای مدل مطرح است. پارامتر تقسیم، درصد داده آموزشی در هر تکرار را مشخص کرده و به‌وسیله کاربر تعیین می‌شود. پارامتر کاهش یا نرخ یادگیری بیان‌گر سهم هر درخت در فرایند مدل‌سازی

پشتیبان نامیده می‌شوند که جزء اعضای کلیدی داده-های آموزشی محسوب می‌شوند (Pourghasemi و همکاران، ۲۰۱۳). هر چه صفحه جدایش بزرگ‌تر باشد، پایداری مدل به اختلال و نویز افزایش یافته و در نتیجه قدرت بالایی برای تعمیم خواهد داشت. الگوریتم‌های مدل SVM، روشی کلی برای تخمین توابع ارائه می‌نمایند که هدف اصلی آن‌ها حل مسائل بهینه‌سازی درجه دوم می‌باشد. طی مراحل مدل‌سازی، مدل SVM مجموعه‌ای از سلول‌های آموزشی مجزای خطی را انتخاب می‌نماید (Xi (i=1,2,...,n)). سلول‌های آموزشی شامل دو طبقه بوده که به صورت $Y_i = \pm 1$ مشخص می‌شوند. هدف مدل SVM، شناسایی یک صفحه جدایش n بعدی است که قادر باشد بیشترین فاصله و شکاف را بین این دو طبقه برقرار کند که نتیجتاً سبب کاهش متغیر $\|W\|$ می‌شود. بیان ریاضی این مطلب در روابط (۲) و (۳) ارائه شده است (Yao و همکاران، ۲۰۰۸؛ Xu و همکاران، ۲۰۱۲؛ Pourghasemi و همکاران، ۲۰۱۳).

$$\frac{1}{2}\|W\|^2 \quad (2)$$

$$Y_i((W.X_i) + b) \geq 1 \quad (3)$$

که در آن، $\|W\|$ قدر مطلق صفحه جدایش نرمال و b پایه (مبنای) عددی می‌باشد. برای حل مسأله فوق، از یک رابطه لاگرانژی استفاده می‌شود که حاوی یک ضریب افزایشنده به نام λ_i می‌باشد. هدف از این رابطه آن است که با کاهش ضرایب W و b و افزایش λ_i مقدار L لاگرانژی کاهش یابد. بنابراین، صورت کلی معادله به شکل رابطه (۴) تبدیل می‌شود (Vapnik، ۲۰۱۳).

$$L = \frac{1}{2}\|W\|^2 - \sum_{i=1}^n \lambda_i (Y_i((W.X_i) + b) - 1) \quad (4)$$

جهت تهیه نقشه حساسیت زمین‌لغزش مبتنی بر روش SVM، از نرم‌افزار R و بسته ksvm استفاده شد. در این سامانه توابع مختلفی برای مدل SVM ارائه شده که در این تحقیق از تابع شعاعی استفاده شد که در رابطه (۵) ارائه شده است، زیرا در بسیاری از موارد عملکرد مناسبی از خود ارائه داده است (ENVI، ۲۰۰۶).

$$K(X_i, X_j) = (-\gamma \|X_i - X_j\|), \quad \gamma > 0 \quad (5)$$

که در آن، γ پارامتر گاما بوده که باید به‌وسیله

¹ Boosting

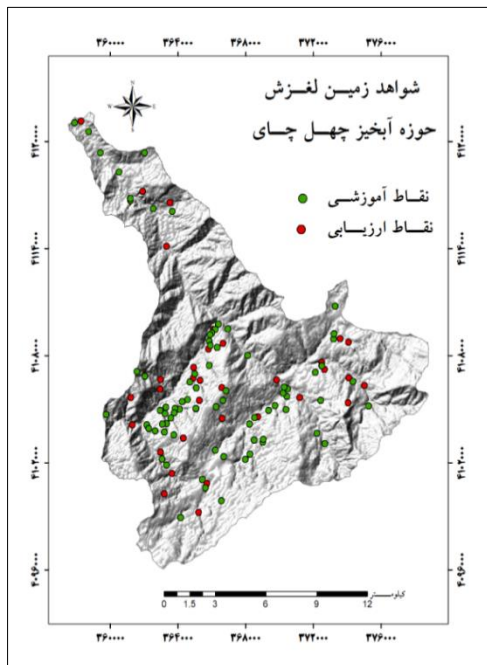
² Model averaging

³ Bag fraction parameter

⁴ Shrinkage parameter

نتایج و بحث

تهیه نقشه شواهد زمین لغزش و عوامل موثر: به طور کلی تعداد ۱۱۱ زمین لغزش به مساحت ۱۱۹۲ هکتار در حوزه آبخیز چهل چای ثبت شد. ۷۰ درصد از زمین لغزشها (۷۷ عدد) برای مدل سازی و بقیه (۳۴ عدد) برای ارزیابی مدلها به شکل تصادفی انتخاب شدند (شکل ۳). لایه های عوامل موثر در سامانه اطلاعات جغرافیایی (ArcGIS 10.2) تهیه و برای مدل سازی آماده شدند (شکل ۴). علائم نقشه های بافت خاک و سازندهای سنگ شناسی در جدول های ۱ و ۲ ارائه شده است.



شکل ۳- پراکنش زمین لغزش های حوزه آبخیز چهل چای

تهیه نقشه حساسیت زمین لغزش با مدل SVM: نقشه حساسیت زمین لغزش با استفاده از مدل SVM و تابع پایه شعاعی در محیط R تهیه شد. جهت خروجی گرفتن از نقشه حاصل از مدل SVM، بعد از مدل سازی و ران در نرم افزار R، خروجی وزن ها برای هر پیکسل (پیکسل به پیکسل) به محیط نرم افزار GIS منتقل و بر اساس آن نقشه نهایی تهیه شد. وزن های خروجی در بازه اعداد صفر و یک قرار دارند. پیکسل های دارای مقدار صفر و یک به ترتیب به منزله مناطق کاملاً پایدار و کاملاً ناپایدار می باشند. نهایتاً

و پیچیدگی درخت (تعداد نودهای هر درخت) می باشد (Elith و همکاران، ۲۰۰۸). نرخ یادگیری جهت جلوگیری از مشکل بیش برآزش داده ها در مدل تعبیه شده که طبق مطالعات تجربی، مقادیر ۰/۱ یا کمتر معمولاً به سمت مدل های مناسب گرایش داشته و برای مجموعه داده های کوچک ($n=500$) این نرخ می تواند برابر با ۰/۰۰۵ و برای داده های بزرگ ($n=5000$) برابر با ۰/۰۵ در نظر گرفته شود (Shataee و همکاران، ۲۰۱۲). مدل BRT توانایی دو الگوریتم را در خود گنجانده است؛ یکی الگوریتم درختان رگرسیون و دیگری یک روش تطبیقی جهت ترکیب کردن تعداد زیادی مدل ساده جهت کسب عملکرد مناسب (Elith و همکاران، ۲۰۰۸). این روش از چندین مزیت برخوردار است، از جمله این که می تواند داده های با حجم زیاد را با سرعت زیاد تجزیه و تحلیل کند، حساسیت کمتری به بیش برآزش^۱ مدلها دارد، نیاز به فرض های توزیع داده ندارد و همچنین، قادر به تعیین مهمترین عوامل در طبقه بندی است (Westreich و همکاران، ۲۰۱۰).

ارزیابی کارایی مدلها: منحنی ROC^۲ مشخصه عملکرد نسبی هر مدل است. منحنی ROC نموداری است که در آن نسبت پیکسل هایی که وقوع یا عدم وقوع زمین لغزش به وسیله مدل به درستی پیش بینی شده روی محور افقی (مثبت درست^۳ یا 1-Specificity) در برابر مقدار مکمل آن یعنی نسبت پیکسل هایی که نادرست پیش بینی شده اند (مثبت غلط^۴ یا Sensitivity) روی محور عمودی ترسیم می شود. این منحنی در نرم افزار SPSS محاسبه و رسم شد. مساحت زیر این منحنی AUC^۵ نام داشته و مدلی که بیشترین مقدار AUC را دارا باشد، دارای عملکرد نسبی بالاتری است. AUC برابر با ۰/۵ معادل مدل خنثی بوده و هر چه این مقدار به یک نزدیک تر شود، کارایی مدل افزایش می یابد (Pontius و Schneider، ۲۰۰۱).

^۱ Overfitting

^۲ Relative Operating Characteristic

^۳ True positive

^۴ False positive

^۵ Area Under Curve

تهیه نقشه حساسیت زمین لغزش با مدل BRT: جهت تهیه نقشه حساسیت با استفاده از مدل BRT، با لحاظ کردن عدد ۰/۰۰۵ برای نرخ یادگیری، عدد پنج برای پیچیدگی درخت و عدد ۰/۰۰۵ برای پارامتر تقسیم، تعداد ۹۰۰ درخت به‌عنوان تعداد درخت بهینه برای مدل‌سازی به‌دست آمد. در نهایت، نقشه حساسیت حاصل مشابه مدل SVM و بر اساس نقاط عطف منحنی تجمعی فراوانی پیکسل‌ها در چهار طبقه حساسیت (کم، متوسط، زیاد و خیلی‌زیاد) طبقه‌بندی شد (شکل ۵-ب).

مدل BRT درصد مشارکت عوامل استفاده شده در مدل‌سازی را نیز محاسبه می‌نماید که بر اساس آن عوامل جهت شیب، فاصله از جاده، ارتفاع و درجه شیب به‌ترتیب با درصد مشارکت ۲۲/۶، ۲۰/۱، ۱۳/۳ و ۸ درصد به‌عنوان مهمترین عوامل به‌وسیله مدل معرفی شده‌اند. میزان اهمیت سایر عوامل نیز در شکل ۶ ارائه شده است.

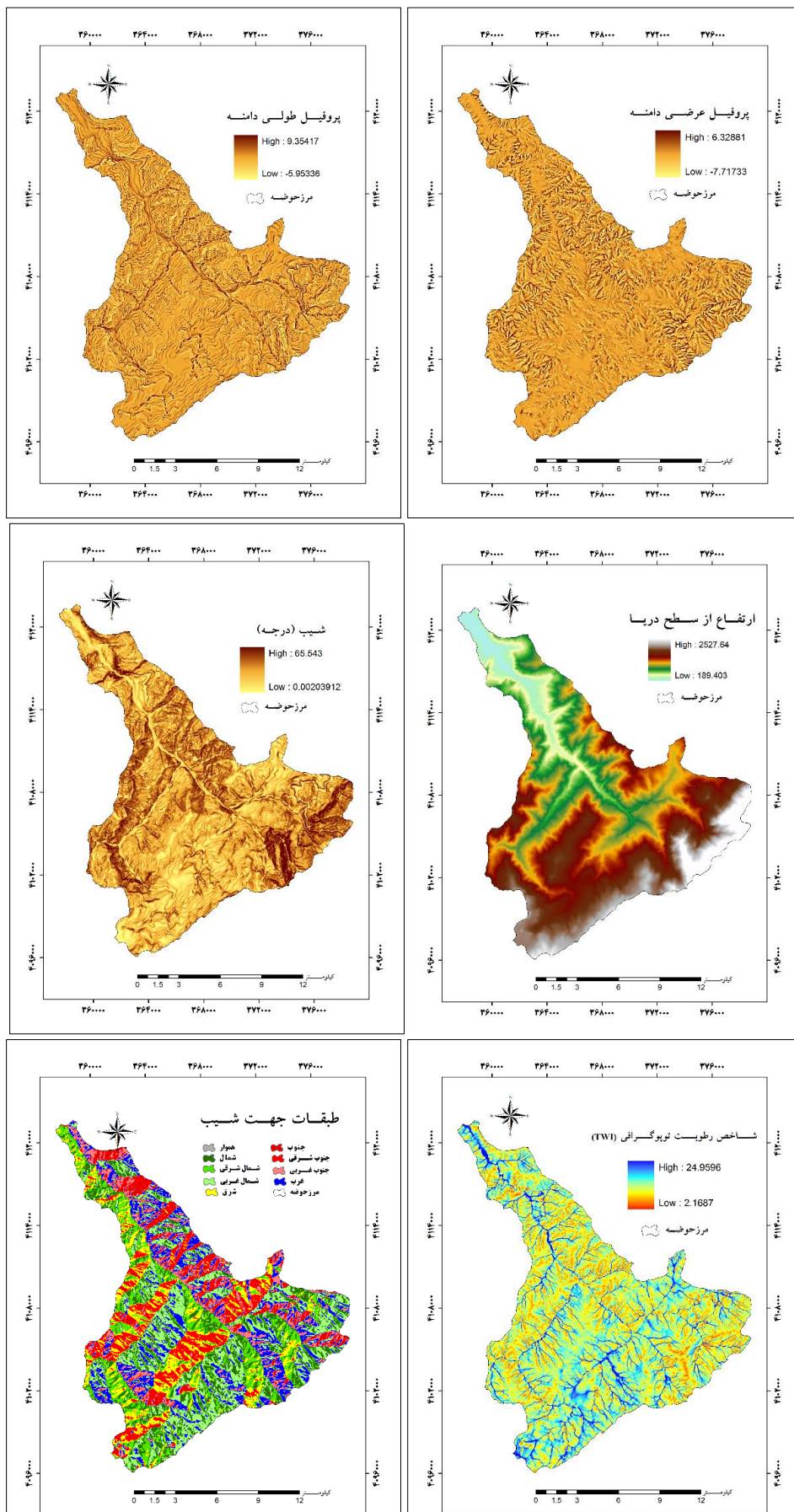
نقشه حساسیت حاصل در چهار طبقه حساسیت نهایی (کم، متوسط، زیاد، خیلی‌زیاد) طبقه‌بندی شد (شکل ۵-الف). طبق نتایج حاصل از مدل SVM، ۴۵ درصد از حوزه چهل‌چای در طبقات حساسیت زیاد و خیلی‌زیاد زمین لغزش قرار گرفته است. خروجی‌های مهم مدل SVM در جدول ۳ ارائه شده است.

جدول ۱- سازندهای سنگی حوزه آبخیز چهل‌چای

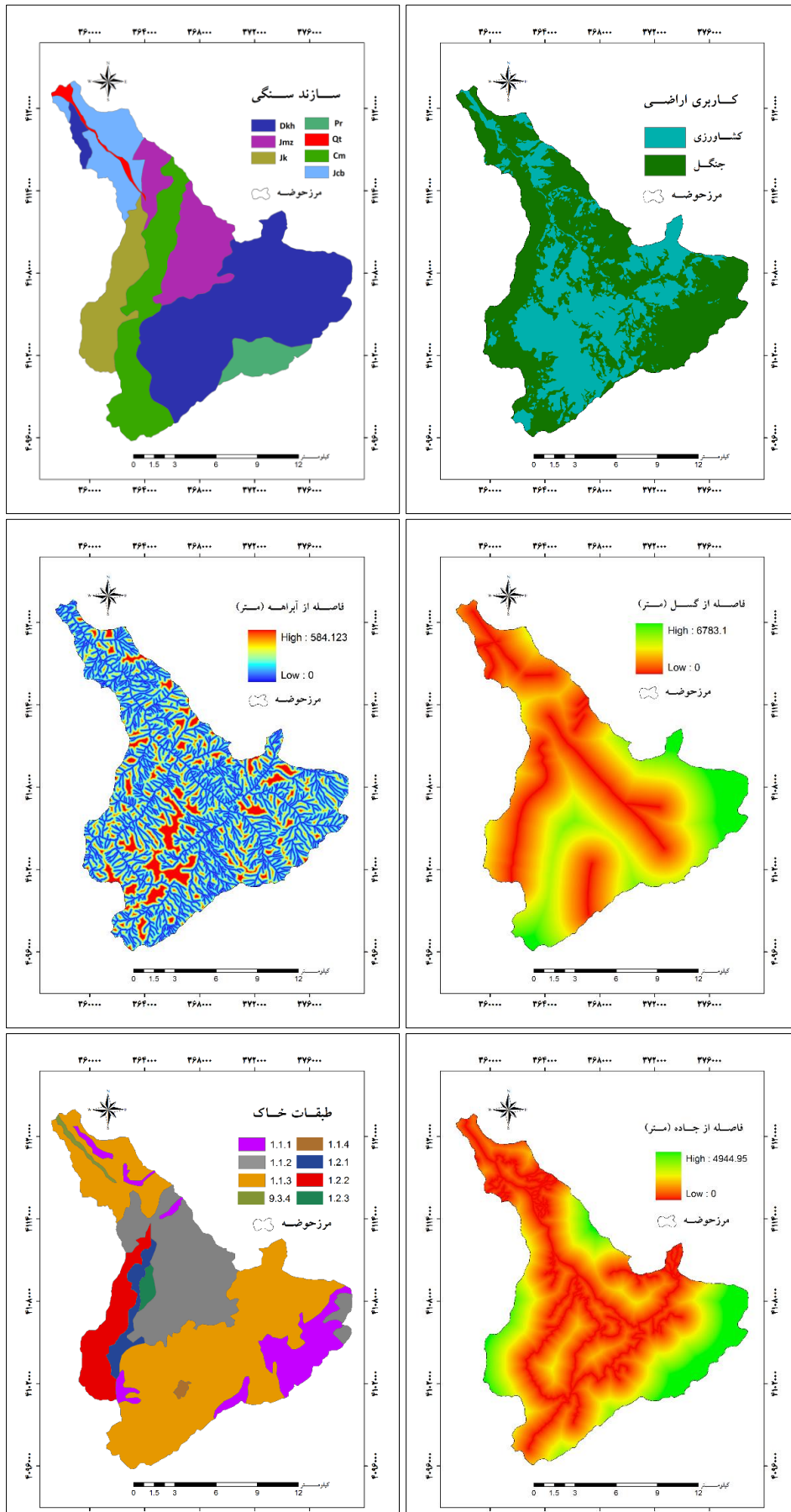
علائم اختصاری	نام سازند سنگ‌شناسی
Qt	پادگانه‌های آبرفتی کواترنری
Jcb	چمن‌بید
Jmz	مزدوران
Jk	کشف‌رود
Pr	روته
Cm	مبارک
Dkh	خوش‌بیلاق

جدول ۲- مشخصات خاک در حوزه چهل‌چای

مشخصات خاک	زیر واحد خاک
حداکثر ضخامت خاک ۳۵ سانتی‌متر، ضخامت لاشبرگ ۰-۲ سانتی‌متر، افق B تکامل نیافته با سنگریزه زیاد	۱،۱،۱
حداکثر ضخامت خاک ۱۰۰ سانتی‌متر، رسی لومی نسبتاً سنگین	۱،۱،۲
حداکثر ضخامت خاک بیش از ۱۰۰ سانتی‌متر، متوسط سیلتی لومی، سیلتی رسی، سیلتی رسی لومی	۱،۱،۳
حداکثر ضخامت خاک ۱۱۰ سانتی‌متر، pH خاک ۷/۲-۶/۴، ساختمان خاک دانه‌ای ریز تا درشت تا مکعبی، بافت خاک متوسط سیلتی لومی	۱،۱،۴
حداکثر ضخامت خاک ۴۰ سانتی‌متر، بافت خاک متوسط سیلتی لومی، نوع هوموس مول	۱،۲،۱
حداکثر ضخامت خاک ۱۰۰ سانتی‌متر، بافت خاک متوسط سیلتی لومی، سیلتی رسی لومی، نوع هوموس مول	۱،۲،۲
حداکثر ضخامت ۱۰۰ سانتی‌متر، بافت خاک سیلتی رسی لومی، درصد رس در بافت خاک در عمق ۹۰-۴۵ سانتی‌متر افزایش چشمگیر دارد.	۱،۲،۳
حداکثر ضخامت بیش از ۱۰۰ سانتی‌متر، بافت خاک سبک، تهویه‌پذیری خاک بالا، پایداری خاک ضعیف، نوع هوموس مول	۹،۳،۴



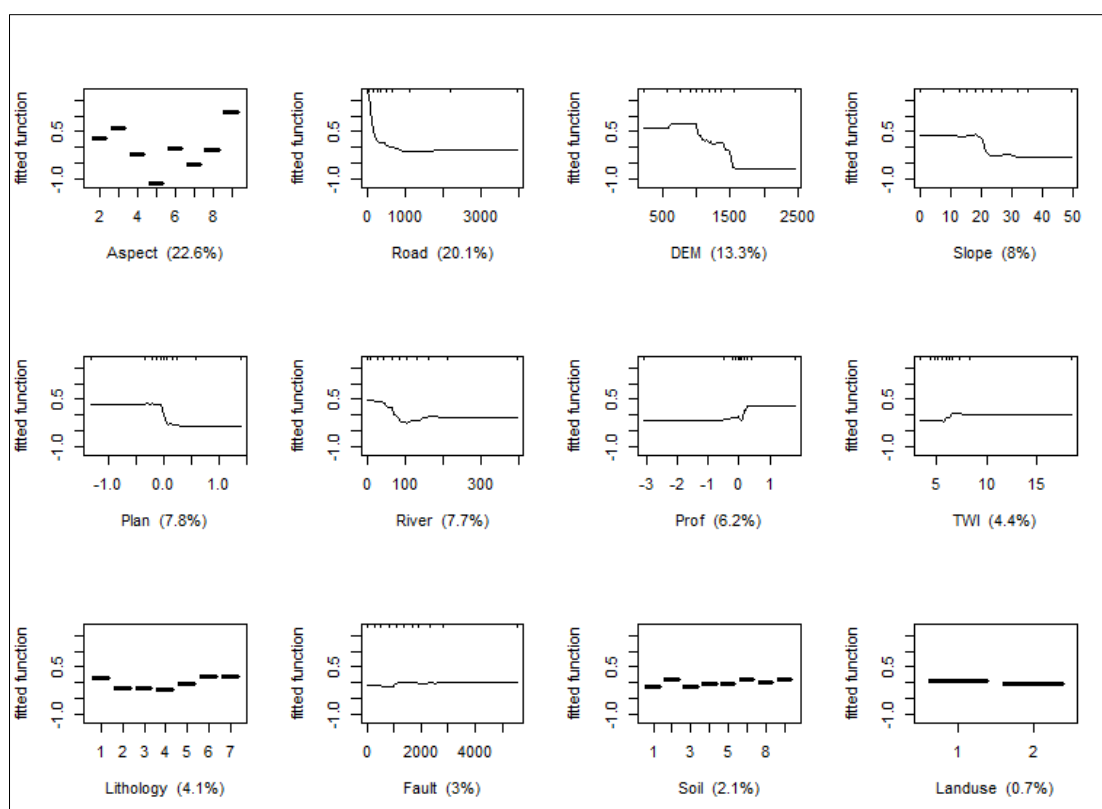
تصویر ۴- نقشه عوامل موثر در وقوع زمین لغزش در حوزه آبخیز چهل چای



شکل ۵- نقشه حساسیت زمین لغزش حوزه چهل چای با استفاده از دو مدل، الف- SVM و ب- BRT

جدول ۳- خروجی‌های مدل SVM در محیط R

نوع تابع	طبقات حساسیت	درصد هر طبقه
	پارامتر سیگما	۰/۰۷۵
شعاعی (RBF)	تعداد بردارهای پشتیبان	۱۲۷
	مقدار تابع هدف	-۹۰/۶۶
	خطای یادگیری	۰/۲۱۸



شکل ۶- میزان اهمیت عوامل موثر بر اساس مدل BRT

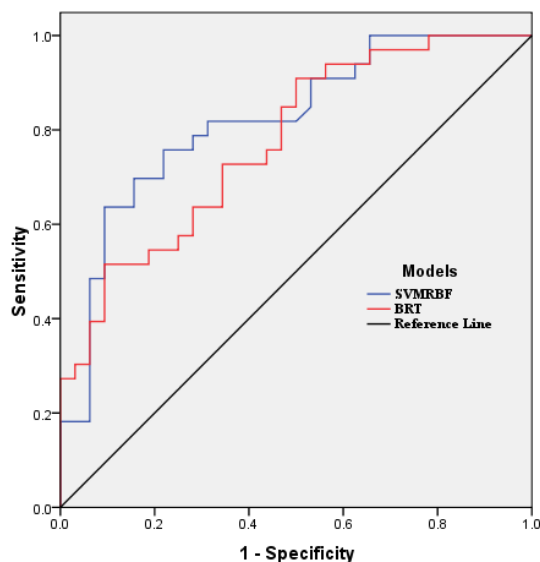
مقعر (اعداد منفی) به انحنای محدب (اعداد مثبت) میل می‌کنیم، حساسیت زمین‌لغزش کاهش یافته است. در انحنای طولی دامنه نیز، انحنای مقعر (اعداد مثبت) نسبت به انحنای محدب (اعداد منفی) حساسیت بیشتری به وقوع زمین‌لغزش داشته‌اند که با نتایج Moradi و Rezaei (۲۰۱۴) و Motamedvaziri و همکاران (۲۰۱۵) هم راستا است. علت این امر را می‌توان در این دانست که در دامنه‌های مقعر (برعکس دامنه‌های محدب) به دلیل پایین آمدن سرعت جریان و افزایش نگره‌داشت، آب زیرسطحی در مساحتی کوچک‌تر متمرکز شده که متعاقباً سبب بالا آمدن

همان‌طور که در شکل ۶ مشاهده می‌شود، جهت شمال غربی (کد ۹) دارای بیشترین حساسیت زمین‌لغزش در عامل جهت شیب می‌باشد. هر چه از جاده فاصله می‌گیریم حساسیت زمین‌لغزش کاهش یافته است. تقریباً همین الگو در عوامل ارتفاع، شیب و فاصله از آبراهه مشاهده شد. نمودار عامل فاصله از آبراهه می‌تواند معیار مناسبی جهت تعیین حریم آبراهه در حوضه چهل‌چای باشد (تنها با در نظر گرفتن خطر زمین‌لغزش)، به طوری که تا فاصله حدوداً ۵۰ متری حساسیت زمین‌لغزش ثابت بوده و بلافاصله پس از آن کاهش یافته است. هر چه از انحنای عرضی

جزئیات خروجی حاصل از آزمون فوق در شکل ۷ و جدول ۵ ارائه شده است.

جدول ۴- درصد مساحت طبقات حساسیت مدل SVM و BRT

مدل	طبقات حساسیت	درصد هر طبقه
	کم	۲۷/۰۶
ماشین بردار پشتیبان	متوسط	۲۷/۹۱
	زیاد	۲۴/۸۸
	خیلی زیاد	۲۰/۱۵
درخت رگرسیون تقویت‌شده	کم	۳۴/۷۸
	متوسط	۲۵/۴۸
	زیاد	۲۰/۷۳
	خیلی زیاد	۱۹/۰۱



شکل ۷- منحنی ROC برای دو مدل SVM و BRT در فرآیند مدل‌سازی حساسیت زمین لغزش حوضه چهل چای

سطح ایستایی، افزایش فشار آب منفذی و در نتیجه ناپایداری دامنه می‌شود. با افزایش مقادیر شاخص TWI و نزدیک شدن به زون اشباع، حساسیت زمین لغزش افزایش یافته است. سازندهای سنگی مبارک (Cm: کد ۱)، روت (Pr: کد ۶) و پادگانه‌های آبرفتی کواترنر (Qt: کد ۷) بیشترین حساسیت زمین لغزش را در عامل سنگ‌شناسی داشته‌اند. حساسیت زمین لغزش در عامل فاصله از گسل نسبتاً ثابت است و با شیب بسیار ملایمی با فاصله گرفتن از گسل در حال افزایش است که بیان‌گر نقش خنثی این عامل در حساسیت حوضه چهل چای به وقوع زمین لغزش می‌باشد. در عامل بافت خاک، طبقات ۹،۳،۴ (کد ۲: خاک سبک و ضعیف)، ۱،۲،۱ (کد ۶: خاک متوسط سیلتی لومی) و طبقه ۱،۲،۲ (کد ۸: خاک متوسط سیلتی لومی و سیلتی‌رسی لومی) بیشترین حساسیت به زمین لغزش را داشته‌اند. همچنین، کاربری کشاورزی نسبت به کاربری جنگل حساسیت بیشتری به زمین لغزش را در منطقه مورد مطالعه از خود نشان داده است. بر اساس نتایج مدل BRT، حدوداً ۴۰ درصد از حوضه چهل چای در پهنه حساسیت زیاد و خیلی زیاد زمین لغزش واقع شده است. درصد مساحت طبقات حساسیت مدل SVM و BRT جهت مقایسه در جدول ۴ ارائه شده است.

ارزیابی کارایی دو مدل SVM و BRT: همان‌طور که در بخش مواد و روش‌ها ذکر شد، از مساحت زیرمنحنی ROC (AUC) جهت ارزیابی کارایی دو مدل SVM و BRT استفاده شد.

جدول ۵- مقایسه مقادیر خروجی آزمون ROC بین دو مدل SVM و BRT

مدل‌های حساسیت	مساحت زیرمنحنی ROC	خطای انحراف معیار	معناداری	سطح اطمینان ۹۵ درصد
				باند پایین
SVM-RBF	۰/۸۲	۰/۰۵۲	۰/۰۰۰	۰/۷۱۷
BRT	۰/۷۷۴	۰/۰۵۷	۰/۰۰۰	۰/۸۸۵

نتیجه‌گیری

در تحقیق حاضر، دو مدل داده‌کاوی شامل ماشین بردار پشتیبان (SVM) و درخت رگرسیون تقویت‌شده (BRT) جهت ارزیابی حساسیت زمین لغزش در حوزه

بر اساس جدول ۵، مدل SVM (تابع شعاعی: RBF) با AUC برابر با ۰/۸۲ نسبت به مدل BRT با AUC برابر با ۰/۷۷۴، کارایی بهتری در مدل‌سازی فرایند زمین لغزش در حوضه چهل چای دارد.

آبخیز چهل چای واقع در استان گلستان به کار گرفته شد. تعداد ۱۱۱ زمین لغزش در حوضه ثبت، ۷۰ درصد آن‌ها برای مدل‌سازی و بقیه برای ارزیابی مدل‌ها استفاده شد. تعداد ۱۲ عامل کلیدی و موثر بر وقوع زمین لغزش شامل فاصله از گسل، فاصله از آبراهه، فاصله از جاده، سنگ‌شناسی، بافت خاک، کاربری اراضی، درجه شیب، جهت شیب، ارتفاع از سطح دریا، انحناء طولی و عرضی دامنه و شاخص رطوبت توپوگرافی (TWI) در فرایند مدل‌سازی استفاده شد. بر اساس نتایج مدل‌های SVM و BRT به ترتیب ۴۵ و ۴۰ درصد از حوضه مورد مطالعه در پهنه حساسیت زیاد و خیلی زیاد قرار گرفته است.

بر اساس نتایج مدل BRT، عوامل جهت شیب، فاصله از جاده، ارتفاع و درجه شیب به ترتیب با درصد مشارکت ۲۲/۶، ۲۰/۱، ۱۳/۳ و هشت درصد به عنوان مهمترین عوامل معرفی شدند. بر اساس نتایج حاصل از ارزیابی کارایی مدل‌ها با شاخص AUC، مدل SVM با مقدار برابر با ۰/۸۲ به عنوان مدل برتر شناخته شده و مدل BRT با مقدار ۰/۷۷۴ در رتبه بعدی اهمیت قرار گرفت.

نتایج این تحقیق با مطالعات Pourghasemi و همکاران (۲۰۱۳)، Yamani و همکاران (۲۰۱۲) و

Ahmadabadi و Rahmati (۲۰۱۵) مبنی بر موفقیت مدل SVM و با تحقیقات Naghibi و Pourghasemi (۲۰۱۵)، Naghibi و همکاران (۲۰۱۶)، Pourtghi و همکاران (۲۰۱۶) و Golshani و همکاران (۲۰۱۴) مبنی بر کارایی قابل قبول مدل BRT هم‌راستا می‌باشد. از علل موفقیت مدل SVM مبتنی بر مقایسه نتایج، می‌توان به پیش‌فرض‌های نظری قوی همراه با الگوریتم غیرخطی، توانایی بالا در یافتن مقادیر پارامترها و توانایی غلبه بر بیش‌برازش (Over-fitting) داده‌های تعلیمی اشاره کرد که سبب برتری این مدل به سایر مدل‌های ماشین یادگیری و رگرسیونی شده است. همچنین، از دلایل کارایی قابل قبول مدل BRT مبتنی بر مقایسه نتایج، می‌توان به قابلیت خودارزیابی و خوداصلاحی مدل در طی ساخت درخت‌های متعدد اشاره نمود، هر چند گاهی ساخت درخت‌های متعدد با شاخ و برگ انبوه (پارامتر عمق در هر درخت) سبب افزایش مدت مدل‌سازی شده و همچنین، مدل را به بیش‌برازش حساس‌تر می‌نماید.

در پایان، پیشنهاد می‌شود تا نتایج این تحقیق با سایر مدل‌های ماشینی و آماری دو متغیره و چند متغیره با ترکیب متفاوت عوامل در حوزه آبخیز مذکور و مناطق دیگر مقایسه شود.

منابع مورد استفاده

- Ahmadabadi, A. and M. Rahmati. 2015. Application of geomorphic quantitative indices in identification of landslide prone areas using SVM model, case study: Khoram Abad-Pol Zal highway. *Journal of Quantitative Geomorphological Researches*, 4: 197-213 (in Persian).
- Beven, K.J. and M.J. Kirkby. 1979. A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Journal*, 24(1): 43-69.
- Breiman, L., J.H. Friedman, R. Olshen and C.J. Stone. 1984. Classification and regression trees. *WIREs Data Mining and Knowledge Discovery*.
- Central Office of Natural Resources and Watershed Management in Golestan Province. 2007. Detailed action plan of Chehel-Chai Basin, Ravanab Consulting Engineers Co. Basic Studies.
- Cristianini, N. and J. Shawe-Taylor. 2000. An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press, 256 pages.
- Elith, J., J.R. Leathwick and T. Hastie. 2008. A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4): 802-813.
- ENVI Version 4.3. 2006. ITT industries, Inc. <http://www.RSInc.com/envi>.
- Ercanoglu, M. and C. Gokceoglu. 2002. Assessment of landslide susceptibility for a landslide-prone area (north of Yenice, NW Turkey) by fuzzy approach. *Environmental Geology*, 41:720-730.
- Golshani, P., A. Fallah, J. Oladi Ghadikolai and S. Kalbi. 2014. Evaluation of eoeye, 1 multispectral imagery data and texture analysis for urban scene classification, region 3 of Tehran City. *Physical Geography Research Quarterly*, 46(2): 157-168 (in Persian).
- Jenness, J. 2013. DEM surface tools for ArcGIS. Jenness Enterprises. Available at http://www.jennessent.com/arcgis/surface_area.htm. Accessed March.
- Koehorst, B.A.N., O. Kjekstad, D. Patel, Z. Lubkowski, J.G. Knoeff and G.J. Akkerman. 2005. Work package 6, determination of socio-economic impact of natural disasters. *Assessing Socio-Economic*

- Impact in Europe, 173 pages.
12. Lee, S., J. Hwang and I. Park. 2013. Application of data-driven evidential belief functions to landslide susceptibility mapping in Jinbu, Korea. *Catena*, 100: 15-30.
 13. Meten, M., N. PrakashBhandary and R. Yatabe. 2015. Effect of landslide factor combinations on the prediction accuracy of landslide susceptibility maps in the Blue Nile gorge of central Ethiopia. *Geo environmental Disasters*, 2(1): 1-17.
 14. Moradi, S. and M. Rezaei. 2014. A GIS-based comparative study of the analytic hierarchy process, bivariate statistics and frequency ratio methods for landslide susceptibility mapping in part of the Tehran metropolis, Iran. *Geopersia*, 4(1): 45-61.
 15. Motamedvaziri, B., S.M. Soleimanpour and A. Eshaghi. 2015. Mass movement. Islamic Azad University Science and Research Branch Press, Tehran, 61 pages.
 16. Naghibi, S.A. and H.R. Pourghasemi. 2015. A comparative assessment between three machine learning models and their performance comparison by bivariate and multivariate statistical methods in groundwater potential mapping. *Water Resources Management*, 29(14): 5217-5236.
 17. Naghibi, S.A., H.R. Pourghasemi and B. Dixon. 2016. GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree and random forest machine learning models in Iran. *Environmental Monitoring and Assessment*, 188(1): 1-27.
 18. Ownegh, M. 2002. Landslide hazard and risk assessment in the southern Suburbs of Newcastle. Sabbatical Research Report. No. 2, University of Newcastle, Australia, 85 pages.
 19. Pontius, R.G. and L.C. Schneider. 2001. Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA. *Agriculture, Ecosystems and Environment*, 85(1): 239-248.
 20. Pourghasemi, H.R., A.G. Jirandeh, B. Pradhan, C. Xu and C. Gokceoglu. 2013. Landslide susceptibility mapping using support vector machine and GIS at the Golestan Province, Iran. *Journal of Earth System Science*, 122(2): 349-369.
 21. Pourghasemi, H.R., H.R. Moradi, M. Mohammadi, B. Pradhan, R. Mostafazadeh and A. Goli Jirandeh. 2012b. Landslide hazard assessment using remote sensing data, GIS and weights-of-evidence model, South of Golestan Province, Iran. In *Asia Pacific Conference on Environmental Science and Technology, Advances in Biomedical Engineering*, 6: 30-36.
 22. Pourghasemi, H.R., M. Mohammady and B. Pradhan. 2012a. Landslide susceptibility mapping using index of entropy and conditional probability models in GIS: Safarood Basin, Iran. *Catena*, 97: 71-84.
 23. Pourtaghi, Z.S., H.R. Pourghasemi, R. Aretano and T. Semeraro. 2016. Investigation of general indicators influencing on forest fire and its susceptibility modeling using different data mining techniques. *Ecological Indicators*, 64: 72-84.
 24. Rahimi, M., M. Bay, M. Ownegh, A. Sadoddin and A. Salmanmahiny. 2010. IRS images data application for land use/ land cover mapping, case study: Chehel-Chai basin. 1st National Conference on Natural Resources, Kurdistan Province, Iran (in Persian).
 25. Shataee, S., S. Kalbi, A. Fallah and D. Pelz. 2012. Forest attribute imputation using machine-learning methods and ASTER data: comparison of k-NN, SVR and random forest regression algorithms. *International Journal of Remote Sensing*, 33(19): 6254-6280.
 26. Stumpf, A. and N. Kerle. 2011. Object-oriented mapping of landslides using random forests. *Remote Sensing of Environment*, 115(10): 2564-2577.
 27. Trigila, A., P. Frattini, N. Casagli, F. Catani, G. Crosta, C. Esposito and D. Spizzichino. 2013. Landslide susceptibility mapping at national scale: the Italian case study. In *Landslide Science and Practice*, Springer Berlin Heidelberg, 12: 287-295.
 28. United Nations Office for Disaster Risk Reduction. 2005. National report of the Islamic Republic of Iran. World Conference on Disaster Reduction, Kobe, Hyogo, Japan.
 29. Van Westen, C.J., T.W. Van Asch and R. Soeters. 2006. Landslide hazard and risk zonation-why is it still so difficult? *Bulletin of Engineering geology and the Environment*, 65(2): 167-184.
 30. Vapnik, V. 2013. *The nature of statistical learning theory*. Springer-Verlag New York, 314 pages.
 31. Varnes, D.J. and D.H. Radbruch-Hall. 1976. Landslides cause and effect. *Bulletin of the International Association of Engineering Geology*, 13(1): 205-216.
 32. Vorpahl, P., H. Elsenbeer, M. Märker and B. Schröder. 2012. How can statistical models help to determine driving factors of landslides? *Ecological Modelling*, 239: 27-39.
 33. Westreich, D., J. Lessler and M.J. Funk. 2010. Propensity score estimation: neural networks, support vector machines, decision trees (CART) and meta-classifiers as alternatives to logistic regression. *Journal of Clinical Epidemiology*, 63(8): 826-833.
 34. Xu, C., F. Dai, X. Xu and Y.H. Lee. 2012. GIS-based support vector machine modeling of earthquake-triggered landslide susceptibility in the Jianjiang River Watershed, China. *Geomorphology*, 145: 70-80.

35. Yamani, M., A. Ahmadabadi and R. Zare. 2012. The Zonation of landslide occurrence using of support vector machines algorithm, case study: Darakeh Basin. *Geography and Environmental Hazards*, 3: 125-142 (in Persian).
36. Yao, X., L.G. Tham and F.C. Dai. 2008. Landslide susceptibility mapping based on support vector machine: a case study on natural slopes of Hong Kong, China. *Geomorphology*, 101(4): 572-582.
37. Zevenbergen, L.W. and C.R. Thorne. 1987. Quantitative analysis of land surface topography. *Earth Surface Processes, Landforms*, 12: 47-56.

Landslide susceptibility assessment using data mining models, a case study: Chehel-Chai Basin

Aiding Kornejady¹ and Hamidreza Pourghasemi^{*2}

¹ PhD Student, Faculty of Range Land and Watershed Management, Gorgan University of Agricultural Sciences and Natural Resources, Gorgan, Iran and ² Assistant Professor, Faculty of Agriculture, Shiraz University, Shiraz, Iran

Received: 15 September 2016

Accepted: 20 December 2016

Abstract

The current study is aimed to map landslide susceptibility in the Chehel-Chai Basin is located in the Golestan Province. To this aim, two data mining models namely Support Vector Machine (SVM) and Boosted Regression Tree (BRT) were employed due to their robust computational algorithm. Landslide inventories were recorded through several field surveys using global positioning system, local information and available organizational resources and corresponding map was created in the geographic information system. Reviewing several worldwide studies, 12 predisposing factors including proximity to fault, proximity to stream, proximity to road, lithological units, soil texture, land use/cover, slope degree, slope aspect, altitude, plan curvature, profile curvature and topographic wetness index were chosen and the corresponding maps were produced in the geographic information system. In order to evaluate models' results the area under the receiver operating characteristic curve and 30% of landslide inventories were used. Results showed that the SVM model with the area under curve value of 0.82 had better performance on landslide susceptibility zonation over the study area and followed by the BRT model with the value of 0.77. Based on the SVM model results, about 45% of the Chehel-Chai Basin has high and very high landslide prone areas.

Keywords: Boosted regression tree, Golestan Province, Landslide, Geographic information system, Support vector machine

* Corresponding author: hr.pourghasemi@shirazu.ac.ir