

ارزیابی حساسیت زمین لغزش با استفاده از مدل جدید ترکیبی الگوریتم مینا (مطالعه موردی: شهرستان کامیاران، استان کردستان)

بهاره قاسمیان - دانش آموخته دکترای ژئومورفولوژی از گروه جغرافیای طبیعی، دانشگاه محقق اردبیلی، اردبیل، ایران.
موسی عابدینی* - استاد گروه جغرافیای طبیعی (ژئومورفولوژی)، دانشگاه محقق اردبیلی، اردبیل، ایران.
شهرام روستایی - استاد گروه جغرافیای طبیعی (ژئومورفولوژی)، دانشگاه تبریز، تبریز، ایران.
عطالله شیرزادی - کارشناس آموزشی و پژوهشی گروه مرتع و آبخیزداری، دانشکده منابع طبیعی، دانشگاه کردستان.

پذیرش مقاله: ۱۳۹۸/۰۲/۲۱ تأیید نهایی: ۱۳۹۹/۰۴/۱۲

چکیده

زمین لغزش‌ها به عنوان یکی از مخرب‌ترین پدیده‌های طبیعی محسوب می‌شوند. به دلیل تهدید آن‌ها، باید یک نقشه جامع حساسیت زمین لغزش برای کاهش آسیب‌های احتمالی به افراد و زیرساخت‌ها تهیه شود. کیفیت نقشه‌های حساسیت زمین لغزش تحت تأثیر بسیاری از عوامل، از جمله کیفیت داده‌های ورودی و انتخاب مدل‌های ریاضی است. هدف اصلی این پژوهش ارائه یک مدل ترکیبی جدید داده‌کاوی به نام *Rotation Forest - Functional Trees (RF-FT)* که یک رویکرد هوشمند ترکیبی از دو تکنیک یادگیری ماشین مدل *Functional Trees (FT)* و تکنیک طبقه‌بندی مدل *Rotation Forest (RF)* برای ارزیابی حساسیت زمین لغزش‌های اطراف شهر کامیاران واقع در استان کردستان می‌باشد. در ابتدا، بیست و یک عامل مؤثر بر وقوع زمین لغزش‌های منطقه مورد مطالعه شامل درجه شیب، جهت شیب، ارتفاع، انحنای شیب، انحنای عرضی شیب، انحنای طولی شیب، تابش خورشید، عمق دره، شاخص قدرت جریان، شاخص نمناکی توپوگرافی، شاخص طول دامنه، کاربری اراضی، تراکم پوشش گیاهی، فاصله از گسل، تراکم گسل، فاصله از جاده، تراکم جاده، فاصله از آبراهه، تراکم آبراهه، همپاران و لیتولوژی به همراه نقشه پراکنش زمین لغزش با ۶۰ نقطه لغزشی برای جمع‌آوری داده‌های آموزشی و آزمون جمع‌آوری شدند. سپس، بر اساس شاخص *Information Gain Ratio* هفده عامل مؤثر از بین آن‌ها انتخاب و جهت مدل سازی به کار گرفته شدند. در مرحله بعد مدل هیبریدی *RFFFT* برای ارزیابی حساسیت زمین لغزش با استفاده از مجموعه داده‌های آموزشی ساخته شد. عملکرد مدل پیشنهادی *RFFFT* با استفاده از چندین پارامتر آماری از جمله حساسیت، شفافیت، صحت، مجذور مربعات خطا، منحنی نرخ موفقیت و سطح زیر این منحنی مورد ارزیابی قرار گرفت. نتایج نشان داد که مدل پیشنهادی در این پژوهش به خوبی انجام شده ($AUC = 0.891$) و عملکرد طبقه‌بندی *FT* را به طور قابل توجهی بهبود داده است ($AUC = 0.819$). بنابراین مدل پیشنهادی *RFFFT* به عنوان یک روش جایگزین عالی برای ارزیابی حساسیت زمین لغزش در ناحیه مستعد زمین لغزش استفاده شود.

واژگان کلیدی: زمین لغزش، مدل ترکیبی، *Functional Trees Rotation Forest*، شاخص *dGR* کردستان، کامیاران.

مقدمه

نواحی که نشانه‌ای از بی‌ثباتی شیب وجود دارد و زمین‌لغزش‌هایی در آن نواحی مشاهده می‌شود، نظارت بر تغییر شکل سطح زمین جهت کاهش خطر مورد نیاز است. زمین لغزش یک فرایند تصادفی^۱، ناپیوسته و ایزوله شده‌ای از مخاطرات طبیعی است که در سطح کوچک اما با فراوانی زیاد اتفاق افتاده و مقدار انتقال دبی رسوب ناشی از آن‌ها به آبراهه در زمان و مکان متغیر (بندا و دون^۲، ۱۹۹۷: ۲۸۵۱) است. زمین لغزش در ایران به عنوان یک بلای طبیعی، سالیانه خسارات جانی و مالی فراوانی به بار می‌آورد. بر اساس یک برآورد اولیه، سالیانه ۵۰۰ میلیارد ریال خسارت‌های مالی از طریق این پدیده بر کشور تحمیل می‌شود و این در صورتی است که از بین رفتن منابع طبیعی غیر قابل بازگشت به حساب آورده نمی‌شود. از آنجایی که امروزه نیاز فزاینده‌ای جهت مدیریت خطر زمین‌لغزش به صورت کمی و پهنه‌بندی آن وجود دارد و تهیه‌ی نقشه‌ی وقوع خطر زمین‌لغزش، ابزاری اساسی برای فعالیت‌های مدیریت بحران در نواحی کوهستانی است (دهال^۳، ۲۰۰۸: ۴۹۶). بنابراین، جهت تهیه‌ی نقشه حساسیت زمین‌لغزش استفاده از روش‌های آماری در منطقه مورد مطالعه ضروری به نظر می‌رسد. روش‌های کمی زیادی در مطالعات پیش‌بینی مکانی زمین‌لغزش‌ها تاکنون استفاده شده است اما اخیراً برخی از محققین رو به استفاده از ترکیب مدل‌ها و روش‌های مختلف جهت افزایش صحت و دقت نقشه مکان‌یابی مخاطرات طبیعی آورده‌اند. به عنوان مثال؛ شیرزادی و همکاران (۱۳۹۶) به ارائه یک مدل جدید ترکیبی الگوریتم مینا به منظور پیش‌بینی حساسیت زمین لغزش‌های سطحی اطراف شهر بیجار پرداختند نتایج ارزیابی صحت نقشه پهنه‌بندی به دست آمده نشان داد که درصد مساحت زیر منحنی ROC (AUROC) برای داده‌های تعلیمی در مدل RF و مدل ترکیبی RF-RS به ترتیب ۰/۷۲۹ و ۰/۷۸۴ و برای مدل ترکیبی جدید ارائه شده به ترتیب ۰/۷۱۷ و ۰/۷۷۱ به دست آمدند. به طور کلی، نتایج نشان داد که تکنیک Random Subspace منجر به افزایش صحت پیش‌بینی مکانی حساسیت زمین لغزش‌های سطحی منطقه مورد مطالعه شده است. در پژوهشی دیگر طالبی و همکاران (۱۳۹۶) به تهیه نقشه خطر زمین لغزش با استفاده از الگوریتم جنگل تصادفی^۴ (محدوده‌ی مورد مطالعه: حوزه آبخیز سردارآباد، استان لرستان) پرداختند که نتایج آن‌ها نشان داد که درصد مساحت زیر منحنی نقشه پهنه بندی به دست آمده در مدل با مقدار ۹۸/۸ درصد بیان‌گر دقت عالی مدل جنگل تصادفی در منطقه مطالعاتی می‌باشد. چن^۵ و همکاران (۲۰۱۷) در پژوهشی به مقایسه مدل‌های لجستیک درختی^۶ (LMT)، جنگل تصادفی (RF) و طبقه‌بندی و رگرسیون درختی^۷ (CART) برای پیش‌بینی فضایی حساسیت زمین‌لغزش در لانگ کانتی^۸ واقع در کشور چین پرداختند. ارزیابی نتایج نشانگر آن بود که هر سه مدل عملکرد خوبی دارند؛ اما مدل RF دارای بالاترین قابلیت پیش‌بینی در مقایسه با مدل‌های LMT و CART است. مدل RF با نرخ موفقیت ۰/۸۳۷ و نرخ پیش‌بینی ۰/۷۸۱، یک روش امیدوارکننده برای ارزیابی حساسیت زمین‌لغزش است. بنابراین، هر سه مدل ابزار مفیدی برای پیش‌بینی فضایی حساسیت زمین‌لغزش هستند با این حال از آنجایی که رویکردهای یادگیری ماشین در مقایسه با رویکردهای دیگر از قبیل روش‌های مبتنی بر قضاوت‌های کارشناسی و روش‌های تحلیلی دیگر جهت پیش‌بینی فضایی زمین‌لغزش‌ها، کارآمدتر است (پرادهان^۹، ۲۰۱۳: ۳۶۳). لذا این مدل‌ها در سال‌های اخیر مورد توجه

1. Stochastic

2. Benda & Dunne

3. Dahal

4. Random Forest

5. Chen

6. Logistic Model Tree

7. Classification And Regression Tree

8. Long County

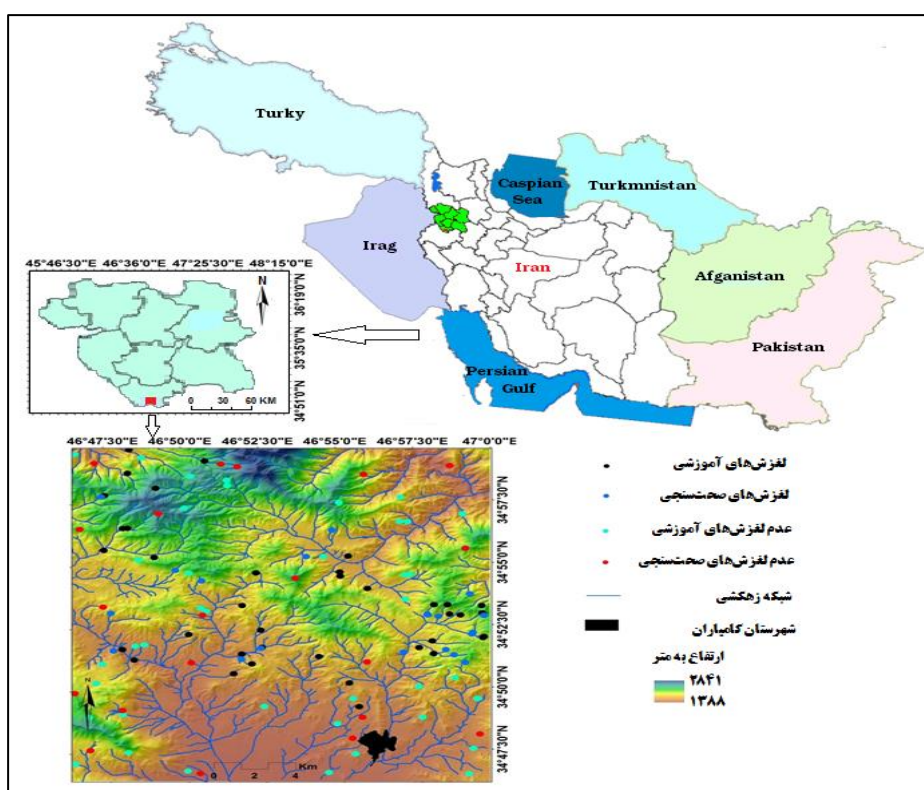
9. Pradhan

بسیاری از پژوهشگران بوده است که از نوین‌ترین و دقیق‌ترین آن‌ها می‌توان به مدل‌های RF^1 و FT^2 اشاره کرد. لذا با توجه به این که استان کردستان از لحاظ وقوع تعداد زمین‌لغزش با ثبت ۲۹۲ زمین‌لغزش بعد از استان‌های مازندران و گلستان در رتبه‌ی ۳ کشوری قرار دارد و چنانچه ملاک رتبه‌بندی مساحت استان‌ها باشد این استان در رتبه‌ی بالاتری قرار می‌گیرد (مرادی و همکاران، ۱۳۹۱: ۲۰۹). لذا اهمیت این پژوهش در شهرستان کامیاران مورد ارزیابی و بررسی قرار گرفته است.

داده‌ها و روش انجام پژوهش

موقعیت جغرافیایی و خصوصیات فیزیکی، زمین‌شناسی و اقلیمی منطقه مورد مطالعه

شهرستان کامیاران، واقع در بخش شرقی استان کردستان، در طول جغرافیایی $30^{\circ} 37' 46''$ تا $30^{\circ} 47' 46''$ شرقی و عرض جغرافیایی $30^{\circ} 34' 57''$ تا $30^{\circ} 47' 30''$ شمالی جای گرفته است. منطقه مورد مطالعه زمین‌لغزش‌های اطراف شهرستان کامیاران در مساحتی حدود $516/44$ کیلومتر مربع را شامل می‌گردد (شکل ۱). از نظر توپوگرافی و فیزیکی، حداکثر، حداقل و میانگین ارتفاع از سطح دریا به ترتیب 2841 ، 1757 و 1388 متر و اختلاف ارتفاع بین بلندترین و پست‌ترین نقطه 1453 متر می‌باشد. از لحاظ زمین‌شناسی منطقه مورد مطالعه در زون ساختاری سنندج - سرجان و زاگرس قرار دارد. مطابق با طبقه‌بندی اقلیمی دومارتن ($D = 23/56$) اقلیم منطقه مدیترانه‌ای می‌باشد. متوسط بارندگی منطقه مطالعاتی در طول دوره آماری سیزده ساله ($2012 - 2000$)، 560 میلیمتر در سال و میانگین درجه حرارت سالانه آن $13/69$ درجه سانتی‌گراد می‌باشد.



شکل ۱: موقعیت جغرافیایی منطقه مورد مطالعه (منبع: نگارندگان)

1. Rotation Forest

2. Functional Tree

نقشه پراکنش زمین لغزش های منطقه مورد مطالعه

روش پژوهش حاضر از نظر هدف کاربردی و از نظر ماهیت توصیفی - تحلیلی است که در آن از روش های کتابخانه‌ای، بازدیدهای میدانی و مدل سازی استفاده شده است. یکی از مهم ترین داده‌ها در ارزیابی حساسیت زمین لغزش، نقشه پراکنش زمین لغزش است. در پژوهش حاضر، تعداد ۶۰ نقطه‌ی لغزشی به‌وقوع پیوسته از سازمان جنگل‌ها، مراتع و آبخیزداری کشور تهیه و سپس با استفاده از پیمایش صحرایی، تفسیر عکس‌های هوایی و استفاده از تصاویر ماهواره‌ای گوگل ارث تفسیر و محل وقوع برخی از پهنه‌های زمین لغزشی اصلاح شد. از مجموع ۶۰ نقطه لغزشی به عنوان مناطق مستعد به زمین لغزش سطحی شناسایی شده در منطقه ۷۰ درصد به عنوان داده های آموزشی^۱ برای تعلیم و اجرای مدل و ۳۰ درصد به عنوان داده های آزمون^۲ جهت صحت‌سنجی طبقه بندی شدند. همچنین، تعداد ۶۰ مکان به عنوان مناطقی که زمین لغزش در آنجا اتفاق نیافتاده اند به صورت کاملاً تصادفی انتخاب و شبیه به مرحله مدل‌سازی و صحت‌سنجی به دو دسته ۷۰ و ۳۰ درصد طبقه بندی شدند.

عوامل مؤثر بر وقوع زمین لغزش

انتخاب عوامل مؤثر بر وقوع زمین لغزش به عنوان متغیرهای ورودی در مدل‌ها گامی اساسی در ارزیابی حساسیت زمین لغزش است (جدول ۱). همچنین کیفیت، وضوح، دقت و تغییرات زمانی داده‌های ورودی بسیار مهم است، زیرا این امر به طور قابل توجهی نتایج مدل آماری را تحت تأثیر خود قرار می‌دهد. اولین گام در پژوهش حاضر آماده‌سازی یک مدل رقومی ارتفاع^۳ (DEM) بود. این نقشه توسط دیجیتالی کردن نقشه توپوگرافی دقیق در مقیاس ۱:۲۵۰۰۰ و تفکیک مکانی ۲۰ متر از منطقه مطالعاتی ترسیم شده است. پارامترهای اولیه مانند شیب، جهت شیب، ارتفاع از سطح دریا، انحنای شیب، انحنای عرضی شیب^۴، انحنای طولی شیب^۵، شدت تابش خورشید^۶، شاخص طول دامنه^۷ و شاخص عمق دره^۸ شاخص قدرت جریان^۹، شاخص نمناکی توپوگرافی^{۱۰}، از نقشه رقومی ارتفاع در محیط Arc GIS 10.2 به طور مستقیم استخراج شدند. نقشه رقومی زمین‌شناسی شهرستان کامیاران با مقیاس ۱:۱۰۰۰۰۰ برگرفته از سازمان زمین‌شناسی کشور در نرم‌افزار Arc GIS 10.2 رقومی شده که منطقه مطالعاتی با انواع مختلف سازه‌های سنگ شناسی پوشیده شده است (جدول ۱). نقشه کاربری ارضی و شاخص پوشش گیاهی^{۱۱} از تصویر ماهواره‌ای ETM⁺ منطقه مورد مطالعه سال ۲۰۱۱، نقشه‌های لیتولوژی، فاصله از گسل و تراکم گسل از نقشه زمین‌شناسی شهرستان کامیاران با مقیاس ۱:۱۰۰۰۰۰، نقشه بارندگی یا خطوط همباران بر اساس رابطه رگرسیونی بین ارتفاع و میانگین بلند مدت سالیانه بارندگی ایستگاه‌های بارن‌سنجی داخل و خارج محدوده مورد مطالعه در طی دوره آماری سیزده ساله (۲۰۱۲ - ۲۰۰۰)، نقشه‌های فاصله از آبراهه و تراکم آبراهه براساس فواصل از اطراف شبکه آبراهه‌ها و نقشه‌های فاصله از شبکه جاده و تراکم جاده ترتیب براساس فواصل از اطراف شبکه جاده منطقه مورد مطالعه تهیه شدند. فاکتور شدت تابش خورشیدی بیانگر میانگین تابش خورشید در یک پیکسل مشخص در یک سال می‌باشد. همچنین این شاخص زاویه تابش و مسیری که تابش خورشیدی تا زمان برخورد با یک

1. Training

2. Testing

3. Digital Elevation Model

4. Profile curvature

5. Plan curvature

6. Solar radiation

7. Length-Slope (LS)

8. Valley Depth (VD)

9. Stream Power Index (SPI)

10. Topographic Wetness Index (TWI)

11. Normalized Difference Vegetation Index (NDVI)

پیکسل دارد را در نظر می‌گیرد. واحد این شاخص بر حسب کیلو وات ساعت بر مترمربع می‌باشد. دلیل اهمیت این شاخص به این دلیل است که مقدار بزرگ‌تر آن نشان‌دهنده بخار بیشتر از سطح خاک در یک ناحیه می‌باشد. به ازای مقدارهای بیشتر این شاخص، منافذ خاک برای تجمع بارندگی درون خاک در دسترس قرار می‌گیرد. مقدار کمتر آن، منافذ کمتری بین ذرات خاک را در اختیار می‌گذارد و هنگامی که باران ببارد، احتمال بالایی برای شکست خاک وجود دارد، زیرا مواد اشباع‌شده به دلیل اثر فشار منفذی سیال ضعیف‌تر می‌گردند. این فاکتور بر اساس شیب و جهت شیب در تحلیل‌گر ARC GIS 10.2 با استفاده از دستور "Area Solar Radiation" در پنچ کلاس طبقه‌بندی شده است. شاخص عمق دره و قدرت حمل جریان بر اساس نقشه رقومی ارتفاع در محیط نرم‌افزار SAGA تهیه و پس از اکسپورت کردن به محیط ARC GIS رستری و به ۵ کلاس به عنوان یکی از لایه‌های ورودی برای مدل‌سازی طبقه‌بندی گردید هم چنین این شاخص بر اساس رابطه (۳-۵) محاسبه می‌شود:

$$TWI = LN \left(\frac{A_a}{\tan \beta} \right) \quad (5)$$

رابطه (۳)

که در آن A_s مساحت ویژه بالادست در یک پیکسل (متر مربع بر متر) و β زاویه شیب در یک نقطه از دامنه می‌باشد. این شاخص بر اساس نقشه رقومی ارتفاع (DEM) در محیط نرم‌افزار SAGA تهیه و پس از اکسپورت کردن به محیط نرم‌افزار ARC GIS رستری و به ۵ کلاس به عنوان یکی از ورودی‌ها برای مدل‌سازی طبقه‌بندی گردید

جدول ۱: عوامل مؤثر بر وقوع زمین‌لغزش و طبقه‌بندی آن‌ها (منبع: نگارندگان)

| ردیف | عوامل مؤثر | طبقه‌بندی |
|------|-----------------------|---|
| ۱ | درجه شیب | (۱) ۰-۸، (۲) ۸-۱۵، (۳) ۱۵-۲۵، (۴) ۲۵-۳۵، (۵) ۳۵-۴۵، (۶) ۴۵-۷۰ |
| ۲ | جهت شیب | (۱) مسطح (۲) شمال (۳) شمال شرق (۴) شرق (۵) جنوب شرق (۶) جنوب (۷) شمال غرب (۸) غرب (۹) جنوب غرب |
| ۳ | ارتفاع به متر | (۱) ۱۶۰۰-۱۴۰۰، (۲) ۱۸۰۰-۱۶۰۰، (۳) ۲۰۰۰-۱۸۰۰، (۴) ۲۲۰۰-۲۰۰۰، (۵) ۲۲۰۰+ |
| ۴ | انحنای شیب | (۱) بسیار مقعر [(-۷/۲) - (-۷/۶)]، (۲) مقعر [(-۲/۴) - (-۷/۱)]، (۳) مسطح [(۰/۱۸) - (-۲/۳)]، (۴) محدب [(-۵/۶) - (۰/۱۸)]، (۵) بسیار محدب [(۰/۱۸) - (۵/۷)] |
| ۵ | انحنای عرضی شیب | (۱) [(-۱/۳) - (-۱/۵)]، (۲) [(-۰/۳۱) - (-۱/۲)]، (۳) [(-۰/۴۲) - (-۰/۳)]، (۴) [(۰/۴۳) - (۱/۵)]، (۵) [(۱/۶) - (۱/۵)] |
| ۶ | انحنای طولی شیب | (۱) [(-۱/۸) - (-۱/۶)]، (۲) [(-۰/۵۳) - (-۱/۷)]، (۳) [(-۰/۲۸) - (-۰/۵۲)]، (۴) [(۰/۲۹) - (۱/۲)]، (۵) [(۱/۳) - (۱/۰)] |
| ۷ | شدت تابش خورشید | (۱) ۵۱۷/۰۰۰ - ۲۳۷/۰۰۰، (۲) ۶۰۳/۰۰۰ - ۵۱۸/۰۰۰، (۳) ۶۷۴/۰۰۰ - ۶۰۴/۰۰۰، (۴) ۷۳۴/۰۰۰ - ۶۷۵/۰۰۰، (۵) ۷۳۵/۰۰۰ - ۸۸۱/۰۰۰ |
| ۸ | عمق دره به متر | (۱) ۰-۵۰، (۲) ۵۰-۱۰۰، (۳) ۱۵۰-۱۰۰، (۴) ۲۰۰-۱۵۰، (۵) ۲۰۰+ |
| ۹ | شاخص قدرت جریان | (۱) ۰-۱۰۰، (۲) ۱۰۰-۵۰۰، (۳) ۱۰۰۰-۵۰۰، (۴) ۱۵۰۰-۱۰۰۰، (۵) ۱۵۰۰+ |
| ۱۰ | شاخص نمناکی توپوگرافی | (۱) ۰-۳، (۲) ۳-۵، (۳) ۵-۷، (۴) ۷-۱۱، (۵) ۱۱+ |
| ۱۱ | شاخص طول دامنه | (۱) ۰-۲، (۲) ۲-۵، (۳) ۵-۱۰، (۴) ۱۰-۱۵، (۵) ۱۵+ |
| ۱۲ | کاربری اراضی | (۱) اراضی بایر، (۲) زراعت دیم، (۳) زراعت آبی، (۴) مرتع، (۵) منطقه مسکونی، (۶) مناطق تحت پوشش آب |
| ۱۳ | تراکم پوشش گیاهی | (۱) [(۰/۵۶) - (۰/۲۶)]، (۲) [(۰/۱۱) - (۰/۵۷)]، (۳) [(۰/۱۷) - (۰/۱۲)]، (۴) [(۰/۲۸) - (۰/۱۸)]، (۵) [(۰/۶) - (۰/۲۹)] |
| ۱۴ | فاصله از گسل به متر | (۱) ۰-۵۰، (۲) ۵۰-۱۰۰، (۳) ۱۵۰-۱۰۰، (۴) ۲۰۰-۱۵۰، (۵) ۲۰۰+ |
| ۱۵ | تراکم گسل | (۱) ۰/۴۳۷ - ۰/۰، (۲) ۱/۱۷ - ۰/۴۳۸، (۳) ۱/۹۳ - ۱/۱۸، (۴) ۲/۸۸ - ۱/۹۴، (۵) ۵/۸۷ - ۲/۸۸ |
| ۱۶ | فاصله از جاده به متر | (۱) ۰-۵۰، (۲) ۵۰-۱۰۰، (۳) ۱۵۰-۱۰۰، (۴) ۲۰۰-۱۵۰، (۵) ۲۰۰+ |

| | |
|------------------------|--|
| تراکم جاده | (۱) ۰-۰/۴۲۸، (۲) ۰/۱۷-۰/۴۲۹، (۳) ۱/۸۷-۱/۱۸، (۴) ۲/۷-۱/۸۸، (۵) ۴/۹۶-۲/۷۱ |
| فاصله از آبراهه به متر | (۱) ۰-۵۰، (۲) ۱۰۰-۵۰، (۳) ۱۵۰-۱۰۰، (۴) ۲۰۰-۱۵۰، (۵) > ۲۰۰ |
| تراکم آبراهه | (۱) ۰/۰۰۰۴۹-۰/۰، (۲) ۰/۰۰۱۳-۰/۰۰۰۵، (۳) ۰/۰۰۲۲-۰/۰۰۱۴، (۴) ۰/۰۰۳۲-۰/۰۰۲۳، (۵) ۰/۰۰۰۳ |
| همباران به میلی متر | (۱) ۵۰۰-۴۷۱، (۲) ۵۵۰-۵۰۱، (۳) ۶۰۰-۵۵۰، (۴) ۶۰۰-۵۵۱، (۵) ۶۵۰-۶۰۱، (۶) ۷۰۰-۶۵۱ |
| لیتولوژی | (۱) دوره کواترنر، (۲) اواخر کرتاسه - اوائل پالتوسن، (۳) اواخر اتوسن - اوائل الیگوسن، (۴) پالتوسن - اتوسن، (۵) اواخر کرتاسه - پالتوسن، (۶) اواخر الیگوسن - میوسن، (۷) دوره کرتاسه، (۸) ژوراسیک - کرتاسه |

روش انجام پژوهش

انتخاب مؤثرترین عوامل در مدل‌سازی حساسیت زمین لغزش

تکنیک‌های متعددی برای تشخیص صلاحیت داشتن قابلیت پیش‌بینی متغیرهای تأثیرگذار بر وقوع یک پدیده وجود دارند؛ از جمله شاخص IGR^۱ که توسط (کوئیلین، ۱۹۹۳) پیشنهاد داده شد که مقادیر IGR بالاتر نشان دهنده توانایی پیش‌بینی بالاتر آن عامل مؤثر برای مدل‌سازی است. بنابراین، جهت شناسایی مهم‌ترین عوامل مؤثر بر وقوع زمین لغزش‌های سطحی در منطقه مورد مطالعه، از میان بیست و یک عامل مؤثر معرفی‌شده، از آزمون IGR استفاده شد. اگر S داده‌های آموزشی با n نمونه ورودی و $n(L_i, S)$ تعداد نمونه‌ها در داده‌های آموزشی S متعلق به کلاس L_i (لغزشی، غیرلغزشی) باشد، در این صورت رابطه (۱) را داریم:

$$Info(S) = -\sum_{i=1}^2 \frac{n(L_i, S)}{|S|} \log_2 \frac{n(L_i, S)}{|S|} \quad (1)$$

با توجه به عوامل مؤثر بر وقوع زمین لغزش، مقدار اطلاعات مورد نیاز برای تقسیم S به مجموعه (S_1, S_2, \dots, S_m) ، از رابطه (۲) برآورد می‌گردد:

$$Info(S, A) = \sum_{j=1}^m \frac{S_j}{|S|} Info(S) \quad (2)$$

شاخص IGR برای یک عامل مؤثر خاص مثل فاکتور A (مثال: درجه شیب) از رابطه (۳) محاسبه می‌شود:

$$InformationGainRatio(S, A) = \frac{Info(S) - Info(S, A)}{SplitInfo(S, A)} \quad (3)$$

که در آن Split Info نشان‌دهنده اطلاعات بالقوه تولید شده توسط تقسیم S که داده‌های آموزشی به m زیر مجموعه Split Info از طریق رابطه (۴) محاسبه می‌گردد:

$$SplitInfo(S, A) = -\sum_{j=1}^m \frac{S_j}{|S|} \log_2 \frac{|S_j|}{|S|} \quad (4)$$

معرفی الگوریتم (Rotation Forest (RF)

این الگوریتم توسط رودریگز و همکاران در سال ۲۰۰۶ ارائه شد (رودریگز^۲ و همکاران، ۲۰۰۶: ۱۶۲۳). این الگوریتم به عنوان یک روش کارآمد برای بهبود طبقه‌بندی‌های ضعیف شناخته شده است (ازچیف و گلتن^۳، ۲۰۱۲). اگر چه RF در

^۱. Information Gain Ratio

^۲. Rodriguez

^۳. Ozcift & Gulten

بسیاری از زمینه‌های مختلف مانند پزشکی، طبقه‌بندی داده‌های سنجش از دور، مدل‌سازی حساسیت زمین‌لغزش به کار برده می‌شود (از چیفت و گلتن، ۲۰۱۱: ۹۴۳)، اما به عنوان نوعی از تکنیک‌های قدرتمند الگوریتم‌های یادگیری ماشین، پیش‌بینی طبقه‌بندی‌های پایه را بهبود می‌بخشند (کاوزوگلو و کلکسن^۱، ۲۰۱۳: ۴۲۳۲). RF با استفاده از تجزیه و تحلیل مؤلفه اصلی^۲ (PCA) جهت استخراج مجموعه داده‌های آموزشی به زیر مجموعه‌هایی جهت آموزش کلاس‌های طبقه‌بندی پایه استفاده می‌کند (فام^۳ و همکاران، ۲۰۱۶). ساختار الگوریتم RF به شرح زیر است:

اگر X داده‌های آموزشی با یک ماتریس $N \times n$ ، Y برچسب کلاس وقوع و عدم وقوع زمین‌لغزش، F مؤلفه‌های مجموعه داده و ω برچسب کلاس Y از یک گروه کلاس شامل $\{\omega_1, \dots, \omega_2\}$ باشد. فرض می‌گردد که یک گروه از مجموعه داده به صورت تصادفی به K زیرگروه و L درخت تصمیم^۴ در این الگوریتم به نام $\{D_1, \dots, D_L\}$ تفکیک می‌گردد. از طرفی L و K دو پارامتری هستند که بایستی تعیین شوند.

این مجموعه با استفاده از یک الگوریتم آموزشی، بر مبنای مشخص برای مجموعه داده‌های مختلف آموزشی ساخته شده است. جهت ساختن مجموعه‌ای از داده‌های آموزشی با هدف طبقه‌بندی D_i ، نیاز به انجام سه مرحله زیر است:

مرحله اول: پارامتر F به طور تصادفی با هر زیر مجموعه‌ای از تعداد مؤلفه‌هایش $M = n/K$ ، به K تقسیم می‌شود. مرحله دوم: در صورتیکه پارامتر F_{ij} در طبقه‌بندی مجموعه آموزشی D_i ، زیر مجموعه‌ای از ویژگی‌های j_{th} و پارامتر X_{ij} مشخص‌کننده ویژگی‌های X در پارامتر F_{ij} باشد: در این حالت، برای پارامتر X'_{ij} یک زیر مجموعه تصادفی با حداقل عناصر طراحی می‌شود و بعد یک مجموعه آموزش جدید که شامل مجموعه‌ای از ۷۵٪ از این فرم از مجموعه داده‌های آموزشی که شالوده (اساس) آن پارامتر X'_{ij} معین شده است، تشکیل می‌شود. در نهایت، برای تولید ضرایب ماتریس C_{ij} ، یک تبدیل خطی بر روی پارامتر X'_{ij} اعمال می‌شود.

اندازه هر ماتریس از X'_{ij} با ضرایب $a^{(1)}_{ij}, \dots, a^{(Mj)}_{ij}$ برابر با $M \times 1$ است

مرحله سوم: برای ایجاد یک ماتریس کوچک چرخشی، یعنی R_i ، به دست آوردن (محاسبه) ضرایب ماتریس C_{ij} به صورتی که در رابطه (۵) آمده، لازم است:

$$\begin{bmatrix} a^{(1)}_{i1}, \dots, a^{(M1)}_{i1} & \{0\} \dots & \{0\} \\ \{0\} & a^{(2)}_{i2}, \dots, a^{(M2)}_{i2} \dots & \{0\} \\ \dots & \dots & \dots \\ \{0\} & \dots & a^{(k)}_{ik}, \dots, a^{(Mk)}_{ik} \end{bmatrix} \quad (5)$$

این موقع، با توجه به مجموعه ویژگی‌های اصلی، ستون‌های از پارامتر R_i دوباره مرتب می‌شوند. ماتریس تنظیم شده چرخشی جدید به صورت زیر R_i^a در نظر گرفته می‌شود. سپس، (xR_i^a) به عنوان بسته آموزشی تبدیل شده برای طبقه‌بندی D_i مشخص خواهد شد. در نهایت، تمام طبقه‌بندی‌ها با روش موازی آموزش خواهند یافت (رودریگز و همکاران ۲۰۰۶). مراحل طبقه‌بندی برای مجموعه داده‌های آزمایشی X زمانی که (xR_i^a) احتمالاً حاصل از طبقه‌بندی D_i باشد، و نیز با این فرض که X متعلق به طبقه ω_j است، محاسبه می‌شود. بعداً، برای پی بردن به اطمینان از یک طبقه با استفاده از روش ترکیبی میانگین به صورت رابطه (۶) عمل می‌شود:

$$\mu_j^{(X)} = \frac{1}{L} \sum_{i=1}^L d_{ij} (xR_i^a), j = 1, \dots, c. \quad (6)$$

1. Kavzoglu & Colkesen

2. Principal Component Analysis

3. pham

4. Decision Tree

الگوریتم RF، سعی بر افزایش تنوع، دقت و صحت کل مجموعه بسته آموزشی دارد. در الگوریتم سنتی RF برای چرخش زیر مجموعه‌ای از سری داده‌های آموزشی، به دلیل حساسیت آنها به چرخش محور مؤلفه‌ها (عوامل یا عناصر)، روش درخت تصمیم (شبکه خوشه‌ای) انتخاب شده است. بنابراین، نام جنگل به دلیل نوع فرایند آن (فرم خوشه‌ای درختی) اطلاق شده است (اوزچیف و گالتین، ۲۰۱۱: ۴۴۵).

معرفی الگوریتم Functional Trees

الگوریتم درخت‌های عملکردی (FT) که توسط گاما پیشنهاد شده، یکی از مدل‌های سلسله مراتبی است که یک چهارچوب جهت ساخت درخت‌های چند متغیره برای مشکلات رگرسیونی و طبقه‌بندی دارد (گاما، ۲۰۰۱: ۱۴۹). تنها تفاوت اصلی بین FT و سایر مدل‌های سلسله مراتبی سنتی در این است که FT از تابع رگرسیون لجستیک^۱ برای تقسیم در گره‌های عملکرد داخلی و پیش‌بینی در برگ‌های عملکردی به جای تقسیم ورودی‌ها در یک گره درخت، با مقایسه مقدار برخی از صفات ورودی با ارزش ثابت استفاده می‌کند (لن^۲ و همکاران، ۲۰۱۱). در پژوهش حاضر عملکرد مدل FT در ارزیابی حساسیت زمین لغزش در مقایسه با مدل RF مورد بررسی قرار گرفت. مدل FT برای طبقه‌بندی در پژوهش حاضر در مراحل زیر انجام شده است: (۱) انتخاب سازنده از تابع تشخیصی خطی برای ساخت مدل $f = f(z_i)$ که احتمال توزیع کلاس‌های لغزشی و غیر لغزشی است (۲) گسترش Z_i با استفاده از عوامل مؤثر لغزشی جدید جهت ایجاد مجموعه داده جدید ساخته شده، در حالی که هر عامل مؤثر لغزشی جدید احتمال این را دارد که Z_i متعلق به لغزش یا غیر لغزش باشد (۳) انتخاب عوامل مؤثر لغزشی از مجموعه داده اصلی و تمام مجموعه داده‌های جدید ساخته شده برای ساخت درخت جهت طبقه‌بندی. در نهایت نقشه حساسیت زمین لغزش پس از آموزش مدل فوق ساخته شد.

مدل ترکیبی Rotation Forest - Functional Tree (RFFT)

مدل ترکیبی RFFT یک رویکرد ترکیبی از گروه FT و طبقه‌بندی FT است که اساس کار در سه مرحله به صورت زیر می‌باشد:

مرحله اول، بهینه‌سازی: در این مرحله مجموعه گروه RF جهت بهینه‌سازی داده‌های ورودی برای مدل‌سازی استفاده می‌شود. مجموعه داده‌های آموزشی اصلی به زیرمجموعه‌های آموزشی فرعی جهت طبقه‌بندی تقسیم می‌شوند. سپس مجموعه داده‌های بهینه زیر آموزشی به عنوان بهترین عملکرد مدل RFFT به دست آمد. در مرحله بعد ۱۶ تکرار جهت آزمون خطا تعیین شد تا بهترین روش برای آموزش مجموعه RF حاصل شود.

مرحله دوم، طبقه‌بندی: در این مرحله طبقه‌بندی FT جهت طبقه‌بندی کلاس‌های (لغزشی و غیر لغزشی) برای پیش‌بینی فضایی زمین لغزش‌ها مورد استفاده قرار گرفت. طبقه‌بندی‌های مختلف FT با استفاده از مجموعه‌های مختلف زیر آموزشی انجام شد. برای آموزش طبقه‌بندی FT، پارامترهای الگوریتم با استفاده از فرآیند آزمون‌های خطا مانند حداقل تعداد نمونه‌هایی که در آن یک گره برای تقسیم‌بندی در نظر گرفته شده که تعداد این تکرارها برای RF برابر با ۱۵ بود.

مرحله سوم، ترکیب: در مرحله‌ی نهایی مدل‌سازی، مجموعه RF برای ترکیب با طبقه‌بندی‌های به دست آمده از FT برای ساخت مدل نهایی RFFT مورد استفاده قرار گرفت.

بر این اساس می‌توان گفت که ابتدا بر روی داده‌های تعلیمی الگوریتم Rotation Forest به عنوان یک Meta classifier اعمال و سپس الگوریتم Functional Tree به عنوان یک decision tree در قسمت classifier

¹. Logistic Regression

². Lan

انتخاب می‌گردد. به این ترتیب یک انتخاب تصادفی در داده‌های تعلیمی ایجاد و با استفاده از الگوریتم Functional Tree برای هر داده زمین‌لغزش یک عدد احتمال بین صفر و یک محاسبه می‌گردد. در مرحله بعد همین روند برای داده‌های صحت‌سنجی تکرار و نتایج مدل‌سازی محاسبه می‌گردند. مدل‌سازی الگوریتم FT و مدل ترکیبی RF-FT در محیط نرم افزار Weka 3.7.12 انجام شد.

بحث و نتایج

معیار و ملاک‌های ارزیابی و مقایسه مدل‌سازی با الگوریتم RF و مدل ترکیبی RFFT

ارزیابی عملکرد الگوریتم‌های پیش‌بینی مکانی زمین‌لغزش‌های سطحی توسط معیارهای آماری کمی درصد مساحت زیر منحنی^۱، دقت^۲، حساسیت^۳، شفافیت^۴، شاخص کاپا^۵ و مجذور مربعات خطا^۶ بررسی شد (رابطه ۷ الی ۱۲). بر این اساس، در مدلی که مقادیر معیارهای اشاره شده به سمت یک تمایل داشته باشند، مدل مناسبی شناخته می‌شود (شیرزادی و همکاران، ۲۰۱۷: ۱۳). منحنی ROC ابزار مفیدی جهت نشان دادن کیفیت مدل‌ها و مساحت زیر منحنی ROC مقدار (AUC) برای اعتبارسنجی عملکرد کلی مدل ترکیبی RFFT در پژوهش حاضر انتخاب شد. نتیجه معیارهای مدل‌سازی در مناطقی که زمین لغزش حضور دارند و یا حضور ندارند به صورت دوگانه مثبت و منفی^۷ می‌باشند. این طبقه‌بندی منجر به ایجاد چهار حالت شامل TP، TN، FP و FN می‌شود. که در آن به ترتیب عبارتند از تعداد پیکسل زمین لغزش‌هایی (۱) که به درستی به عنوان زمین لغزش طبقه‌بندی شده‌اند (TP)، تعداد پیکسل عدم حضور زمین لغزش‌هایی (۰) است که به اشتباه به عنوان زمین لغزش طبقه‌بندی شوند (FP)، تعداد پیکسل عدم حضور زمین لغزش‌هایی که به درستی به عنوان عدم حضور زمین لغزش طبقه‌بندی شده‌اند (TN)، تعداد پیکسل حضور زمین لغزش‌هایی که به اشتباه به عنوان عدم حضور زمین لغزش طبقه‌بندی شده‌اند (FN) (تسانگاراتوس و بناردوس^۸، ۲۰۱۴: ۱۴۹۳). در نهایت بهترین نتیجه از این چهار حالت زمانی است که مقدار TP زیاد و مقدار FP کم باشد (عمر^۹ و همکاران، ۲۰۱۴).

$$AUC_{ROC} = \frac{\sum TP + \sum TN}{P + N} \quad (7)$$

که در آن P و N به ترتیب تعداد کل پیکسل‌های زمین لغزش و عدم حضور زمین لغزش می‌باشند. همچنین $X_{Predicted}$ نشان‌دهنده مقادیر پیش‌بینی شده در مجموعه داده‌های آموزشی یا مجموعه داده‌های اعتبارسنجی و \bar{X}_{Actual} نشان‌دهنده مقادیر واقعی (خروجی) مدل‌های حساسیت زمین لغزش است و Π تعداد کل نمونه‌ها در مجموعه داده‌های آموزش یا مجموعه داده‌های اعتبارسنجی است.

$$\text{Sensitivity} \quad SST = \frac{TP}{TP+FN} \quad (8)$$

$$\text{Specificity} \quad SPF = \frac{TN}{FP+TN} \quad (9)$$

$$\text{Accuracy} \quad ACC = \frac{TP+TN}{TP+TN+FN+FN} \quad (10)$$

1. Area Under the ROC Curve

2. Accuracy

3. Sensitivity

4. Specificity

5. Kappa

6. Root Mean Squared Error (RMSE)

7. Positive and Negative

8. Tsangaratos & Benardos

9. Omar

$$RMSE = \left(\frac{1}{n}\right) \sum_{i=1}^n (X_{Predicted} - \bar{X}_{actual})^2 \quad (11)$$

$$k = Pc - P \exp / 1 - P \exp \quad (12)$$

$$P \exp = \left((TP + FN) \times (TP + FP) + (FP + TN) \times (FN + TN) \right) / \sqrt{(TP + TN + FP + FN)}$$

تعیین مهم ترین عوامل مؤثر بر وقوع زمین لغزش های منطقه مورد مطالعه

عوامل مؤثر بر وقوع زمین لغزش برای ارزیابی توانایی پیش بینی مدل های لغزشی در منطقه مطالعاتی با استفاده از روش IGR ارزیابی شده است. (جدول ۲) نتایج میانگین شاخص IGR برای بیست و یک عامل مؤثر انتخاب شده بر وقوع زمین لغزش های سطحی منطقه مورد مطالعه آمده است. این نتایج نشانگر آن است بیشترین مقادیر این شاخص به ترتیب به عوامل فاصله از آبراهه، تراکم آبراهه، شاخص قدرت حمل جریان، بارندگی، عمق دره، شاخص نمناکی توپوگرافی، شدت تابش خورشیدی و انحنای طولی شیب اختصاص یافتند. در صورتی که عوامل دیگر به خاطر دارا شدن مقدار صفر برای این شاخص با ایجاد یک حالت نویز منجر به اثرگذاری منفی بر نتایج مدل سازی شده و همچنین به دلیل عدم تأثیرگذاری بر وقوع زمین لغزش های سطحی از مدل سازی حذف و مدل سازی با ۸ فاکتور باقیمانده انجام شد.

جدول ۲: تعیین مهم ترین عوامل اثرگذار بر لغزش های منطقه مطالعاتی بر اساس شاخص (IGR) (منبع: نگارندگان)

| عوامل مؤثر | میانگین IGR | عوامل مؤثر | میانگین IGR |
|-----------------------|-------------|-------------------|-------------|
| فاصله از آبراهه | ۰/۹۵۱ | انحنای معمولی شیب | ۰/۰۰۰ |
| تراکم آبراهه | ۰/۴۵۶ | انحنای عرضی شیب | ۰/۰۰۰ |
| شاخص قدرت حمل جریان | ۰/۱۲۸ | شاخص طول دامنه | ۰/۰۰۰ |
| بارندگی | ۰/۱۲۳ | کاربری اراضی | ۰/۰۰۰ |
| عمق دره | ۰/۰۸۴ | لیتولوژی | ۰/۰۰۰ |
| شاخص نمناکی توپوگرافی | ۰/۰۶۳ | شاخص پوشش گیاهی | ۰/۰۰۰ |
| شدت تابش خورشیدی | ۰/۰۲۵ | فاصله از گسل | ۰/۰۰۰ |
| انحنای طولی شیب | ۰/۰۰۸ | فاصله از جاده | ۰/۰۰۰ |
| درجه شیب | ۰/۰۰۰ | تراکم گسل | ۰/۰۰۰ |
| جهت شیب | ۰/۰۰۰ | تراکم جاده | ۰/۰۰۰ |
| ارتفاع از سطح دریا | ۰/۰۰۰ | ----- | ----- |

مدل سازی و ارزیابی الگوریتم FT و مدل ترکیبی RFFT با داده های تعلیمی (Training dataset)

(جدول ۳) نتایج ارزیابی مدل سازی با دو روش FT و RFFT برای داده های تعلیمی و صحت سنجی را نشان می دهد. مشاهده جدول نشان می دهد که معیارهای کاپا، دقت، حساسیت، شفافیت، مجذور مربعات خطا و مساحت زیر منحنی ROC برای داده های تعلیمی در مدل FT به ترتیب ۰/۹۲۵، ۰/۹۷۵، ۰/۹۵۰، ۱، ۰/۱۷۲ و ۰/۹۸۶ می باشند. در حالی که این مقادیر به ترتیب برای مدل ترکیبی جدید ارائه شده دارای مقادیر ۰/۹۷۵، ۰/۹۹۰، ۰/۹۸۳، ۱، ۰/۱۳۹ و ۰/۹۹۹ می باشند. از طرفی نتایج این مقادیر برای داده های صحت سنجی در مدل FT شامل ۰/۵۵۰، ۰/۸۳۳، ۰/۷۰۰، ۰/۹۶۷، ۰/۴۱۱ و ۰/۸۵۵ می باشند. در حالی که این مقادیر به ترتیب برای مدل ترکیبی جدید ارائه شده دارای مقادیر ۰/۶۰۰، ۰/۸۶۷، ۰/۷۶۷، ۰/۹۶۷، ۰/۳۷۸ و ۰/۹۲۰ می باشند. این به این معنی می باشد که تمامی این مقادیر در مدل ترکیبی نسبت به مدل دیگر دارای مقادیر بالاتری می باشند. بنابراین اگرچه مدل FT نیز کارایی خوبی از خود نشان داد اما کارایی تکنیک FT به کمک RF به میزان کارایی مدل ترکیبی افزود.

جدول ۳: ارزیابی مدل‌سازی داده‌های تعلیمی و آموزشی با الگوریتم FT و مدل ترکیبی RFFT (منبع: نگارندگان)

| داده‌های صحت سنجی | | داده‌های تعلیمی | | معادل لاتین | شاخص |
|-------------------|-------|-----------------|-------|----------------|------------------|
| RFFT | FT | RFFT | FT | | |
| ۰/۶۰۰ | ۰/۵۵۰ | ۰/۹۷۵ | ۰/۹۲۵ | Kappa | کاپا |
| ۲۳ | ۲۱ | ۵۹ | ۵۷ | True positive | مقدار TP |
| ۲۹ | ۲۹ | ۴۰ | ۶۰ | True negative | مقدار TN |
| ۱ | ۱ | ۰ | ۰ | False positive | مقدار FP |
| ۷ | ۹ | ۱ | ۳ | False negative | مقدار FN |
| ۰/۸۶۷ | ۰/۸۳۳ | ۰/۹۹۰ | ۰/۹۷۵ | Accuracy | دقت |
| ۰/۷۶۷ | ۰/۷۰۰ | ۰/۹۸۳ | ۰/۹۵۰ | Sensitivity | حساسیت |
| ۰/۹۶۷ | ۰/۹۶۷ | ۱ | ۱ | Specificity | شفافیت |
| ۰/۳۷۸ | ۰/۴۱۱ | ۰/۱۳۹ | ۰/۱۷۲ | RMSE | مجدور مربعات خطا |
| ۰/۹۲۰ | ۰/۸۵۵ | ۰/۹۹۹ | ۰/۹۸۶ | ROC (AUC) | مساحت زیر منحنی |

تهیه نقشه‌های پیش‌بینی مکانی حساسیت زمین‌لغزش‌های سطحی منطقه مورد مطالعه

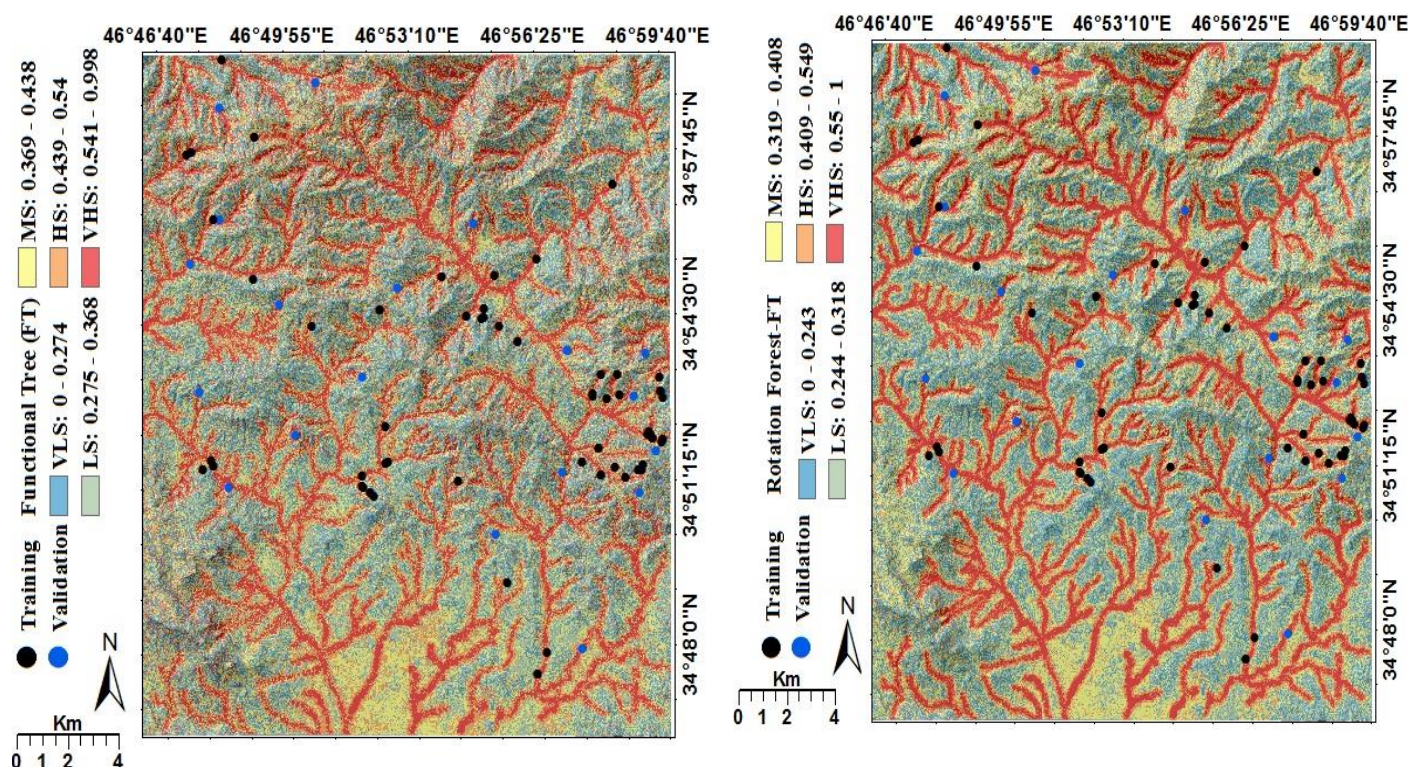
مدل‌های لغزشی بر اساس الگوریتم‌های FT و ترکیب دو الگوریتم یعنی RFFT با استفاده از مجموعه داده‌های آموزشی برای پیش‌بینی فضایی زمین‌لغزش‌های منطقه مطالعاتی ترسیم شدند. نتایج مربوط به قابلیت پیش‌بینی تکنیک‌های جدید یادگیری ماشین در مدل RFFT به دو پارامتر بستگی دارد: ۱: تعداد سیدها^۱ و ۲: تعداد تکرارها^۲ در مجموعه داده‌های آموزش و اعتبارسنجی. در پژوهش حاضر برای ساخت این مدل‌های لغزشی، محاسبه پارامترها از طریق روند آموزش با استفاده از فرآیند آزمایش و خطا بهینه‌سازی شده است (جدول ۴) که در آن تعداد فولدها برای هرس کاهش خطا و تعداد سیدها جهت تقسیم داده‌ها مورد استفاده قرار گرفته است. علاوه بر این، با استفاده از پارامترهای بهینه انتخاب شده، مدل ترکیبی پیشنهاد شده و سایر مدل‌های مقایسه‌ای بر اساس مجموعه داده‌های آموزشی و اعتبارسنجی ساخته شد.

جدول ۴: محاسبه پارامترهای الگوریتم‌های یادگیری ماشین به کار رفته در منطقه مورد مطالعه (منبع: نگارندگان)

| الگوریتم‌ها | پارامترها |
|-------------------------|---|
| FT (as base classifier) | Bin Split, False; error on probabilities, False; min Number Instance, 15, Model type, FT; number Boosting Iterations, 15; use AIC, False |
| RF | Max Group, 3; Min Group, 3; Number of iterations, 15; Projection filter, Principal Component Analysis (PAC); Removed percentage, 50; Number of seed, 3; use a base classifier, Functional Tree (FT) |

1. seed

2. iterations



شکل ۲: نقشه‌های حساسیت زمین لغزش شهرستان کامیاران با مدل‌های RF و مدل ترکیبی RS-RF (منبع: نگارندگان)

به طور کلی، در مدل‌سازی حساسیت زمین لغزش، دو فرآیند اصلی وجود دارد: ساخت و ارزیابی مدل و به دنبال آن ساخت و ارزیابی نقشه حساسیت. ساخت مدل و ارزیابی آن که در طی فرآیند مدل‌سازی در بالا ذکر شد و در آن مدل جدید RFFT که یک رویکرد ترکیبی از گروه RF و طبقه‌بندی FT است ایجاد شد، عملکرد بسیار خوبی نسبت به مدل‌های تکی FT و RF جهت مدل‌سازی زمین لغزش در منطقه مطالعاتی دارد در نهایت نقشه‌های حساسیت زمین لغزش سطحی بر اساس الگوریتم‌های FT و ترکیب دو الگوریتم یعنی RFFT برپایه روش natural breaks در محیط ARC GIS در کلاس‌های حساسیت بسیار کم، کم، متوسط، زیاد و بسیار زیاد، ترسیم شدند. شکل شماره ۲ این نقشه‌ها را برای روش RF و مدل ترکیبی RS-RF نشان می‌دهند.

ارزیابی صحت نقشه‌های حساسیت زمین لغزش‌های منطقه مورد مطالعه

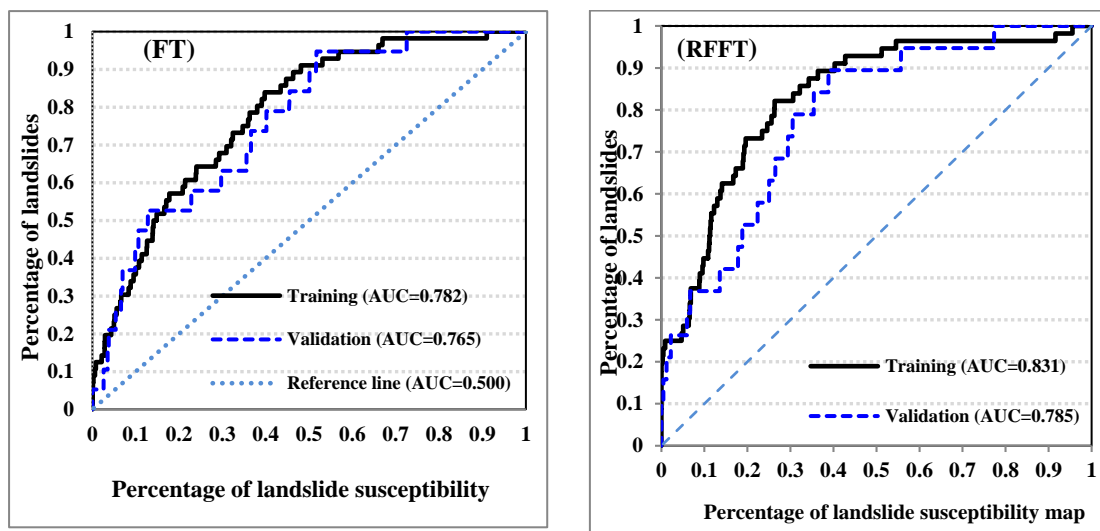
اعتبارسنجی مدل‌های پیش‌بینی شده برای بررسی صحت نقشه‌های حساسیت زمین لغزش ضروری است (چانگ و فابری^۱، ۲۰۰۳، ۴۵۹). اعتبار نقشه‌های حساسیت زمین لغزش را می‌توان به صورت گرافیکی با روش درصد مساحت زیر منحنی ROC و مقدار (AUC) محاسبه کرد. منحنی ROC یک منحنی است که بر روی محور X مقدار Specificity (نسبت پیکسل‌های عدم حضور زمین لغزش که به درستی به عنوان عدم حضور زمین لغزش طبقه‌بندی شده‌اند) (رابطه ۱۳) و بر روی محور Y مقدار Sensitivity (نسبت پیکسل‌های زمین لغزش که به درستی به عنوان زمین لغزش طبقه‌بندی شده‌اند) (رابطه ۱۴) مشخص می‌باشد.

¹. Chang & Fabbri

$$\text{Specificity} = n(TN) / [n(TN) + n(FP)] \quad (۱۳)$$

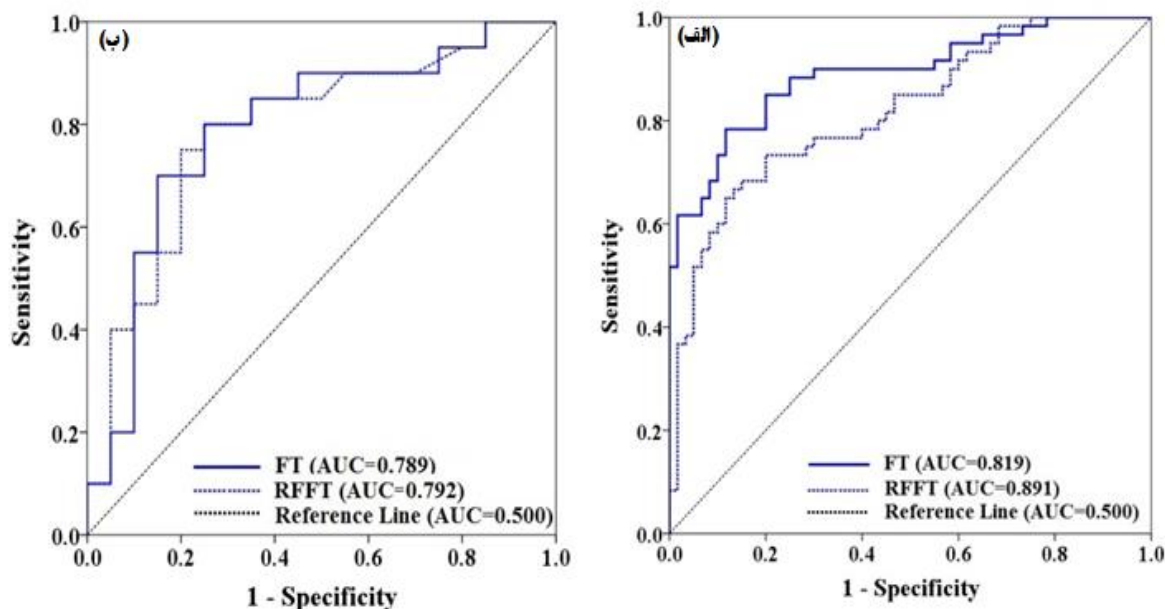
$$\text{Sensitivity} = n(TP) / [n(TP) + n(FN)] \quad (۱۴)$$

و از مقدار AUC، درصد نرخ موفقیت و قابلیت پیش‌بینی هر دو مجموعه داده‌های آموزشی و داده‌های اعتبارسنجی به دست آمد. در موضوع ارزیابی صحت نقشه مکان یابی مناطق حساس به وقوع زمین‌لغزش هم از داده‌های تعلیمی و هم از داده‌های صحت سنجی استفاده می‌گردد. بر این اساس، هنگامی که از داده‌های تعلیمی استفاده شود منحنی ارزیابی صحت success rate curve (SRC) نام دارد و در هنگام استفاده از داده‌های صحت‌سنجی به منحنی ارزیابی صحت نقشه پیش‌بینی مکانی، واژه prediction rate curve (PRC) اطلاق می‌گردد (عابدینی و همکاران ۲۰۱۷، ۸؛ شیرزادی و همکاران، ۲۰۱۷: ۱۳). مقادیر AUC مابین ۰/۵ الی ۱ است و ایده‌آل‌ترین مدل، بیشترین سطح زیر منحنی را دارد (پورقاسمی، ۱۳۸۹: ۷۷). چنان‌چه مدلی نتواند رخداد لغزشی را بهتر از دیدگاه احتمالی تخمین زند. مقدار AUC آن ۰/۵ است و زمانی که منحنی ROC سطح زیر منحنی برابر با یک داشته باشد، بهترین دقت را از نقشه پهنه‌بندی نشان می‌دهد (شیرزادی و همکاران، ۲۰۱۷: ۱۳). همبستگی کیفی - کمی سطح زیر منحنی و ارزیابی تخمین بدین صورت است: (۱ - ۰/۹، عالی؛ ۰/۸ - ۰/۹، خیلی خوب؛ ۰/۷ - ۰/۸، خوب؛ ۰/۶ - ۰/۷، متوسط؛ و ۰/۵ - ۰/۶، ضعیف) (یاسلینسار^۱، ۲۰۰۵: ۴۲۳). در پژوهش حاضر مجموعه لغزش‌های رخ داده (۶۰ نقطه) در منطقه مطالعاتی به دو گروه تقسیم شدند، که ۴۰ (۷۰٪) نقطه لغزشی به صورت تصادفی از مجموع ۶۰ نقطه لغزشی به عنوان داده‌های آموزشی و باقی‌مانده ۲۰ (۳۰٪) نقطه لغزشی جهت آزمون انتخاب شدند. سپس مدل ارزیابی حساسیت زمین‌لغزش با استفاده از مدل جدید ترکیبی RFFT با استفاده از مجموعه داده‌های آموزشی و صحت‌سنجی ترسیم شد که (شکل ۳) نتایج ارزیابی صحت نقشه‌های حساسیت زمین‌لغزش‌های منطقه مورد مطالعه را با داده‌های تعلیمی و صحت‌سنجی نشان می‌دهد. مشاهده این شکل نشان می‌دهد که در منحنی SRC بر اساس داده‌های تعلیمی مقدار سطح زیر منحنی با استفاده از الگوریتم FT، ۰/۷۸۲ می‌باشد که به این معنی است این الگوریتم ۷۸/۲ درصد توانایی پیش‌بینی مناطق حساس به وقوع زمین‌لغزش را داشته است. در حالی که مدل ترکیبی RFFT برای همین نوع داده با دارا شدن مقدار ۰/۸۳۱ مساحت زیر این منحنی، ۸۳/۱ درصد توانایی در پیش‌بینی مناطق حساس را داشته است. این در حالی است که برای داده‌های صحت‌سنجی مقدار سطح زیر منحنی PRC به ترتیب برای الگوریتم FT و مدل ترکیبی RFFT، ۰/۷۶۵ و ۰/۷۸۵ به دست آمد. به طور کلی نتایج صحت‌سنجی نقشه‌های به دست آمده با هر دو روش تأیید شد؛ با این توضیح که مدل ترکیبی ارائه شده منجر به افزایش مقدار درصد مساحت زیر منحنی هم برای داده‌های تعلیمی و هم داده‌های آموزشی شده است.



شکل ۳: ارزیابی نقشه‌های پیش‌بینی مناطق حساس به وقوع زمین لغزش‌ها بر اساس الگوریتم FT و مدل ترکیبی RFFT با داده‌های تعلیمی (سمت راست) و با داده‌های صحت‌سنجی (سمت چپ) (منبع: نگارندگان)

همچنین با استفاده از تجزیه و تحلیل منحنی ROC می‌توان مشاهده کرد که مدل RFFT دارای AUC بالا برای مجموعه داده‌های آموزشی (۰/۸۹۱) و آزمون (۰/۷۹۲) دارد که این مقادیر بالاتر از مدل FT (۰/۸۱۹) برای مجموعه داده‌های آموزشی و (۰/۷۸۹) برای آزمون داده‌ها) می‌باشد (شکل ۴). لذا این نتایج بیانگر آن است که مدل ترکیبی جدید RFFT برآورد موفقیت‌آمیز و عملکرد خوبی در پیش‌بینی حساسیت زمین لغزش در منطقه مطالعاتی دارد.



شکل ۴: عملکرد مدل و اعتبارسنجی با استفاده از منحنی ROC و تجزیه و تحلیل مقدار AUC با استفاده از مدل‌های FT و RFFT با داده‌های تعلیمی (الف) و با داده‌های صحت‌سنجی (ب) (منبع: نگارندگان)

نتیجه‌گیری

در پژوهش حاضر یک رویکرد جدید مبتنی بر RF و FT که ترکیبی از دو روش پیشرفته گروه یادگیری ماشین (RF) و طبقه‌بندی (FT) است جهت پیش‌بینی حساسیت زمین‌لغزش‌های شهرستان کامیاران مورد ارزیابی قرار گرفت. در منطقه مورد مطالعه ۶۰ نقطه لغزشی شناسایی و ۲۱ عامل مؤثر شامل: درجه شیب، جهت شیب، ارتفاع به متر، انحنای شیب، انحنای عرضی شیب، انحنای طولی شیب، تابش خورشید، عمق دره، شاخص قدرت جریان، شاخص نمناکی توپوگرافی، شاخص طول دامنه، کاربری اراضی، تراکم پوشش گیاهی، فاصله از گسل، تراکم گسل، فاصله از جاده، تراکم جاده، فاصله از آبراهه، تراکم آبراهه، همباران و لیتولوژی به کار گرفته شد. نتایج بررسی مهمترین عوامل اثرگذار از میان بیست و یک عامل مؤثر شناخته شده بر وقوع زمین‌لغزش‌های سطحی منطقه مورد مطالعه بر اساس شاخص Information Gain Ratio (IGR) نشان داد که هشت متغیر (فاصله از آبراهه، تراکم آبراهه، شاخص قدرت حمل جریان، بارندگی، عمق دره، شاخص نمناکی توپوگرافی، شدت تابش خورشیدی و انحنای طولی شیب) شناخته شدند. سپس برای بدست آوردن مقادیر بهینه برای مدل‌سازی زمین‌لغزش، از دو پارامتر تعداد سیدها و تعداد تکرارها استفاده شد. نتایج نشان داد که مقادیر بهینه برای تعداد تکرارها و تعداد دانه‌ها به ترتیب ۱۵ و ۳ بود. نقشه‌های حساسیت تهیه شده توسط مدل‌های RF و RFFFT به پنج کلاس حساسیت: بسیار کم، کم، متوسط، زیاد و بسیار زیاد تقسیم شدند. علاوه بر این، به منظور ارزیابی مدل‌های ارائه شده از معیارهای کاپا، Sensitivity, Specificity, Accuracy, RMSE و مساحت زیر منحنی (ROC) استفاده شدند. مساحت زیر منحنی، نرخ کمی از عملکرد نقشه‌های حساسیت را نشان می‌دهد. ارزیابی نتایج نشان داد که مدل RFFFT با نرخ موفقیت ۸۳/۱٪ و دقت پیش‌بینی ۷۸/۲٪ بهتر از مدل FT با نرخ موفقیت ۷۸/۵٪ و دقت پیش‌بینی ۷۶/۵٪ می‌باشد. لذا نتایج بیانگر آن است که مدل پیشنهادی RFFFT از عملکرد بسیار خوب و قابلیت پیش‌بینی بهتری جهت ارزیابی حساسیت زمین‌لغزش در منطقه مطالعاتی دارد. بنابراین مدل پیشنهادی RFFFT به عنوان یک روش جایگزین مناسب برای ارزیابی حساسیت زمین‌لغزش در مناطق مستعد لغزشی استفاده می‌شود. نتایج این تحقیق هم‌چنان نشان می‌دهد که مدل ترکیبی با استفاده از تکنیک‌های گروهی مانند RF، یک رویکرد امیدوار کننده برای مدل‌سازی حساسیت زمین‌لغزش است. چرا که نقشه‌ای که دارای صحت بالا و معقول در شناسایی و تفکیک مناطق حساس به وقوع زمین‌لغزش باشد، به مهندسین جهت معرفی مناطق با حساسیت‌های مختلف وقوع زمین‌لغزش به منظور احداث مکان مناسب برای جلوگیری از تخریب سازه‌های رسوبگیر و اصلاحی، مدیریت دامنه‌ها، زهکشی و انتقال آب از مناطق حساس و نزدیک به اجرای سازه، توسعه بهینه شبکه جاده‌های شهری و روستایی، استفاده در برنامه ریزی شهری و مسائل آمایش سرزمین و غیره کمک شایانی می‌نماید.

منابع

- پورقاسمی، حمیدرضا، ۱۳۸۹، پهنه‌بندی حساسیت زمین‌لغزش با استفاده از مدل احتمالاتی وزن واقعه، مهندسی فناوری اطلاعات مکانی، سال یکم، شماره نهم، ص ۸۰ - ۶۹
- شیرزادی عطاله، سلیمانی کریم، حبیب‌نژاد محمود، کاویان عطاله، چپی کامران، ۱۳۹۶، معرفی یک مدل جدید ترکیبی الگوریتم مینا به منظور پیش‌بینی حساسیت زمین‌لغزش‌های سطحی اطراف شهر بیجار، جغرافیا و توسعه، شماره ۴۶، صفحات ۲۴۶ - ۲۲۵.

- طالبی علی، گودرزی سحر، پورقاسمی حمید رضا، ۱۳۹۶، بررسی امکان تهیه نقشه خطر زمین لغزش با استفاده از الگوریتم جنگل تصادفی (محدوده ی مورد مطالعه: حوزه آبخیز سردارآباد، استان لرستان)، مجله مخاطرات محیط طبیعی.
- مرادی حمیدرضا، محمدی مجید، پورقاسمی حمید.رضا، ۱۳۹۱، حرکات دامنه‌ای با تأکید بر روشهای کمی تحلیل وقوع زمین لغزش. انتشارات سمت، ص ۲۰۹.
- *Abedini M, Ghasemyan B, Rezaei Mogaddam M H, 2017, Landslide susceptibility mapping in Bijar city, Kurdistan Province, Iran: a comparative study by logistic regression And AHP models, Environ Earth Sci, 76:308, DOI 10.1007/s12665-017-6502-3*
- *Benda, L., Dunne, T. 1997. Stochastic forcing of sediment supply to channel networks from landsliding and debris flow, Water Resources Research, 33: 2849–2863.*
- *Chen W, Xie X, Wang J, Pradhan B, Hong H, Bui DT, Duan Z, 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. 151:147–160.*
- *Dahal, R.K., Hasegava, Sh., Nonoura, A., Yamanka, M., Dhakal, S., Pauudyal, P., 2008. Predictive Modeling of Rainfall-Induced Landslide Hazard in the Lesser Himalaya of Nepal Based on Weights of Evidence, Geomorphology, Vol.102, NO.3-4, and PP: 496-510.*
- *Gama, J.: Functional trees for classification. In: Proceedings IEEE International Conference on Data Mining, 2001, ICDM 2001, pp. 147–154.*
- *Kavzoglu, T., Colkesen, I.: 2013, an assessment of the effectiveness of a rotation forest ensemble for land-use and land-cover mapping. Int. J. Remote Sens. 34, 4224–4241*
- *Lan H, Frank E, Hall M, 2011, Data mining: Practical machine learning tools and techniques. Morgan Kaufman, Boston.*
- *Omar F. Althuwaynee I Biswajeet Pradhan I Hyuck-Jin Park I Jung Hyun Lee. ,2014, A novel ensemble decision tree-based CHi-squared Automatic Interaction Detection (CHAID).*
- *Ozcift, A., Gulten, A.: 2012, robust multi-class feature selection strategy based on rotation forest ensemble algorithm for diagnosis of Erythematous-Squamous diseases. J. Med. Syst. 36, 941–949.*
- *Ozcift, A., Gulten, A.: 2011, Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms. Comput. Methods Programs Biomed. 104, 443–451.*
- *Pham, B.T., Bui, D.T., Dholakia, M.B., Prakash, I., Pham, H.V., Mehmood, K., Le, H.Q. 2016, A novel ensemble classifier of rotation forest and Naïve Bayer for landslide susceptibility assessment at the Luc Yen district, Yen Bai Province (Viet Nam) using GIS. Geomat. Nat. Hazards Risk, 1–23.*
- *Pradhan Biswajeet ,2013, a comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS, Computers & Geosciences 51, PP: 350–365.*
- *Quinlan JR, 1993, C4.5: programs for machine learning. Morgan Kaufmann, San Mateo, CA, USA.*
- *Rodriguez JJ, Kuncheva LI, Alonso CJ. 2006. Rotation forest: a new classifier ensemble method. Pattern Anal Mach Intell IEEE Trans. 28:1619–1630.*
- *Shirzadi Ataollah, Dieu Tien Bui, Binh Thai Pham, Karim Solaimani, Kamran Chapi, Ataollah Kavian, Himan Shahabi, 2017. Shallow landslide susceptibility assessment*

using a novel hybrid intelligence approach, Environmental Earth Sciences 76:60. Pp 1-18.

- *Tsangaratos P, Ilia I, 2015, Landslide susceptibility mapping using a modified decision tree classifier in the Xanthi Prefecture, Greece Landslides: 1-16 doi: 10.1007/s10346-015-0565-6.*
- *Yesilnacar E, Topal T, 2005, Landslide susceptibility mapping: comparison of logistic regression and neural networks in a medium scale study, Hendek region (Turkey) Engineering Geology 79:251-266.*
-