

(OPEN ACCESS)

Comparison and integration of machine learning and object-based algorithms for screening underlying factors and preparing a landslide classification map

Marzieh Nikjoui¹, Ali Najafinejad^{*2}, Hamid Reza Pourghasemi³,
Chooghi Bairam Komaki⁴

1. Ph.D. Student in Watershed Management, Faculty of Rangeland and Watershed Management, Gorgan University of Agricultural Sciences and Natural Resources, Gorgan, Iran. E-mail: nikjoi.marzi@gmail.com
2. Corresponding Author, Professor, Dept. of Watershed Management, Faculty of Rangeland and Watershed Management, Gorgan University of Agricultural Sciences and Natural Resources, Gorgan, Iran. E-mail: najafinejad@gau.ac.ir
3. Professor, Dept. of Soil Science, College of Agriculture, Shiraz University, Shiraz, Iran. E-mail: hm_porghasemi@yahoo.com
4. Associate Prof., Dept. of Arid Regions Management, Faculty of Rangeland and Watershed Management, Gorgan University of Agricultural Sciences and Natural Resources, Gorgan, Iran. E-mail: komaki@gau.ac.ir

Article Info	ABSTRACT
Article type: Research Full Paper	Background and Objectives: Landslides, among the most destructive natural hazards following earthquakes, cause irreversible damage to both the environment and infrastructure. Due to the unique geological and climatic conditions, Iran experiences numerous landslides annually, necessitating the needs for careful studies and preventive measures. This study presents an integrated approach that leverages the capabilities of machine learning algorithms to identify effective features for landslide detection. Furthermore, it compares object-based algorithms to generate a landslide classification map utilizing Gaofen-1 satellite images. Eventually, the study includes partial dependence plot illustrating the relationship between landslides and various independent conditional factors.
Article history: Received: 09.22.2024 Revised: 10.24.2024 Accepted: 10.28.2024	
Keywords: Basic object, Data mining, Landslide, Random forest, Support vector system	
	Materials and Methods: To identify landslides in the Mohammadabad watershed of Golestan, two Gaofen-1 satellite images from March 2023 and June 2024 were employed. Due to seasonal variations between the images, all processing was conducted separately. In the first phase, 218 landslides were identified through field visits, with 70% of these used for model training and the remaining 30% reserved for evaluating the results. The classification of satellite images and the extraction of landslide and non-landslide classes were performed using basic object-oriented classification approach involving two stages: segmentation and classification. Following image segmentation with a multi-scale segmentation algorithm, feature selection was conducted using three algorithms: neural networks, decision trees, and random forests. Factors influencing landslide identification were extracted from each segment based on the satellite images, and collinearity among features was assessed. Subsequently, these features were employed in the classification and identification of landslides using four object-based algorithms: support vector machine, nearest neighbor, decision tree, and random forest. The performance of these algorithms was compared using overall accuracy indices, Kappa coefficient, Sorensen coefficient, true positive rate, and false positive rate. Ultimately, the most effective algorithm for landslide detection using satellite imagery was determined.

Results: The results of the feature selection analysis indicated that out of the three methods examined in this study, the random forest algorithm identified the most effective features for landslide detection using satellite images. The classification of landslides utilizing four object-oriented algorithms, support vector machine, decision tree, random forest, and nearest neighbor, revealed that the support vector machine algorithm achieved the highest accuracy of 92% and a Kappa coefficient of 0.85. This performance showed its superior capability in identifying landslides within the studied area compared to the other algorithms.

Conclusion: This study showed that the integrating of machine learning algorithms with object-based methods provides a reliable and cost-effective approach for the rapid identification of landslides using satellite images. Identifying landslides is an important step in studying this natural hazard, as the findings can offer valuable insights to managers and practitioners to enhance planning and management strategies to mitigate landslide-induced damages. For future study, utilizing high-resolution images is recommended to enable more detailed landslide identification.

Cite this article: Nikjou, Marzieh, Najafinejad, Ali, Pourghasemi, Hamid Reza, Komaki, Chooghi Bairam. 2025. Comparison and integration of machine learning and object-based algorithms for screening underlying factors and preparing a landslide classification map. *Journal of Water and Soil Conservation*, 32 (1), 31-55.



© The Author(s).

DOI: 10.22069/jwsc.2025.22812.3759

Publisher: Gorgan University of Agricultural Sciences and Natural Resources

مقایسه و تلفیق الگوریتم‌های یادگیری ماشین و شیء پایه برای غربالگری عوامل زمینه‌ساز و تهیه نقشه طبقه‌بندی زمین لغزش

مرضیه نیکجوی^۱، علی نجفی‌نژاد^{۲*}، حمیدرضا پورقاسمی^۳، چوقی بایرام کمکی^۴

۱. دانشجوی دکتری مدیریت حوزه‌های آبخیز، دانشکده مرتع و آبخیزداری، دانشگاه علوم کشاورزی و منابع طبیعی گرگان، گرگان، ایران. رایانامه: nikjoi.marzi@gmail.com
۲. نویسنده مسئول، استاد گروه آبخیزداری، دانشکده مرتع و آبخیزداری، دانشگاه علوم کشاورزی و منابع طبیعی گرگان، گرگان، ایران. رایانامه: najafinejad@gu.ac.ir
۳. استاد بخش علوم خاک، دانشکده کشاورزی، دانشگاه شیراز، شیراز، ایران. رایانامه: hm_porghasemi@yahoo.com
۴. دانشیار گروه مدیریت مناطق بیابانی، دانشکده مرتع و آبخیزداری، دانشگاه علوم کشاورزی و منابع طبیعی گرگان، گرگان، ایران. رایانامه: komaki@gu.ac.ir

اطلاعات مقاله	چکیده
نوع مقاله: مقاله کامل علمی- پژوهشی تاریخ دریافت: ۱۳/۰۷/۰۱ تاریخ ویرایش: ۱۳/۰۸/۰۳ تاریخ پذیرش: ۱۳/۰۸/۰۷	سابقه و هدف: زمین‌لغزش به‌عنوان یکی از مخرب‌ترین مخاطرات طبیعی پس از زلزله، خسارات جبران‌ناپذیری به محیط‌زیست و زیرساخت‌ها وارد می‌کند. ایران به‌دلیل شرایط زمین‌شناختی و اقلیمی خاص خود، هر ساله شاهد وقوع زمین‌لغزش‌های متعددی است که نیازمند مطالعات دقیق و اقدامات پیشگیرانه است. در مطالعه حاضر یک روش تلفیقی با استفاده از قابلیت و توانایی الگوریتم‌های یادگیری ماشین برای انتخاب ویژگی‌های مؤثر در شناسایی زمین‌لغزش و مقایسه الگوریتم‌های شیء پایه جهت تهیه نقشه طبقه‌بندی زمین‌لغزش با استفاده از تصاویر ماهواره‌ای گائوفن-۱ توسعه داده شد. در نهایت نمودار وابستگی جزئی زمین‌لغزش‌ها با هر یک از متغیرهای مستقل استخراج‌شده ترسیم گردید.
واژه‌های کلیدی: جنگل تصادفی، داده‌کاوی، زمین‌لغزش، شیء پایه، ماشین بردار پشتیبان	مواد و روش‌ها: به‌منظور شناسایی زمین‌لغزش‌های حوزه آبخیز محمدآباد گلستان از دو تصویر ماهواره گائوفن-۱ مربوط به اسفند ۱۴۰۱ و خرداد ۱۴۰۲ استفاده شد. به‌علت متفاوت بودن فصول دو تصویر، تمام پردازش‌های دو تصویر به‌صورت مجزا صورت گرفت. در اولین مرحله ۲۱۸ زمین‌لغزش با استفاده از بازدید میدانی شناسایی گردید که ۷۰ درصد آن برای آموزش مدل و ۳۰ درصد باقی‌مانده برای ارزیابی نتایج استفاده شد. طبقه‌بندی تصاویر ماهواره‌ای و استخراج کلاس‌های لغزش و فاقد لغزش با استفاده از طبقه‌بندی شیء پایه انجام شد؛ که شامل دو مرحله قطعه‌بندی و طبقه‌بندی است. بعد از انجام مرحله قطعه‌بندی تصویر با الگوریتم قطعه‌بندی چندمقیاسه، انتخاب ویژگی با استفاده از سه الگوریتم شبکه عصبی، درخت تصمیم و جنگل

تصادفی انجام شد و عوامل مؤثر بر شناسایی زمین لغزش با استفاده از تصاویر ماهواره‌ای از هر کدام از قطعات استخراج گردید سپس ویژگی‌های انتخاب شده براساس ویژگی هم‌خطی بررسی شدند. در مرحله بعد، این ویژگی‌ها در طبقه‌بندی و شناسایی زمین لغزش با چهار الگوریتم شیء پایه ماشین بردار پشتیبان، نزدیک‌ترین همسایه، درخت تصمیم و جنگل تصادفی مورد استفاده قرار گرفتند سپس عملکرد الگوریتم‌ها با شاخص‌های صحت کلی، ضریب کاپا، ضریب سورنسن، نرخ مثبت صحیح و نرخ مثبت کاذب مقایسه شد و مناسب‌ترین الگوریتم برای شناسایی زمین لغزش با استفاده از تصویر ماهواره‌ای انتخاب شد.

یافته‌ها: نتایج انتخاب ویژگی نشان داد از بین سه روش مورد بررسی در این مطالعه، الگوریتم جنگل تصادفی با دقت بالاتری ویژگی‌های مؤثر در شناسایی زمین لغزش با استفاده از تصاویر ماهواره‌ای را تعیین کرد. بررسی طبقه‌بندی زمین لغزش‌ها با چهار الگوریتم شیء گرا ماشین بردار پشتیبان، درخت تصمیم، جنگل تصادفی و نزدیک‌ترین همسایه نشان داد، الگوریتم ماشین بردار پشتیبان با صحت بالای ۹۲ درصد و ضریب کاپا بالای ۰/۸۵ به نسبت سایر الگوریتم‌ها شناسایی زمین لغزش‌های منطقه مورد مطالعه را به‌خوبی انجام داده است.

نتیجه‌گیری: مطالعه حاضر نشان داد که استفاده از قابلیت و توانایی الگوریتم‌های یادگیری ماشین و تلفیق آن با الگوریتم‌های شیء پایه می‌تواند به‌عنوان روش قابل اعتماد برای شناسایی سریع و ارزان زمین لغزش‌ها با استفاده از تصاویر ماهواره‌ای مورد استفاده قرار گیرد. شناسایی زمین لغزش‌ها اولین گام مطالعه این مخاطره طبیعی است به‌طوری‌که نتایج آن می‌تواند در بهبود برنامه‌ریزی، مدیریت و کاهش خسارات ناشی از زمین لغزش به مدیران امر و کارشناسان اجرایی کشور کمک کنند. به‌منظور انجام مطالعات آتی پیشنهاد می‌گردد که تصاویر با قدرت تفکیک مکانی بالا تهیه شود تا با جزئیات بالا بتوان به شناسایی زمین لغزش‌ها اقدام نمود.

استناد: نیکجوی، مرضیه، نجفی‌نژاد، علی، پورقاسمی، حمیدرضا، کمکی، چوقی بایرام (۱۴۰۴). مقایسه و تلفیق الگوریتم‌های یادگیری ماشین و شیء پایه برای غربالگری عوامل زمینه‌ساز و تهیه نقشه طبقه‌بندی زمین لغزش. پژوهش‌های حفاظت آب و خاک، ۳۲ (۱)، ۵۵-۳۱.

DOI: 10.22069/jwsc.2025.22812.3759



© نویسندگان.

ناشر: دانشگاه علوم کشاورزی و منابع طبیعی گرگان

مقدمه

براساس آمار جهانی، زمین‌لغزش بعد از زلزله بیش‌ترین خسارت و آسیب را به محیط‌زیست، زیرساخت‌ها و میراث فرهنگی و تاریخی وارد می‌کند (۱). کشور ایران نیز به‌علت شرایط کوهستانی، شرایط اقلیمی، زمین‌شناختی و لرزه‌خیزی خاص هر ساله شاهد وقوع زمین‌لغزش‌های فراوانی است و خسارات مالی و جانی متعددی به بار می‌آورد (۲). هم‌چنین زمین‌لغزش اثرات مخرب زیادی روی محیط‌زیست می‌گذارد که باعث از بین رفتن حجم قابل‌توجهی از خاک و پوشش گیاهی می‌شود، به‌طوری‌که از بین رفتن این منابع طبیعی غیرقابل‌برگشت بوده و یا زمان بسیار زیادی را برای ترمیم نیاز دارد (۳). بنابراین لازم است مطالعات و اقدامات مؤثری در خصوص کنترل و پیشگیری از این بلایای طبیعی و گاهی با منشأ انسانی صورت گیرد. یکی از اولین اقدامات در این رابطه، تهیه نقشه محل وقوع زمین‌لغزش‌های اتفاق افتاده در گذشته است که می‌تواند به‌عنوان ابزاری قدرتمند در مدیریت و برنامه‌ریزی جهت جلوگیری از توسعه مناطق مستعد زمین‌لغزش و کاهش خسارات در آینده مورد استفاده مدیران و متولیان امر قرار گیرد (۴). شناسایی محل وقوع زمین‌لغزش‌ها در گذشته بیش‌تر از طریق بازدیدهای میدانی، تفسیر عکس‌های هوایی و اطلاعات ثبت‌شده رقومی و یا مکانی سازمانی موجود انجام می‌گرفته است که در خیلی از موارد تحت‌تأثیر زمان قرار می‌گرفتند و به‌دلیل حجم بالای اطلاعات، پردازش آن‌ها در زمان کوتاه مقدور نبوده و شاید کارایی خود را برای برنامه‌ریزی‌های مدیریتی از دست می‌داده است. اما اکنون با پیشرفت علم، روش‌های نوین مانند استفاده از تصاویر ماهواره‌ای و الگوریتم‌های یادگیری ماشین جایگزین این روش‌های کلاسیک شده‌اند. یکی از موارد بسیار مهم در شناسایی

زمین‌لغزش با استفاده از تصاویر ماهواره‌ای انتخاب مؤثرترین عامل از بین عوامل زمینه‌ساز و الگوریتم مناسب برای تهیه نقشه طبقه‌بندی زمین‌لغزش است. این امر به‌خصوص زمانی که یک مدیر با یک حجم داده زیاد و مجموعه‌ای از عوامل زمینه‌ساز محلی با روابط پیچیده روبه‌رو است دارای کاربرد فراوان است (۵). در سالیان اخیر مطالعات متعددی با استفاده از الگوریتم‌های یادگیری ماشین، روش‌های تصمیم‌گیری چندمعیاره و روش‌های آماری در داخل و خارج کشور انجام شده است. به‌عنوان مثال، مارتا و همکاران (۲۰۱۰) در مطالعه خود با استفاده از روش شیء‌پایه ویژگی‌های طیفی، مکانی و مورفومتریک زمین‌لغزش‌های منطقه هیمالیا در هندوستان را جهت تشخیص خودکار پدیده زمین‌لغزش با استفاده از داده‌های چندطیفی ۵/۸ متر Resourcesat-1 استفاده نمودند. براساس نتایج درصد تشخیص زمین‌لغزش‌ها با این روش ۷۶/۴ درصد و صحت طبقه‌بندی ۶۹/۱ درصد بوده است (۶). بلاشکا و همکاران (۲۰۱۴) به بررسی زمین‌لغزش‌های حوزه دریاچه ارومیه با استفاده از روش تجزیه و تحلیل تصویر مبتنی بر روش شیء‌پایه پرداختند. در این پژوهش، از ترکیب شاخص‌های NDVI، روشنایی و ویژگی‌های بافتی مشتق شده از تصاویر ماهواره‌ای (IRS-ID و SPOT-5) در ترکیب با عوامل شیب و جهت جریان برای شناسایی و ترسیم زمین‌لغزش استفاده شد. جهت اعتبارسنجی، نقشه تهیه‌شده با داده‌های ثبت‌شده زمینی مقایسه و دقت کلی آن ۹۳/۰۷ درصد به‌دست آمد (۷). عابدینی و همکاران (۲۰۱۶) زمین‌لغزش‌های منطقه نصیرآباد تا سد ستارخان اهر را با استفاده از ویژگی‌های طیفی و مکانی و روش شیء‌گرا شناسایی و طبقه‌بندی نمودند. آن‌ها ۷۰ درصد نمونه‌های زمین‌لغزش را برای آموزش مدل و ۳۰ درصد را برای آزمون مدل استفاده کردند.

براساس نتایج آن‌ها دقت شناسایی زمین‌لغزش‌ها ۷۱/۱۱ درصد و دقت کلاس‌بندی ۹۱/۴ درصد بوده است (۸). فتحی و همکاران (۲۰۱۹) مناطق مستعد زمین‌لغزش در کرانه جنوبی حوزه آبریز اهرچای از روستای نصیرآباد تا سد ستارخان را با استفاده از تئوری احتمالات شرطی مکان‌یابی و پهنه‌بندی کردند. آن‌ها ارتباط بین مناطق لغزشی و پارامترهای مؤثر را با احتمالات تئوری بیزی تعیین و وزن هر طبقه از پارامترها را به‌دست آوردند (۹). قنوتی و همکاران (۲۰۱۹) به مطالعه حساسیت زمین‌لغزش آبخیز کن در استان تهران با استفاده از الگوریتم یادگیری ماشین پرداختند. در این پژوهش، هشت عامل مؤثر بر وقوع زمین‌لغزش انتخاب و براساس میزان تأثیر بر وقوع زمین‌لغزش طبق نظر کارشناسی امتیازدهی شد. براساس نتایج اعتبارسنجی، صحت روش ماشین بردار پشتیبان با $AUC=0/91$ از مدل شاخص آنتروپی با $AUC=0/86$ بالاتر بوده است (۱۰). کرنزادی و پورقاسمی (۲۰۱۹) ارزیابی حساسیت زمین‌لغزش حوزه آبخیز چهل‌چای را با استفاده از مدل‌های داده‌کاوی انجام دادند. آن‌ها برای این منظور از ۱۲ عامل مؤثر بر وقوع زمین‌لغزش و دو الگوریتم ماشین بردار پشتیبان و درخت رگرسیون تقویت شده استفاده کردند. براساس نتایج پژوهش آن‌ها الگوریتم ماشین بردار پشتیبان با صحت ۰/۸۲ به‌عنوان روش بهتر انتخاب شد (۱۱). قاسمیان و همکاران (۲۰۲۱) با ترکیب الگوریتم مدل درخت عملکرد (FT) و مدل جنگل چرخشی (RF) نقشه حساسیت زمین‌لغزش در شهرستان کامیاران، استان کردستان را تهیه نمودند. مدل‌سازی مذکور با انتخاب ۱۷ عامل مؤثر از بین ۲۱ عامل پیشنهادی با الگوریتم شاخص نسبت اطلاعات

انجام شد. نتایج آن‌ها نشان از دقت بالای روش پیشنهادی با $AUC=0/89$ بوده است (۱۲). آمتیا و همکاران (۲۰۲۱) یک سیستم شناسایی نیمه‌خودکار زمین‌لغزش (SALaD) مبتنی بر تحلیل شیء‌پایه و یادگیری ماشین را معرفی کردند که قادر به شناسایی ۷۰ درصد از زمین‌لغزش‌های منطقه بود؛ بنابراین به‌عنوان یک جایگزین سریع و کارآمد به‌جای روش‌های سنتی نقشه‌برداری پیشنهاد شد (۱۳). گلی‌مختاری و ناعمی‌تبار (۲۰۲۲) پیش‌بینی و مدل‌سازی مکانی زمین‌لغزش شهرستان کلات را با استفاده از سه الگوریتم داده‌کاوی تابع شواهد قطعی، شواهد وزنی و ماشین بردار پشتیبان انجام دادند. برای این منظور از ۳۶ مورد زمین‌لغزش که با استفاده از تصاویر ماهواره لندست و گوگل ارث شناسایی شد و ۱۷ عامل مؤثر بر وقوع زمین‌لغزش استفاده کردند. براساس نتایج پژوهش آن‌ها، متغیرهای ارتفاع، زمین‌شناسی، خاک‌شناسی و کاربری اراضی به‌عنوان مهم‌ترین عوامل مؤثر در وقوع زمین‌لغزش شناسایی شدند. هم‌چنین تحلیلی نتایج با استفاده از AUC نشان داد الگوریتم ماشین بردار پشتیبان با AUC برابر ۰/۸۹ بهترین عملکرد را داشته است (۱۴). سان و همکاران (۲۰۲۳) ارزیابی حساسیت زمین‌لغزش در امتداد بزرگراه‌های کوهستانی شهرستان چنکو در چین را بر اساس الگوریتم‌های مختلف یادگیری ماشین و واحدهای نقشه‌برداری با غربالگری عوامل ترکیبی انجام دادند. ابتدا براساس تحقیقات میدانی، تصاویر ماهواره‌ای و سوابق تاریخی زمین‌لغزش‌های منطقه را شناسایی کردند. آن‌ها ۲۰ عامل مؤثر بر وقوع زمین‌لغزش را شناسایی و با روش آشکارساز جغرافیایی، عواملی کم‌اهمیت‌تر را حذف کردند. نتایج نشان داد پس از حذف عوامل کم‌اهمیت مدل آموزش داده‌شده توسط

- 1- Area Under the Curve
- 2- Functional Trees
- 3- Rotation Forest

مواد و روش‌ها

معرفی منطقه مورد مطالعه: منطقه مورد مطالعه در این پژوهش، حوزه آبخیز محمدآباد است که یکی از زیرحوزه‌های گرگان‌رود در استان گلستان به‌شمار می‌آید. این حوزه در محدوده طول جغرافیایی $42^{\circ} 54'$ تا $55^{\circ} 00'$ شرقی و عرض جغرافیایی $39^{\circ} 36'$ تا $50^{\circ} 36'$ شمالی واقع شده است (شکل ۱). حدود ۸۰ درصد حوزه مذکور پوشش جنگلی متشکل از گونه‌های خزان‌کننده پهن‌برگ است. وسعت کل حوزه در حدود $444/3$ کیلومتر مربع بوده که ۵۵ درصد آن را سازندهای با حساسیت زیاد و خیلی زیاد به فرسایش تشکیل می‌دهد. این حوزه شامل رودخانه‌ای به طول $30/5$ کیلومتر است. متوسط ارتفاع از سطح دریا 1890 متر می‌باشد. شیب متوسط حوزه $55/2$ درصد است که جهت غالب آن نیز، شمال غربی بوده که در مجموع $18/6$ درصد از کل مساحت حوزه را به خود اختصاص می‌دهد. متوسط بارش سالانه در این منطقه 530 میلی‌متر است که بیش‌ترین حجم بارش در فصل زمستان و کم‌ترین آن در تابستان اتفاق می‌افتد (۱۶).

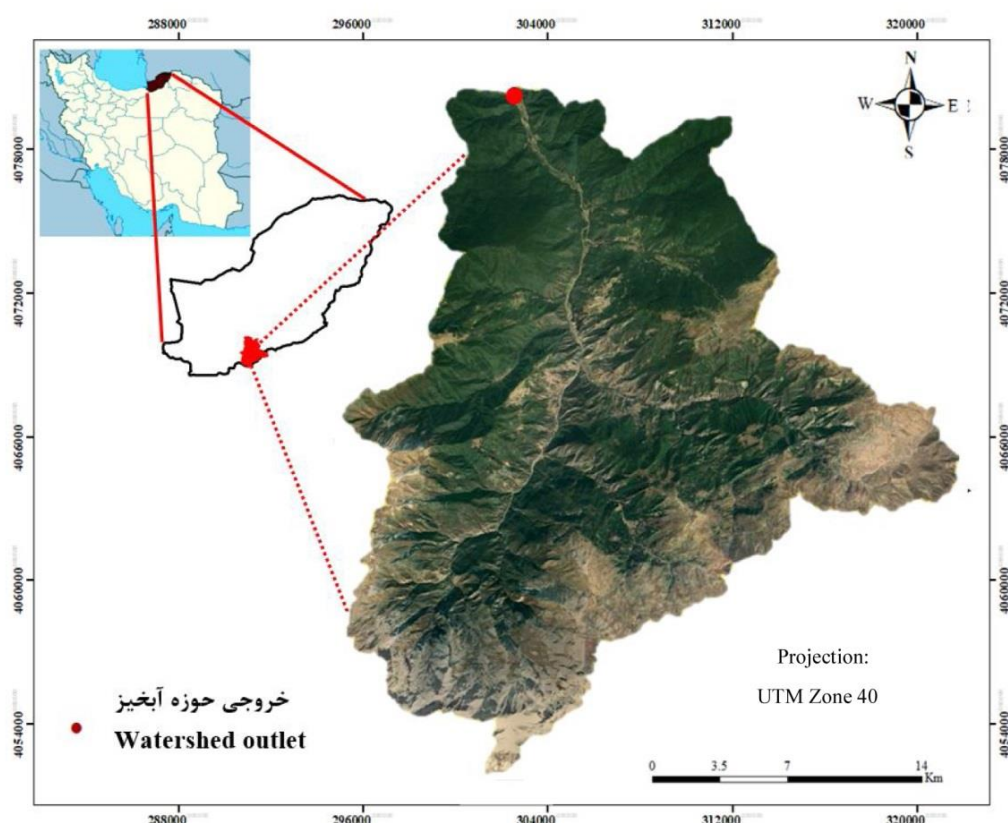
روش کار: شکل ۲ مراحل کار این پژوهش را نشان می‌دهد. به‌منظور مطالعه زمین‌لغزش‌های حوزه آبخیز مورد بررسی، از دو تصویر ماهواره گائوفن-۱ مربوط به اسفند ۱۴۰۱ و خرداد ۱۴۰۲ که از سازمان فضایی ایران تهیه گردید، استفاده شد. تصاویر موردنظر ابتدا به لحاظ تصحیحات هندسی مورد بررسی قرار گرفتند. به علت متفاوت بودن فصول دو تصویر تمام پردازش‌ها به‌صورت مجزا روی دو تصویر انجام شد.

۱۱ عامل مؤثر بر زمین‌لغزش دقت پیش‌بینی بهتری داشته است (۱۵).

مرور منابع انجام‌شده نشان می‌دهد، انتخاب عوامل مؤثر در اکثر مطالعات مربوط به زمین‌لغزش، به‌منظور تهیه نقشه حساسیت و یا خطر زمین‌لغزش بوده است اما در زمینه استفاده از عوامل زمینه‌ساز برای مرحله شناسایی که مرحله اول توسعه مدل‌های حساسیت و ریسک‌پذیری است، مطالعات اندکی انجام گرفته است. در این مطالعه این مهم به‌منظور کاهش ابعاد مسأله و در نظر گرفتن پارامترهای تأثیرگذار با استفاده از سه الگوریتم، شبکه عصبی، درخت تصمیم و جنگل تصادفی انجام شد. در مرحله بعد ویژگی‌های انتخاب‌شده از نظر هم‌خطی بررسی شدند، خروجی این مرحله در طبقه‌بندی زمین‌لغزش بر روی تصاویر ماهواره‌ای گائوفن، با استفاده از چهار الگوریتم شیء‌پایه ماشین بردار پشتیبان، نزدیک‌ترین همسایه، درخت تصمیم و جنگل تصادفی استفاده شد و عملکرد آن‌ها باهم مقایسه گردید سپس رابطه بین ویژگی‌های مؤثر در طبقه‌بندی با بهترین الگوریتم طبقه‌بندی بررسی گردید. مدل‌سازی انجام‌شده در حوزه آبخیز محمدآباد، در استان گلستان مورد آزمون واقع شده است. بررسی‌های میدانی انجام‌شده در حوزه مورد مطالعه نشان می‌دهد وقوع زمین‌لغزش در این منطقه دارای فراوانی زیاد و تکرار سالیانه است. با این حال، تاکنون هیچ مطالعه مدونی در خصوص شناسایی زمین‌لغزش‌های این منطقه صورت نگرفته است. حوزه مورد مطالعه شامل ۱۳ روستا است، هم‌چنین شهر فاضل‌آباد در خروجی آن واقع شده که اهمیت این حوزه را دوچندان می‌کند. بنابراین ضرورت تحقیق در زمینه شناسایی زمین‌لغزش‌های این منطقه به‌وضوح احساس می‌شود.

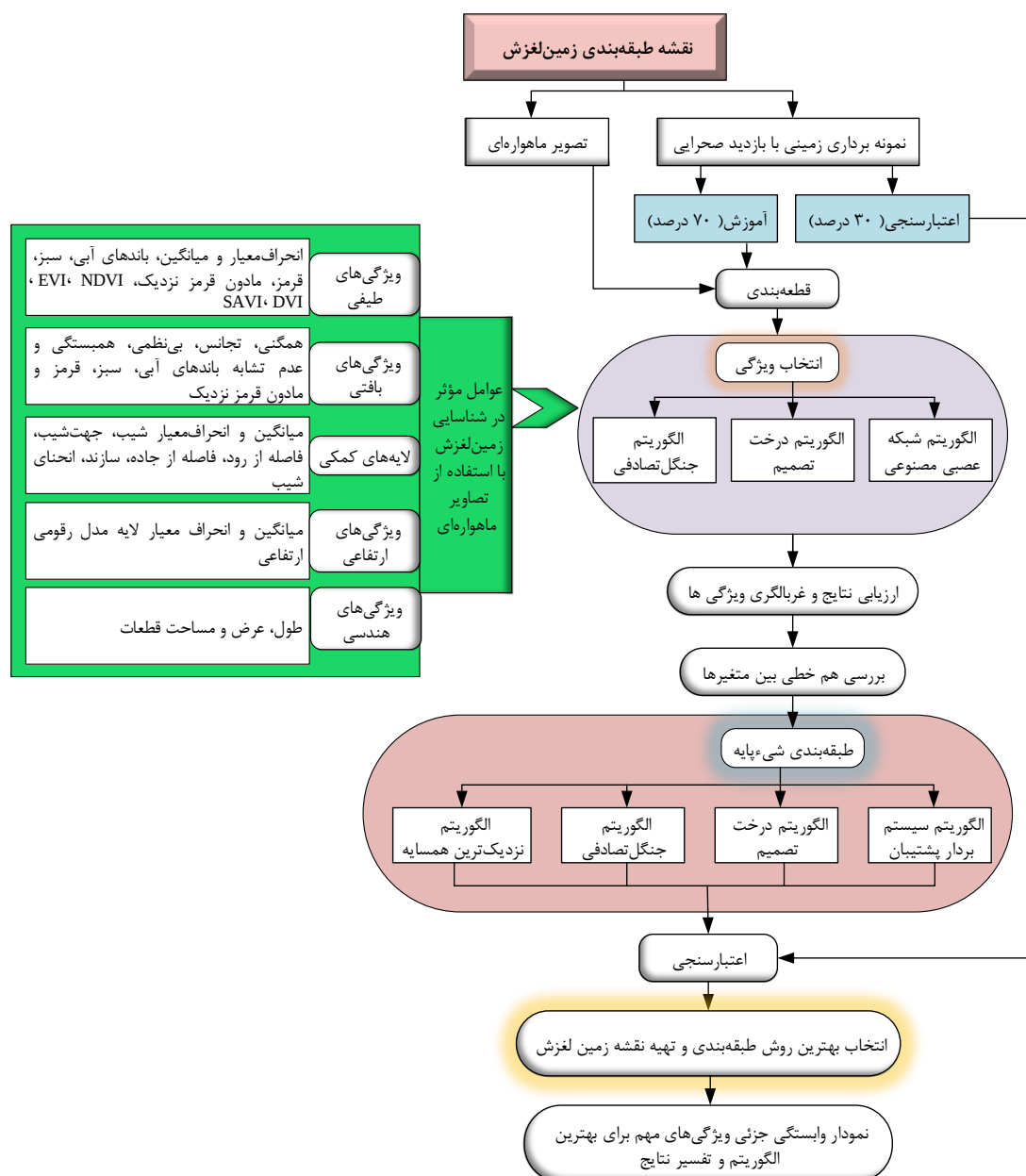
تصاویر مورد مطالعه جانمایی گردید. ۷۰ درصد زمین‌لغزش‌های شناسایی شده برای آموزش مدل و ۳۰ درصد برای ارزیابی نتایج به صورت تصادفی جدا شدند (۱۷). شکل ۳ چند مورد از زمین‌لغزش‌های منطقه مورد مطالعه را نشان می‌دهد.

مهم‌ترین داده در تهیه نقشه زمین‌لغزش، شناسایی زمین‌لغزش‌هایی موجود در منطقه است. بدین منظور طی بازدیدهای میدانی متعدد در فصول بهار و تابستان سال ۱۴۰۲، موقعیت ۲۱۸ واقعه زمین‌لغزش در منطقه مورد مطالعه با دستگاه سامانه موقعیت‌یاب جهانی و خطای ± 3 متر ثبت و موقعیت آن بر روی



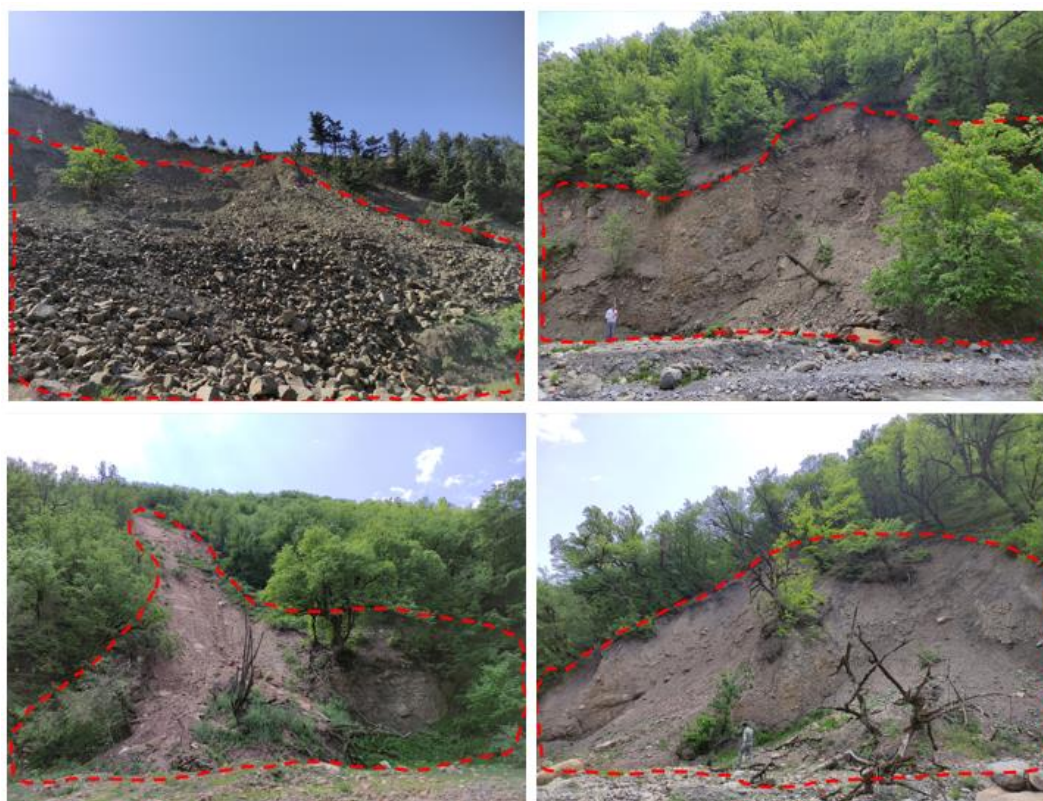
شکل ۱- حوزه آبخیز محمدآباد استان گلستان.

Figure 1. Mohammadabad watershed of Golestan province.



شکل ۲- مراحل اجرایی مطالعه حاضر.

Figure 2. Main steps of the present study.



شکل ۳- شمایی از زمین لغزش‌های شناسایی شده در بازدید میدانی.

Figure 3. Examples of landslides detected in the field survey.

با توجه به تعدد ویژگی قطعات مورد بررسی و حذف ویژگی‌های نامربوط (ویژگی‌های که بر خروجی نهایی تأثیری ندارند)، فرایند انتخاب ویژگی انجام شد (۲۱). هدف از انجام این پیش‌پردازش کمک به کاهش ابعاد مسئله، کاهش هزینه محاسباتی و افزایش دقت نقشه طبقه‌بندی بود. انتخاب زیرمجموعه‌ای از ویژگی‌های مفید از تعداد زیادی متغیر ممکن، به‌منظور شناسایی ورودی‌های کلیدی، کشف پاسخ مدل به یکسری ورودی‌های خاص و درنهایت افزایش کارایی الگوریتم‌های یادگیری ماشین انجام می‌شود و به‌عنوان یکی از حوضه‌های جدید در مبحث یادگیری ماشین مطرح است (۲۲). با وجود وسعت پژوهش در حوضه انتخاب ویژگی و توسعه روش‌های متعدد در دهه‌های گذشته، الگوریتم واحدی برای همه مطالعات وجود ندارد و بسته به شرایط مختلف هر مسأله،

به‌منظور طبقه‌بندی تصاویر ماهواره‌ای و استخراج کلاس‌های لغزش و فاقد لغزش از روش طبقه‌بندی شیء پایه استفاده شد (۱۸). طبقه‌بندی تصویر با الگوریتم طبقه‌بندی چندمقیاسه انجام شد. بر اساس مرور منابع و قابلیت داده‌های در دسترس در هریک از قطعات شناسایی‌شده بر روی تصاویر ماهواره‌ای، ویژگی‌های طیفی (انحراف معیار و میانگین باندهای اصلی و شاخص‌های گیاهی: ۱۶ ویژگی)، بافتی (همگنی، تجانس، بی‌نظمی، همبستگی و عدم تشابه باندهای تصویر: ۲۰ ویژگی)، ارتفاعی (انحراف معیار و میانگین لایه مدل رقومی ارتفاعی: ۲ ویژگی)، هندسی (طول، عرض و مساحت قطعه: ۳ ویژگی) و لایه‌های کمکی (میانگین و انحراف معیار شیب، جهت شیب، فاصله از رود، فاصله از جاده، سازند و انحنای شیب: ۱۲ ویژگی) استخراج شد (۱۳، ۱۹ و ۲۰).

الگوریتم‌های مختلف می‌توانند عملکرد متفاوتی را از خود بروز دهند. در مطالعه حاضر، به منظور انتخاب یک زیرمجموعه حداقلی از عوامل مؤثر با حفظ دقت و به حداکثر رساندن کارایی مدل در شناسایی زمین‌لغزش از سه الگوریتم جنگل تصادفی، شبکه عصبی و درخت تصمیم استفاده و نتایج آن‌ها باهم مقایسه شد. این فرایند در نرم‌افزار RStudio 2024.04.0 استفاده از بسته‌های نرم‌افزاری NeuralNetTools (۲۳)، psych، rpart، randomForest، caret انجام شد (۲۲، ۲۴ و ۴۸).

جنگل تصادفی: این الگوریتم یادگیری نظارت‌شده درخت پایه، شامل انبوهی از درخت‌های تصمیم کلاس‌بندی و رگرسیونی است و از میانگین‌گیری نتیجه همه درخت‌های تصمیم مرتبط ساخته می‌شود. این الگوریتم یکی از بهترین سیستم‌های یادگیری ماشین کنونی است که با استفاده از ایده کیسه‌گذاری و انتخاب تصادفی متغیرها به انجام مدل‌سازی می‌پردازد. مهم‌ترین ویژگی جنگل تصادفی عملکرد بالای آن‌ها در اندازه‌گیری اهمیت متغیرها برای مشخص کردن این‌که هر متغیر چه نقشی در پیش‌بینی پاسخ دارد، می‌باشد (۲۵). این الگوریتم درخت‌های طبقه‌بندی زیادی تولید می‌کند. هر درخت توسط یک نمونه Bootstrap متفاوت از روی داده اولیه ساخته می‌شود. این الگوریتم قابلیت مدل‌سازی روابط پیچیده غیرخطی از متغیرها را داراست و با تعیین اهمیت متغیرها، صحت نتایج حاصل بالاتر از الگوریتم‌های دیگر خواهد بود. این شیوه مدل‌سازی آماری در مطالعات مربوط به شناسایی زمین‌لغزش در مطالعات وورپاهل و همکاران (۲۰۱۲)، کیم و همکاران (۲۰۱۷) و چانگ و همکاران (۲۰۲۲) به طور گسترده مورد استفاده واقع شده است (۲۶، ۲۷ و ۲۸).

شبکه عصبی مصنوعی (ANN): شبکه عصبی مجموعه‌ای از گره‌ها است که عملکرد آن تا حد

زیادی شبیه سیستم عصبی و مغز موجودات زنده است (۲۹). این روش را می‌توان به عنوان شکل خاصی از رگرسیون غیرخطی تفسیر کرد که به طور خودکار تمام تعاملات ممکن بین پارامترهای مسئله را در یک ساختار شبکه‌ای متشکل از داده‌های ورودی، لایه‌های پنهان، ارتباطات و لایه‌های خروجی در نظر می‌گیرد، طوری که برتری فراوانی را از این نظر در مقایسه با روش‌های معمول مانند رگرسیون خطی ساده و مدل جمعی تعمیم‌یافته دارد (۲۶). شبکه‌های عصبی می‌توانند داده‌های حضور و عدم حضور را بهتر از درخت‌های تصمیم مدیریت کنند، اما نمی‌توانند مقادیر طبقه‌بندی را به خوبی روش درخت تصمیم مدیریت کنند (۳۰). این شیوه مدل‌سازی آماری در مطالعات مربوط به شناسایی زمین‌لغزش در مطالعات (سامن و همکاران (۲۰۲۰) و وورپاهل و همکاران (۲۰۱۲)) به طور گسترده مورد استفاده واقع شده است (۳۱ و ۲۶).

درخت تصمیم: یکی از روش‌های نوین یادگیری ماشین نظارت‌شده است که از یک ساختار درختی برای مشخص کردن دنباله‌ای از تصمیمات و برآمدها استفاده می‌کند. در این الگوریتم رشد یک درخت با تقسیم گره‌ها (متغیر مستقلی که آزمون بر روی آن صورت می‌گیرد) آغاز و با انتخاب متغیرهای تأثیرگذار در یک رویکرد بالا به پایین به استخراج اطلاعات از داده‌های ورودی می‌پردازد (۳۲). این روش به خصوص زمانی که حجم داده‌ها بالا، روابط بین داده‌ها پیچیده و دسترسی به اطلاعات نهفته بین داده‌ها مشکل باشد دارای کاربرد فراوان است. این روش توانایی کار با داده‌های گسسته و پیوسته را داراست. در این شیوه پیش‌بینی به دست آمده از درخت در قالب یک سری قواعد توضیح داده می‌شود. هر مسیر از ریشه تا یک برگ درخت تصمیم، یک قانون را بیان می‌کند و در نهایت برگ با کلاسی که بیش‌ترین مقدار رکورد در

آن تعلق گرفته برچسب می‌خورد. درخت تصمیم نیاز به محاسبات پیچیده برای دسته‌بندی داده‌ها ندارد. این شیوه مدل‌سازی آماری در مطالعات مربوط به شناسایی زمین‌لغزش در مطالعات تای و همکاران (۲۰۱۸) و رونگ و همکاران (۲۰۲۰) به‌طور گسترده مورد استفاده واقع شده است (۳۳ و ۳۴). این روش در عین سادگی دارای قدرت رقابت بالا با الگوریتم‌های دیگر است اما هرس درخت تصمیم نهایی در صورت تعدد متغیرها هزینه محاسباتی بالایی دارد و با اعمال تغییرات کوچک در داده‌ها، نتایج آن ممکن است دچار تغییر شود.

به‌منظور مقایسه کمی نتایج حاصل از اجرای این سه الگوریتم، از شاخص‌های صحت کلی، ضریب کاپا^۱، حساسیت^۲ و تشخیص‌پذیری^۳ استفاده گردید. در مرحله بعد، هم‌خطی بین متغیرهای منتخب با استفاده از دو شاخص آماری خطای مجاز^۴ و عامل تورم واریانس^۵ با استفاده از بسته *olsrr* انجام شد. براساس این آزمون، اگر شاخص خطای مجاز کوچک‌تر از ۰/۱ و عامل تورم واریانس بزرگ‌تر از ۵ باشد، بیانگر ارتباط هم‌خطی بین متغیرهای مستقل و متغیر وابسته می‌باشد و باید از فرایند مدل‌سازی حذف شوند (۳۵). در مرحله بعد، طبقه‌بندی زمین‌لغزش‌های اتفاق افتاده با چهار الگوریتم شیء‌گرای سیستم بردار پشتیبان، درخت تصمیم، جنگل تصادفی و نزدیک‌ترین همسایه در محیط نرم‌افزار *eCognition* که نرم‌افزار تخصصی طبقه‌بندی شیء‌پایه است انجام شد.

سیستم بردار پشتیبان: این الگوریتم یکی از نیرومندترین روش‌های طبقه‌بندی نظارت‌شده است که به‌طور مشترک می‌تواند برای حل مدل‌های خطی و

غیرخطی و به‌طور مشترک برای دو عملیات رگرسیون و طبقه‌بندی استفاده شود (۳۶). این روش درواقع یک سیستم یادگیری کارآمد است که بر مبنای تئوری بهینه‌سازی و اصل استقرایی کمینه‌سازی خطای ساختاری بوده و منجر به تولید یک جواب بهینه کلی می‌شود. ایده اصلی این الگوریتم به این شکل است که به کمک یک نگاشت غیرخطی فضای داده‌های آموزشی را به یک بعد بالاتر تبدیل می‌کند و در این بعد جدید، به کمک بردارهای پشتیبان و حاشیه‌ها به دنبال ابرصفحه‌ای است که نمونه‌های یک کلاس را از کلاس دیگر جدا کند طوری که فاصله بین صفحاتی که عمل جداسازی را ایجاد می‌کند، حداکثر شود و از این نظر خطای مدل حداقل گردد (۳۷ و ۳۸). این روش یادگیری ماشین در مطالعات مربوط به زمین‌لغزش جهت انتخاب ویژگی و طبقه‌بندی از محبوبیت بالایی برخوردار است (۳۹، ۳۸ و ۴۰).

روش K- نزدیک‌ترین همسایه: این الگوریتم یکی از تکنیک‌های قدیمی پرطرفدار است که اولین بار در سال ۱۹۵۰ معرفی شد. این روش با بررسی تعدادی (K) از شبیه‌ترین همسایه کار طبقه‌بندی را آغاز می‌کند (۴۱) و تعداد موارد برای هر کلاس برچسب می‌خورد و مورد جدید به رده‌ای که تعداد بیش‌تر از همسایه‌ها به آن تعلق دارد نسبت داده می‌شود. این الگوریتم اغلب برای طبقه‌بندی به‌کار می‌رود اگرچه برای پیش‌بینی هم استفاده می‌شود اما در مقایسه با الگوریتم‌های حریصانه یادگیری نظیر درخت تصمیم و شبکه عصبی از قدرت رقابت کم‌تری برخوردار است. اگرچه الگوریتم دارای اجرای ساده و عدم نیاز به برآورد اولیه پارامترها ندارد، اما فرایند یادگیری و روند استنباط و تعمیم آن تا زمان اجرا هزینه محاسباتی بالایی دارد (۴۲).

- 1- Kappa coefficient
- 2- Sensitivity
- 3- Specificity
- 4- Tolerance
- 5- VIF

مفید است. به بیانی دیگر مدل‌های ناپارامتریک شبیه شبکه عصبی، جنگل تصادفی و ماشین بردار پشتیبان، مدل‌های رایج در تجزیه و تحلیل‌های پیش‌بینی هستند، به‌خصوص زمانی که محقق با حجم زیادی از مشاهدات روبرو است و پیش‌فرض آزمون‌های سنتی آماری مانند (رگرسیون خطی چندگانه که نیازمند خطی بودن، همگنی واریانس و نرمال بودن داده‌ها است) در مورد آن‌ها قابل اجرا نیست. ضمن این‌که نتایج این مدل‌ها پیچیده بوده و فهم آن برای مدیران بالادستی سخت می‌باشد. نمودارهای وابستگی جزئی تأثیر هر متغیر پیش‌بینی‌کننده را با ثابت نگه‌داشتن سایر متغیرها بر متغیر پاسخ نشان می‌دهند (۴۴). نمودارهای وابستگی جزئی برای هر متغیر با استفاده از بسته "psych" در نرم‌افزار RStudio 2024.04.0 تولید شد.

نتایج و بحث

انتخاب ویژگی با الگوریتم‌های یادگیری ماشین: نتایج اجرای انتخاب ویژگی‌های مؤثر در شناسایی زمین‌لغزش با استفاده از تصاویر ماهواره‌ای، با استفاده از الگوریتم‌های جنگل تصادفی، درخت تصمیم و شبکه عصبی به‌ترتیب اولویت و اهمیت آن‌ها در جدول ۱ ارائه شده است.

در مرحله بعد بهترین الگوریتم طبقه‌بندی با بررسی و مقایسه درصد صحت، ضریب کاپا، ضریب سورنسن^۱، نرخ مثبت صحیح^۲ و نرخ مثبت کاذب^۳ انتخاب شد. ضریب سورنسن، معیاری برای تشابه بین دو مجموعه A و B است. مقدار عددی این ضریب از ۰ تا ۱ متغیر است که ۱ نشان می‌دهد که دو مجموعه یکسان هستند و عدد صفر نشان می‌دهد که این دو مجموعه هیچ همپوشانی ندارند (۴۳). مقدار ضریب Dice بالا نشان‌دهنده شباهت بالایی بین قطعات/ نمونه‌های پیش‌بینی‌شده بر روی تصویر در مقایسه با واقعیت زمینی است، به این معنی که مدل یا الگوریتم تقسیم‌بندی به‌خوبی عمل می‌کند. برعکس، مقدار ضریب پایین نشان‌دهنده عملکرد ضعیف در تشخیص عوارض در مقایسه با نمونه‌های زمینی بوده است. مقدار ضریب Dice از رابطه ۱ محاسبه شد.

$$\text{Dice coefficient} = 2 * |A \cap B| / (|A| + |B|) \quad (1)$$

در این رابطه، |A| بیانگر تعداد عناصر موجود در واحد A، |B| بیانگر تعداد عناصر در مجموعه B و عبارت (|A| + |B|) بیانگر تعداد عناصر موجود در هر دو مجموعه را نشان می‌دهد (۱۹).

سپس آزمون وابستگی جزئی بین متغیرها و متغیر وابسته انجام شد. هدف از آزمون وابستگی جزئی، بررسی ارتباط بین زیرمجموعه‌ای از متغیرها با متغیر وابسته با در نظر گرفتن میانگین تأثیر سایر متغیرهای پیش‌بینی‌کننده در مدل می‌باشد. این تحلیل برای تفسیر نتایج مدل‌های پیچیده و دارای جعبه سیاه مانند جنگل تصادفی و ماشین بردار پشتیبان بسیار مؤثر و

-
- 1- Sorensen coefficient
 - 2- True positive rate
 - 3- False positive rate

جدول ۱- معیارها و ویژگی‌های انتخاب‌شده با سه روش یادگیری ماشین: جنگل تصادفی، درخت تصمیم و شبکه عصبی.

Table 1. Criteria and selected features by three machine-learning algorithms: random forest, decision tree and neural network.

ترتیب اهمیت	الگوریتم جنگل تصادفی Random forest	الگوریتم درخت تصمیم Decision tree	الگوریتم شبکه عصبی Neural network
1	میانگین مدل رقومی ارتفاع ^۱ Mean DEM	میانگین مدل رقومی ارتفاع Mean DEM	میانگین باند آبی Mean of blue band
2	میانگین باند آبی Mean blue band	میانگین باند آبی Mean blue band	میانگین باند سبز Mean green band
3	میانگین شیب Mean slope	میانگین شیب Mean slope	انحراف معیار سازند STD ^۲ of formation
4	انحراف معیار DEM STD of DEM	میانگین باند قرمز Mean red band	انحراف معیار باند آبی STD of blue band
5	میانگین جهت شیب Mean slope direction	انحراف معیار DEM STD of DEM	آنتروپی باند آبی Entropy of blue band
6	میانگین باند قرمز Mean red band	میانگین باند NIR Mean NIR band	انحراف معیار باند قرمز STD red band
7	میانگین باند NIR Mean NIR band	میانگین DVI Mean DVI	همگنی باند قرمز Homogeneity of the red band
8	میانگین DVI Mean DVI	میانگین باند سبز Mean green band	انحراف معیار جهت شیب STD of slope direction
9	میانگین باند سبز Mean green band	عرض لغزش Widths of landslide	میانگین سازند Mean formation
10	میانگین EVI Mean EVI	میانگین NDVI Mean NDVI	انحراف معیار باند سبز STD of green band
11	میانگین فاصله تا رودخانه Mean distance to the river	میانگین EVI Mean EVI	مساحت لغزش Area of landslide
12	عرض لغزش Widths of Landslide	میانگین جهت شیب Mean slope direction	میانگین باند قرمز Mean red band
13	مساحت لغزش Area of landslide	انحراف معیار جهت شیب STD of slope direction	تجانس باند قرمز
14	انحراف معیار جهت شیب STD of slope direction	آنتروپی باند NIR Entropy of NIR band	میانگین فاصله تا رود Mean distance to the river
15	میانگین SAVI Mean SAVI	میانگین SAVI Mean SAVI	عدم تشابه باند NIR dissimilarity of the NIR band
16	میانگین NDVI Mean NDVI	میانگین فاصله تا جاده Mean distance to the road	عدم تشابه باند آبی dissimilarity of the blue band

1- Digital height model (DEM)

2- Standard deviation

انتخاب ویژگی با استفاده از ۵۳ ویژگی انتخاب‌شده به‌عنوان عوامل مؤثر در شناسایی زمین‌لغزش با استفاده از تصاویر ماهواره‌ای به‌صورت حذف تک‌تک ویژگی‌ها و تأثیری که حذف ویژگی‌ها بر روی صحت کلی و ضریب کاپا داشتند انجام شد. بر اساس نتایج بالاترین میزان صحت کلی و ضریب کاپا، با تعداد ۱۶ ویژگی به‌دست آمد و کم‌تر یا بیش‌تر شدن تعداد ویژگی‌ها از ۱۶ ویژگی، منجر به کاهش میزان صحت کلی شد. براساس نتایج مطالعات آماتیا و همکاران (۲۰۲۱) و شیروانی و همکاران (۲۰۱۹) انتخاب ویژگی و کاهش ابعاد مسأله منجر به بهبود نتایج الگوریتم‌های یادگیری ماشین به‌خصوص هنگام کار با داده‌های زیاد خواهد شد که با نتایج پژوهش حاضر مطابقت دارد (۲۱، ۱۳ و ۴۵). همان‌طور که در جدول ۱ مشاهده می‌شود، به‌جز دو ویژگی میانگین فاصله تا جاده و شاخص آنتروپی باند مادون‌قرمز نزدیک، سایر ویژگی‌های انتخاب‌شده با روش درخت تصمیم، مشابه روش جنگل تصادفی بوده است و تنها در ترتیب اهمیت آن‌ها تفاوت وجود دارد. در روش جنگل تصادفی دو ویژگی میانگین فاصله تا رودخانه و مساحت لغزش متفاوت از روش درخت تصمیم بوده است. هم‌چنین در الگوریتم جنگل تصادفی و درخت تصمیم سه ویژگی میانگین لایه ارتفاعی، میانگین باند آبی و میانگین لایه شیب به‌ترتیب به‌عنوان مهم‌ترین ویژگی در شناسایی زمین‌لغزش انتخاب شدند. میانگین لایه ارتفاعی در تجزیه و تحلیل‌های جغرافیایی اهمیت دارد و بر طیف گسترده‌ای از فرآیندهای طبیعی و انسانی تأثیر می‌گذارد (۱۴). اهمیت بالای آن در مدل‌های یادگیری ماشین نشان‌دهنده تطبیق‌پذیری و نقش اساسی آن در ثبت و پیش‌بینی دقیق پدیده‌های محیطی و مکانی است. بازتاب باند آبی و تأثیر آن بر سلامت و نوع پوشش گیاهی می‌تواند به شناسایی زمین‌لغزش‌ها با استفاده از تصاویر ماهواره‌ای کمک کند. کاهش

بازتاب باند آبی می‌تواند نشان‌دهنده تنش گیاهی و تغییرات در رطوبت خاک باشد که در نتیجه می‌تواند منجر به وقوع زمین‌لغزش شود. هم‌چنین انتخاب میانگین لایه شیب به‌عنوان یک عامل مهم می‌تواند به‌این‌علت باشد که با تحلیل میانگین لایه شیب، می‌توان مناطقی را که احتمال وقوع زمین‌لغزش در آن‌ها بیش‌تر است شناسایی کرد. لایه شیب می‌تواند به تحلیل الگوهای زمین‌شناسی و نوع خاک نیز کمک کند، این اطلاعات می‌تواند منجر به درک بهتر رفتار زمین و پایداری آن شود. در الگوریتم جنگل تصادفی هیچ‌یک از ویژگی‌های بافت (عدم تشابه باندهای تصویر، آنتروپی و ...) تصویر جزو ویژگی‌های پراهمیت انتخاب نشدند. در الگوریتم درخت تصمیم نیز به‌جز ویژگی آنتروپی باند مادون‌قرمز نزدیک سایر ویژگی‌های بافت تصویر جزو ویژگی‌های بی‌اهمیت شناسایی شدند. برخلاف این دو الگوریتم، در الگوریتم شبکه عصبی ویژگی‌های بافت تصویر سهم زیادی در شناسایی زمین‌لغزش‌ها داشتند و ویژگی‌هایی همچون شاخص آنتروپی باند آبی، همگنی باند قرمز، تجانس باند قرمز، عدم تشابه باند مادون‌قرمز نزدیک^۱ و عدم تشابه باند آبی به‌عنوان ویژگی‌های مؤثر در شناسایی انتخاب شدند. هم‌چنین در این روش میانگین باند آبی به‌عنوان مهم‌ترین ویژگی انتخاب شد که در دو روش دیگر نیز در رده دوم اهمیت قرار دارد. شیروانی و همکاران (۲۰۱۹) در مطالعه خود ویژگی‌های بافت تصویر و ویژگی‌های طیفی تصویر را مهم‌ترین ویژگی برای تمایز زمین‌لغزش از مناطق فاقد زمین‌لغزش عنوان کردند (۴۵). در این مطالعه نیز ویژگی‌های طیفی جزو ویژگی‌های پراهمیت در شناسایی زمین‌لغزش انتخاب شدند اما ویژگی‌های بافت تصویر در امر شناسایی زمین‌لغزش کمک‌کننده نبوده و جزو عوامل مؤثر انتخاب نشدند.

1- Near infrared band (NIR)

میزان صحت کلی ۹۲ درصد و ضریب کاپا ۰/۴۱ در رتبه دوم قرار گرفت. روش درخت تصمیم به دلیل پایین بودن شاخص صحت کلی و شاخص تشخیص پذیری در اولویت آخر قرار گرفت.

جدول ۲ نتایج حاصل از مقایسه کمی نتایج حاصل از سه الگوریتم انتخاب ویژگی را نشان می‌دهد. براساس نتایج الگوریتم جنگل تصادفی با میزان صحت کلی ۹۴ درصد و ضریب کاپا ۰/۴۸ به‌عنوان روش مناسب در رتبه اول و شبکه عصبی با

جدول ۲- مقایسه کمی الگوریتم‌های یادگیری ماشین در انتخاب ویژگی.

Table 2. Quantitative comparison of machine learning algorithms in feature selection.

روش	صحت کلی (%)	ضریب کاپا	حساسیت	تشخیص پذیری
Method	Overall accuracy (%)	kappa coefficient	Sensitivity	Specificity
جنگل تصادفی Random forest	94	0.48	0.96	0.62
شبکه عصبی Neural network	92	0.41	0.93	0.55
درخت تصمیم Decision tree	68	0.05	0.80	0.25

چهار الگوریتم طبقه‌بندی شیء پایه درخت تصمیم، سیستم بردار پشتیبان، جنگل تصادفی و نزدیک‌ترین همسایه نشان می‌دهد. به لحاظ بصری، الگوریتم‌های درخت تصمیم و جنگل تصادفی بخش زیادی از منطقه مورد مطالعه را به‌عنوان قطعات زمین‌لغزش شناسایی کرده است. این مسأله ناشی از حساسیت بالای این مدل‌ها و تعدد درخت‌های تصمیم تولیدشده به دلیل شدت بالای پیچیدگی مسأله بوده است. تراکم قطعات شناسایی‌شده در روش نزدیک‌ترین همسایه نسبت به الگوریتم‌های درخت تصمیم و جنگل تصادفی کمتر بوده و براساس اصل نزدیک بودن و شباهت قطعات نزدیک به هم، طبقه‌بندی زمین‌لغزش انجام شده است. این الگوریتم نتوانست مناطق فاقد پوشش گیاهی اما بدون لغزش را از مناطق زمین‌لغزش به‌خوبی تفکیک کند. بهترین نتیجه شناسایی موقعیت مکانی زمین‌لغزش‌ها با استفاده از الگوریتم سیستم بردار پشتیبان به‌دست آمد، طوری‌که قطعات شناسایی‌شده بر

پس از انتخاب زیرمجموعه از عوامل مؤثر در شناسایی زمین‌لغزش، هم خطی بین متغیرهای مستقل بررسی شد (جدول ۳). از آنجایی‌که مقادیر عامل تورم واریانس بزرگ‌تر از ۵ و خطای مجاز کم‌تر از ۰/۱ نیست، بنابراین می‌توان نتیجه گرفت که ارتباط هم خطی بین هیچ‌یک از متغیرهای مورد بررسی با متغیر پاسخ وجود ندارد. کم‌ترین مقدار عامل تورم واریانس ۱/۰۵۶ مربوط به ویژگی انحراف معیار جهت شیب و بیش‌ترین مقدار آن ۳/۰۸۶ مربوط به متغیر میانگین باند NIR بوده است. مقادیر مربوط به تورم واریانس برای سایر متغیرها بین این دو مقدار نوسان داشته است. هم‌چنین کم‌ترین میزان خطای مجاز برابر ۰/۳۲۴ مربوط به عامل میانگین باند NIR و بیش‌ترین مقدار آن برابر ۰/۹۴۷ مربوط به عامل انحراف معیار جهت شیب بوده است.

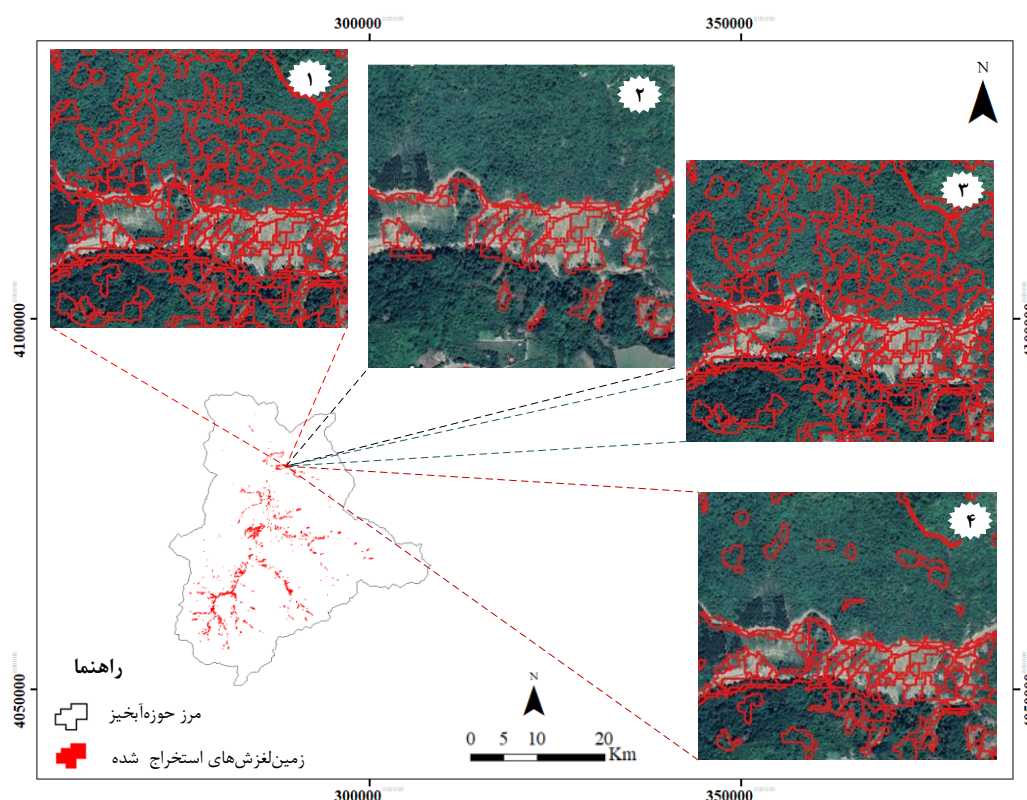
طبقه‌بندی زمین‌لغزش: شکل ۴ بخش واحدی از منطقه مورد مطالعه را در شناسایی زمین‌لغزش برای

روی تصویر تطابق بیشتری را با زمین‌لغزش‌های مشاهده‌شده در بازدید صحرایی از خود نشان داده است. عابدینی و همکاران (۲۰۱۶) نیز در مطالعه خود استفاده از روش شیء پایه را برای شناسایی زمین‌لغزش مناسب دانستند (۸).

جدول ۳- نتایج بررسی هم خطی بین متغیرهای مستقل در شناسایی زمین‌لغزش.

Table 3. Collinearity analysis between independent variables in detecting landslides.

خطای مجاز Tolerance	عامل تورم واریانس (VIF) Variance inflation factor	ویژگی‌های مؤثر Effective features
0.445	2.245	عرض لغزش Widths of landslide
0.884	1.131	انحراف معیار DEM STD of DEM
0.947	1.056	انحراف معیار جهت شیب STD of slope direction
0.846	1.182	میانگین فاصله تا رود Mean distance to the
0.465	2.151	میانگین باند قرمز Mean red band
0.324	3.086	میانگین باند NIR Mean NIR band
0.510	1.962	میانگین باند سبز Mean green band
0.847	1.180	میانگین DVI Mean DVI
0.725	1.380	میانگین باند آبی Mean blue band
0.921	1.085	میانگین جهت شیب Mean slope direction
0.495	2.019	مساحت لغزش Area of landslide
0.398	2.510	میانگین شیب Mean slope
0.652	1.533	میانگین SAVI Mean SAVI
0.505	1.980	میانگین NDVI Mean NDVI
0.833	1/200	میانگین EVI Mean EVI
0.725	1/380	میانگین DEM Mean DEM



شکل ۴- زمین لغزش‌های شناسایی شده با روش‌های مختلف طبقه‌بندی، (۱): الگوریتم درخت تصمیم،

(۲) الگوریتم ماشین بردار پشتیبان، (۳): الگوریتم جنگل تصادفی و (۴): الگوریتم نزدیک‌ترین همسایه.

Figure 4. Landslides detected by different classification methods, (1): decision tree algorithm, (2) support vector system algorithm, (3): random forest algorithm & (4): nearest neighbor algorithm.

همسایه و درخت تصمیم در اولویت‌های بعدی و الگوریتم جنگل تصادفی کم‌ترین مقدار صحت (۸۲ درصد) و ضریب کاپا (۰/۶۵) را از خود نشان داده است. شیروانی و همکاران (۲۰۱۹) با الگوریتم جنگل تصادفی زمین لغزش‌های پارک ملی گلستان را برای مناطق حفاظت شده و غیرحفاظتی به ترتیب با صحت برابر ۸۶/۶ و ۸۰/۳ درصد شناسایی کردند (۴۵). در مطالعه حاضر نیز روش جنگل تصادفی با صحت ۸۲ درصد توانست زمین لغزش‌ها را شناسایی کند اما عملکرد الگوریتم‌های دیگر بهتر بوده است. سان و همکاران (۲۰۲۳) در مطالعه خود عملکرد مدل جنگل تصادفی را بهتر از روش ماشین بردار پشتیبان گزارش دادند که با نتیجه پژوهش حاضر یکسان نیست (۱۵). گلی‌مختاری و ناعمی‌تبار (۲۰۲۲) در پژوهش خود به

جدول ۴ ارزیابی کمی نتایج و مقایسه مربوط به شناسایی زمین لغزش بر روی تصاویر ماهواره‌ای با چهار الگوریتم طبقه‌بندی، ماشین بردار پشتیبان، درخت تصمیم، جنگل تصادفی و نزدیک‌ترین همسایه برای دو تصویر مورد مطالعه را در حوزه آبخیز محمدآباد نشان می‌دهد. در اکثر مطالعات ارزیابی نتایج طبقه‌بندی تنها با شاخص AUC انجام شده است (۱۳، ۱۴ و ۱۱) اما در این مطالعه نتایج با پنج شاخص بررسی شده است. براساس نتایج مشاهده می‌شود که بر روی هر دو تصویر مطالعه شده، الگوریتم سیستم بردار پشتیبان بهترین عملکرد (صحت بالای ۹۲ درصد و ضریب کاپا بالای ۰/۸۵) را نسبت به سه روش دیگر در شناسایی زمین لغزش‌ها از خود نشان داده است. الگوریتم‌های نزدیک‌ترین

است بر روی هر دو تصویر برای الگوریتم سیستم بردار پشتیبان مشاهده شد. در مقابل نرخ مثبت صحیح، نرخ مثبت کاذب قرار دارد که یک معیار منفی بوده، کم‌ترین آن برای الگوریتم نزدیک‌ترین همسایه بر روی تصویر ۱ (۰/۰۶) و بیش‌ترین آن الگوریتم بردار پشتیبان بر روی تصویر یک (۰/۳۴) مشاهده گردید. پراتون و ساجیتامپان (۲۰۲۴) برای پهنه‌بندی زمین‌لغزش در منطقه کوهستانی با استفاده از مدل ماشین بردار پشتیبان بهترین نتیجه را گرفتند که با نتایج مطالعه حاضر مطابقت دارد (۴۶).

این نتیجه رسیدند که الگوریتم تابع شواهد قطعی نسبت به الگوریتم‌های ماشین بردار پشتیبان و تابع شواهد وزنی با صحت بیش‌تری مناطق حساس زمین‌لغزش را شناسایی کرده است (۱۴). نتایج شاخص/ضریب سورنسن که به اندازه‌گیری شباهت تصویر طبقه‌بندی‌شده با واقعیت زمینی منطقه می‌پردازد، نشان داد که بیش‌ترین مقدار این ضریب با استفاده از الگوریتم ماشین بردار پشتیبان (بالای ۰/۹۲) در هر دو تصویر و کم‌ترین آن برای الگوریتم جنگل تصادفی بر روی هر دو تصویر مشاهده شد. هم‌چنین بیش‌ترین مقدار نرخ مثبت صحیح که یک معیار مثبت

جدول ۴- نتایج اجرای الگوریتم‌های طبقه‌بندی برای شناسایی زمین‌لغزش‌های بر روی تصاویر مورد مطالعه.

Table 4. The results of classification algorithms to identify landslide events on the studied images.

نرخ مثبت کاذب False Positive Rate (FPR)	نرخ مثبت صحیح True Positive Rate (TPR) or Recall	ضریب سورنسن Dice-Sørensen coefficient	ضریب کاپا kappa coefficient	صحت (%) accuracy	الگوریتم طبقه‌بندی Classification algorithm	
0.34	0.91	0.93	0.87	93	ماشین بردار پشتیبان Support vector machine	تصویر ۱ Image 1
0.06	0.81	0.86	0.74	87	نزدیک‌ترین همسایه Nearest neighbor	
0.22	0.89	0.84	0.67	83	درخت تصمیم Decision tree	
0.20	0.86	0.83	0.65	82	جنگل تصادفی Random forest	
0.30	0.89	0.92	0.85	92	ماشین بردار پشتیبان Support vector machine	تصویر ۲ Image 2
0.15	0.87	0.86	0.71	85	نزدیک‌ترین همسایه Nearest neighbor	
0.20	0.90	0.85	0.70	85	درخت تصمیم Decision tree	
0.32	0.96	0.84	0.64	82	جنگل تصادفی Random forest	

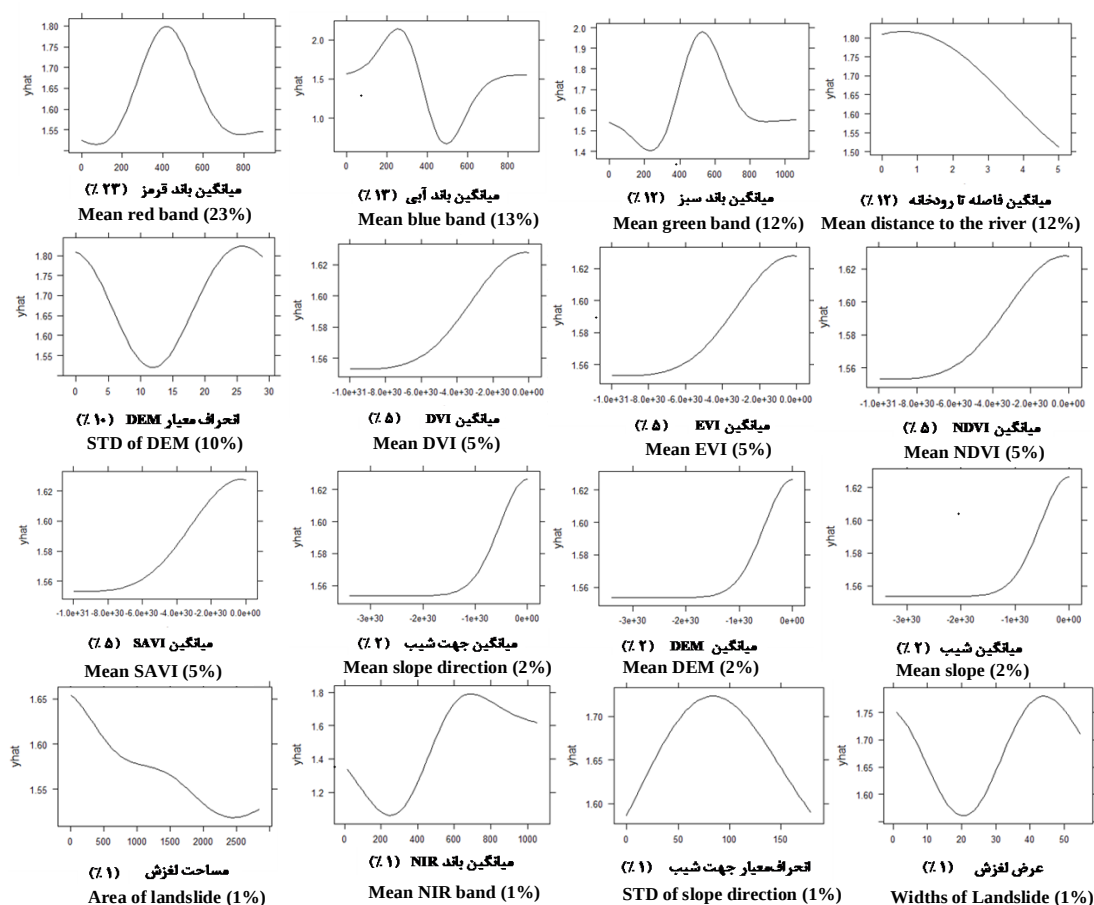
در شکل ۵ ارائه شده است. درواقع این نمودار، برای هر ویژگی، سهمی از آن متغیر را که ممکن است شانس مدل را در پیش‌بینی کلاس‌های متغیر پاسخ بالا برد نشان می‌دهد. بر این اساس متغیرهای میانگین

نمودار وابستگی جزئی^۱ بین متغیر وابسته (شناسایی زمین‌لغزش بر روی تصاویر ماهواره‌ای) با متغیرهای مستقل استخراج شده از تصویر مورد مطالعه

1- Partial dependence plots

رودخانه‌ها معمولاً تحت تأثیر فرسایش و تغییرات هیدرولوژیکی قرار دارند که می‌تواند منجر به وقوع زمین‌لغزش شود. انحراف معیار DEM نشان‌دهنده تغییرات ارتفاعی در سطح زمین است و می‌تواند به شناسایی نواحی با شیب‌های تند و ناپایدار کمک کند. همچنین متغیرهای عرض لغزش، انحراف معیار جهت شیب، میانگین باند NIR و مساحت لغزش کم‌ترین سهم را در شناسایی زمین‌لغزش با استفاده از تصاویر ماهواره‌ای داشته‌اند.

باند قرمز، میانگین باند آبی، میانگین باند سبز، میانگین فاصله تا رودخانه و انحراف معیار DEM بیش‌ترین سهم را در شناسایی زمین‌لغزش داشته است. باندهای قرمز، آبی و سبز به دلیل توانایی بالای خود در تشخیص تغییرات پوشش زمین و ویژگی‌های سطحی، نقش مهمی در شناسایی زمین‌لغزش‌ها دارند. باند قرمز به‌طور خاص به تشخیص نواحی با رطوبت بالا و تغییرات در پوشش گیاهی کمک می‌کند که می‌تواند نشانه‌ای از وقوع زمین‌لغزش باشد. نواحی نزدیک به



شکل ۵- همبستگی جزئی میان عوامل مؤثر در شناسایی زمین‌لغزش بر اساس مدل ماشین بردار پشتیبان و متغیر پاسخ (حضور و عدم حضور). محور Y بیانگر تابع برازش و محور X سهم هر یک از متغیرهای وابسته در شناسایی می‌باشد.

Figure 5. Partial correlation between the effective factors in landslide detection based on the support vector machine model and the response variable (presence and absence). The Y-axis represents the fitting function and the X-axis represents the contribution of each of the dependent variables in identification.

با توجه به شکل ۳ مشاهده می‌شود، ویژگی‌های هندسی با امتیاز ۱٪ جزو کم‌اهمیت‌ترین ویژگی‌های مؤثر در شناسایی زمین‌لغزش بودند که با مطالعات شیروانی و همکاران (۲۰۱۹) هم‌خوانی دارد (۴۵). در مطالعه حاضر برای بهبود تفکیک زمین‌لغزش با زمین‌های بایر شاخص‌های پوشش گیاهی وارد مدل شد براساس نتایج نمودارهای وابستگی جزئی نیز مشاهده می‌شود این شاخص‌ها جزو شاخص‌های پراهمیت در شناسایی زمین‌لغزش بوده است که با مطالعات ونا و همکاران (۲۰۱۶) و سان و همکاران (۲۰۱۷) مطابقت دارد (۲۰ و ۴۷).

نتیجه‌گیری کلی

در این مطالعه، روش تلفیقی از الگوریتم‌های یادگیری ماشین و شیء‌گرا برای تهیه نقشه طبقه‌بندی زمین‌لغزش استفاده شد. انتخاب ویژگی با استفاده از ۵۳ ویژگی اولیه، حذف تدریجی ویژگی‌ها و بررسی تأثیر آن‌ها بر صحت کلی و ضریب کاپا با سه روش جنگل تصادفی، درخت تصمیم و شبکه عصبی انجام شد. ارزیابی انتخاب ویژگی با این سه روش نشان داد بهترین روش انتخاب ویژگی، الگوریتم جنگل تصادفی با صحت بالای ۹۴ درصد بود که قادر به تفکیک ۱۶ ویژگی مؤثر در شناسایی زمین‌لغزش بوده است. الگوریتم‌های جنگل تصادفی و درخت تصمیم در شناسایی ویژگی‌های مؤثر در شناسایی زمین‌لغزش شباهت زیادی با هم داشتند و ویژگی‌های میانگین لایه ارتفاعی، میانگین باند آبی و میانگین لایه شیب به‌عنوان مهم‌ترین عوامل شناسایی شدند. در مقابل، الگوریتم شبکه عصبی تأکید بیشتری بر ویژگی‌های بافت تصویر داشت و ویژگی‌هایی مانند شاخص آنتروپی باند آبی و همگنی باند قرمز را به‌عنوان ویژگی‌های مؤثر شناسایی کرد. این تفاوت‌ها نشان‌دهنده اهمیت انتخاب ویژگی مناسب برای هر الگوریتم

است. به‌طورکلی، انتخاب ویژگی‌ها و کاهش ابعاد داده‌ها می‌تواند منجر به بهبود عملکرد الگوریتم‌های یادگیری ماشین شود، به‌ویژه زمانی که با داده‌های حجیم و پیچیده سروکار داریم. پس از انتخاب عوامل مؤثر در شناسایی زمین‌لغزش، هم‌خطی بین متغیرهای مستقل بررسی شد. نتایج این بخش نشان داد بین ویژگی‌های مورد بررسی و متغیر پاسخ هم‌خطی وجود ندارد. سپس طبقه‌بندی زمین‌لغزش‌ها با چهار الگوریتم شیء‌گرا ماشین بردار پشتیبان، درخت تصمیم، جنگل تصادفی و نزدیک‌ترین همسایه انجام شد. الگوریتم ماشین بردار پشتیبان با صحت بالای ۹۲ درصد و ضریب کاپا بالای ۰/۸۵ به نسبت سایر الگوریتم‌های آزمون شده نتایج بهتری از خود نشان داد. درنهایت نمودار وابستگی جزئی ویژگی‌ها، به‌منظور نمایش بصری سهم هر یک از ویژگی‌های مؤثر بر شناسایی زمین‌لغزش، تفسیر و فهم بهتر مدل و تعاملات بین ویژگی‌ها انجام شد. مطالعه حاضر نشان داد که استفاده از قابلیت و توانایی الگوریتم‌های یادگیری ماشین می‌تواند به‌عنوان روش قابل‌اعتماد برای تشخیص سریع‌تر زمین‌لغزش‌ها مورد استفاده مدیران و کارشناسان اجرایی کشور قرار گیرد. از جمله محدودیت‌های این پژوهش نبود تصاویر با قدرت تفکیک مکانی بالا بوده است که پیشنهاد می‌گردد در مطالعات آینده از تصاویر با قدرت تفکیک مکانی بالا استفاده شود تا با جزئیات بالا بتوان به شناسایی زمین‌لغزش‌ها اقدام نمود.

تقدیر و تشکر

نویسندگان این مقاله از سازمان فضایی ایران به‌دلیل ارائه تصاویر ماهواره‌ای گائوفن-۱ صمیمانه تشکر می‌کنند. همچنین از دانشگاه علوم کشاورزی و منابع طبیعی گرگان قدردانی می‌شود. تشکر ویژه‌ای نیز از اداره کل منابع طبیعی و آبخیزداری استان گلستان و قرقبان‌های حوزه آبخیز محمدآباد می‌گردد

مشارکت نویسندگان

مرضیه نیکجوی (جمع‌آوری داده، پردازش، اجرای مدل و نوشتن نسخه اولیه مقاله)، علی نجفی‌نژاد (مدیریت و مسئول مکاتبات)، حمیدرضا پورقاسمی (مدیریت و بازبینی مقاله) و چوقی بایرام کمکی (ارائه مشاوره در انجام تحقیق و بازبینی مقاله).

اصول اخلاقی

نویسندگان اصول اخلاقی را در انجام و انتشار این اثر عملی رعایت نموده‌اند و این موضوع مورد تأیید همه آن‌ها می‌باشد.

حمایت مالی

این پژوهش از حمایت معاونت آموزشی و پژوهشی دانشگاه علوم کشاورزی و منابع طبیعی گرگان در قالب پژوهانه دانشجویی نوسنده اول برخوردار بوده است. همچنین تصاویر ماهواره‌ای به‌صورت رایگان توسط سازمان فضایی ایران در اختیار قرار گرفت.

که با همکاری و اطلاعات میدانی ارزشمندشان در اجرای دقیق این پژوهش کمک کردند.

داده‌ها و اطلاعات و دسترسی

پژوهش حاضر، بخشی از رساله دکتری نویسنده اول می‌باشد تصاویر ماهواره‌ای مورد استفاده در این پژوهش از سازمان فضایی ایران تهیه شد. همچنین، بخش دیگری از داده‌ها از مطالعات آبخیزداری حوزه آبخیز محمدآباد استخراج شد که از اداره کل منابع طبیعی و آبخیزداری استان گلستان دریافت گردید. داده‌های خام این پژوهش با مکاتبه با نویسنده اول قابل دسترسی است؛ اما به دلیل تعهد به سازمان فضایی ایران از ارائه تصاویر ماهواره‌ای معذور هستیم.

تعارض منافع

در این مقاله تضاد منافی وجود ندارد و این مسأله مورد تأیید همه نویسندگان است.

منابع

1. Wang, L., Xiao, T., Liu, S., Zhang, W., Yang, B., & Chen, L. (2023). Quantification of model uncertainty and variability for landslide displacement prediction based on Monte Carlo simulation. *Gondwana Research*, 123, 27-40.
2. Heydari, N., Habibnejad, M., Kavian, A., & Pourqasmi, H. R. (2019). Landslide Susceptibility Modelling Using the Random Forest Machine Learning Algorithm in the Watershed of Rais-Ali Delvari Reservoir. *Watershed research (research and development)*, 33 (1), 13-2. SID. [https:// sid.ir/ paper/ 386906/ fa](https://sid.ir/paper/386906/fa). [In Persian]
3. Heydari, N., Habibnejad, M., Kavian, A., & Pourqasmi, H. R. (2019). Landslide Susceptibility Modelling Using the Random Forest Machine Learning Algorithm in the Watershed of Rais-Ali Delvari Reservoir. *Watershed research (research and development)*, 33(1), 126 series), 13-2. SID. <https://sid.ir/paper/386906/fa>. [In Persian]
4. Karakas, G., Unal, E. O., Cetinkaya, S., Ozcan, N. T., Karakas, V. E., Can, R., & Kocaman, S. (2024). Analysis of landslide susceptibility prediction accuracy with an event-based inventory: The 6 February 2023 Turkiye earthquakes. *Soil Dynamics and Earthquake Engineering*, 178, 108491.
5. Chen, W., Xie, X., Wang, J., Pradhan, B., Hong, H., Tien Bui, D., 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.

6. Martha, T. R., Kerle, N., Jetten, V., van Westen, C. J., & Kumar, K. V. (2010). Characterising spectral, spatial and morphometric properties of landslides for semi-automatic detection using object-oriented methods. *Geomorphology*, 116, 24-36.
7. Blaschke, T., Hay, G. J., Kelly, M., Lang, S., Hofmann, P., Addink, E., & Tiede, D. (2014). Geographic object-based image analysis—towards a new paradigm. *ISPRS journal of photogrammetry and remote sensing*, 87, 180-191.
8. Abedini, M., Roustai, SH., & Fathi, M. H. (2016). Identification and classification of landslide types using Spectral and spatial features with an object-oriented method approach (Nasirabad to Sattarkhan Ahar Dam). *Scientific-research journal of geography and planning*, 22 (66), 187-205. [In Persian]
9. Fathi, M., H. Abedini, M., & Roostaei, Sh. (2019). Identification and zonation landslide prone areas using object oriented method and conditional probability theory (Bayesian theorem) Case Study: Ahar drainage basin South boundary (From Nasirabad to Sattar Khan dam). *Journal of Geographic space*, 18 (64), 20-40. [In Persian]
10. Ghanavati, E., Ahmadabadi, A., & Gholami, M. (2019). Landslide susceptibility mapping of Kan using index of Entropy and LSM. *Quantitative geomorphological researches*. 8 (1), 16-33. [In Persian]
11. Kornejady, A., & Pourghasemi, H. R. (2019). Landslide susceptibility assessment using data mining models, a case study: Chehel-Chai Basin. *Watershed Engineering and Management*, 11 (1), 28-42. [In Persian]
12. Gasemyan, B., Abedini, M., & Roostaei, Sh. (2021). Landslide susceptibility assessment using a novel ensemble algorithm based model (Case Study: Kamyaran city, Kurdistan province). *Quantitative geomorphological researches*, 9 (4), 130-146. [In Persian]
13. Amatya, P., Kirschbaum, D., Stanley, T., & Tanyas, H. (2021). Landslide mapping using object-based image analysis and open source tools. *Engineering Geology*, 288, 106000. <https://doi.org/10.1016/j.enggeo.2021.106000>.
14. GoliMokhtari, L., & NaemiTabar, M. (2022). Spatial modeling and prediction of landslide risk using advanced data mining algorithms (case study: Kalt city). *Quantitative geomorphology research*, 10 (4), 137-116. [In Persian]
15. Sun, D., Gu, Q., Wen, H., Xu, J., Zhang, Y., Shi, S., & Zhou, X. (2023). Assessment of landslide susceptibility along mountain highways based on different machine learning algorithms and mapping units by hybrid factors screening and sample optimization. *Gondwana Research*, 123, 89-106.
16. Mohammad Abad Kalate watershed study project (2015). Khorasan flood control consulting engineers company. [In Persian]
17. Pourghasemi, H. R., Moradi, M., Mohammadi, B., Pradhan, R., Mostafazadeh, & Goli Jirandeh, A. (2012). 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.
18. Agarwal, S., Vailshery, L., Jaganmohan, M., & Nagendra, H. (2013). Mapping Urban Tree Species Using Very High Resolution Satellite Imagery: Comparing Pixel-Based and Object-Based Approaches. *ISPRS International Journal of Geo-Information*, 2 (1), 220-236. <https://doi.org/10.3390/ijgi2010220>.
19. Eelbode, T., Bertels, J., Berman, M., Vandermeulen, D., Maes, F., Bisschops, R., & Blaschko, M. B. (2020). Optimization for medical image segmentation: theory and practice when evaluating with dice score or jaccard index. *IEEE transactions on medical imaging*, 39 (11), 3679-3690.

20. Veena, V. S., Sai, S. G., Tapas, R. M., Deepak, M., & Rama, R. N. (2016). Automatic detection of landslides in object-based environment using open source tools. In *Proceedings of the GEOBIA 2016, Solutions and synergies, Enschede, The Netherlands*, 14-16.
21. Altarabichi, M. G., Nowaczyk, S., Pashami, S., & Sheikholharam Mashhadi, P. (2023). Fast Genetic Algorithm for feature selection-A qualitative approximation approach. In *Proceedings of the companion conference on genetic and evolutionary computation*. 11-12.
22. Pourgholam-Amiji, M., Ahmadaali, Kh., & Liaghat, A. M. (2021). Sensitivity Analysis of Parameters Affecting the Early Cost of Drip Irrigation Systems Using Meta-Heuristic Algorithms. *Iranian Journal of Irrigation and Drainage*, 15 (4), 737-756. Doi: <https://sid.ir/paper/1054363/fa>. [In Persian]
23. Zhang, W., Wu, C., Li, Y., Wang, L., & Samui, P. (2019). Assessment of pile drivability using random forest regression and multivariate adaptive regression splines. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, 15 (1), 27-40.
24. R Core Team. (2024). R: A Language and Environment for Statistical Computing_R Foundation.
25. Noori, S., Nourijelyani, K., Mohammad, K., Niknam, M., Mahmoudi, M., & Andonian, L. (2012). Random forests analysis: A modern statistical method for screening in high-dimensional studies and its application in a population-based genetic association study. *North Khorasan University of Medical Sciences*. 3 (5), 93-101. [In Persian]
26. Vorpahl, P., Elsenbeer, H., Märker, M., & Schröder, B. (2012). How can statistical models help to determine driving factors of landslides?. *Ecological Modelling*, 239, 27-39.
27. Kim, J. C., Lee, S., Jung, H. S., & Lee, S. (2017). Landslide susceptibility mapping using random forest and boosted tree models in PyeongChang, Korea. *Geocarto International*, 33(9), 1000-1015.
28. Chang, Z., Catani, F., Huang, F., Liu, G., Meena, S. R., Huang, J., & Zhou, C. (2023). Landslide susceptibility prediction using slope unit-based machine learning models considering the heterogeneity of conditioning factors. *Journal of Rock Mechanics and Geotechnical Engineering*. 15 (5), 1127-1143.
29. Wang, S.C. (2003). Artificial neural network. In *Interdisciplinary computing in java programming* (pp. 81-100). springer, boston, MA.
30. Liu, S., Wang, L., Zhang, W., He, Y., & Pijush, S. (2023). A comprehensive review of machine learning-based methods in landslide susceptibility mapping. *Geological Journal*, 58 (6), 2283-2301.
31. Sameen, M. I., Pradhan, B., & Lee, S. (2020). Application of convolutional neural networks featuring Bayesian optimization for landslide susceptibility assessment. *Catena*, 186, 104249.
32. Tsangaratos, P., & Ilia, I. (2015). Landslide susceptibility mapping using a modified decision tree classifier in the Xanthi Prefecture, Greece. *Landslides*, 13(2), 305-320.
33. Thai Pham, B., Tien Bui, D., & Prakash, I. (2018). Landslide susceptibility modelling using different advanced decision trees methods. *Civil Engineering and Environmental Systems*, 35(1-4), 139-157.
34. Rong, G., Alu, S., Li, K., Su, Y., Zhang, J., Zhang, Y., & Li, T. (2020). Rainfall induced landslide susceptibility mapping based on Bayesian optimized random forest and gradient boosting decision tree models-A case study of Shuicheng County, China. *Water*, 12 (11), 3066.
35. Grzywiński, W., Turowski, R., Naskrent, B., Jelonek, T., & Tomczak, A. (2019). The effect of season of the year on the frequency and degree of damage during commercial thinning in black alder stands in Poland. *Forests*, 10 (8), 668.
36. Vapnik, V. N. (1995). Introduction: Four periods in the research of the learning

- problem. In The nature of statistical learning theory. *Springer*.
37. Peng, L., Niu, R., Huang, B., Wu, X., Zhao, Y., & Ye, R. (2014). Landslide susceptibility mapping based on rough set theory and support vector machines: A case of the Three Gorges area, China. *Geomorphology*, 204, 287-301.
 38. Tien Bui, D., Pradhan, B., Lofman, O., & Revhaug, I. (2012). Landslide susceptibility assessment in vietnam using support vector machines, decision tree, and Naive Bayes Models. *Mathematical problems in Engineering*, 2012 (1), 974638.
 39. Ghasemian, B., Abedini, M., & Roostaei, Sh. (2016). Landslide sensitivity assessment using support vector machine algorithm (case study: Kamiyaran city, Kurdistan province), *Quantitative Geomorphology Research*, 6 (3), 15-36. [In Persian]
 40. Chen, W., Chai, H., Zhao, Z., Wang, Q., & Hong, H. (2016). Landslide susceptibility mapping based on GIS and support vector machine models for the Qianyang County, China. *Environmental Earth Sciences*, 75 (6).
 41. Fix, E., & Hodges, J. (1951). Discriminatory Analysis, Non Parametric Discrimination: Consistency Properties". *Technical Report 4, USA, School of Aviation Medicine Randolph Field Texas*.
 42. Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160 (1), 3-24.
 43. Annathurai, K. S., & Angamuthu, T. (2022). Sorensen-dice similarity indexing based weighted iterative clustering for big data analytics. *Int. Arab J. Inf. Technol.* 19 (1), 11-22.
 44. Greenwell, B. M. (2017). pdp: An R package for constructing partial dependence plots. *R J.* 9 (1), 421.
 45. Shirvani, Z., Abdi, O., & Buchroithner, M. (2019). A synergetic analysis of Sentinel-1 and-2 for mapping historical landslides using object-oriented Random Forest in the Hyrcanian forests. *Remote Sensing*, 11 (19), 2300.
 46. Prathom, K., & Sujitapan, C. (2024). Performance of logistic regression and support vector machine conjunction with the GIS and RS in the landslide susceptibility assessment: Case study in Nakhon Si Thammarat, southern Thailand. *Journal of King Saud University-Science*, 103306.
 47. Sun, W., Tian, Y., Mu, X., Zhai, J., Gao, P., & Zhao, G. 2017. Loess landslide inventory map based on GF-1 satellite imagery. *Remote Sensing*, 9, 314.
 48. Pourghasemi, H. R. (2022). GIS, Remote Sensing, and Spatial Modeling in R. *Shiraz University. Press*, 254p. [In Persian]

