

علوم و تکنولوژی محیط زیست، دوره بیست و شش، شماره دوازده، اسفند ماه ۱۴۰۳ (۳۷-۲۳)

تهیه نقشه پیش‌بینی مناطق مستعد سیل به کمک یادگیری ماشین جنگل تصادفی (مطالعه موردی: حوزه آبخیز سیاه خور استان کرمانشاه)

علی کیانی^۱

بهارک معتمدوزیری^{۲*}

bmvaziri@gmail.com

محمدرضا خالقی^۳

حسن احمدی^۴

تاریخ پذیرش: ۱۴۰۳/۱۰/۱۹

تاریخ دریافت: ۱۴۰۳/۵/۹

چکیده

زمینه و هدف: سیل ناگهانی یکی از پویاترین بلایای طبیعی در نظر گرفته می‌شود که برای به حداقل رساندن خسارات اقتصادی، اثرات نامطلوب و پیامدهای آن با ترسیم حساسیت به سیل باید تدابیری اتخاذ شود. بنابراین تهیه نقشه حساسیت مناطق مستعد سیل گامی مهم در راستای مدیریت سیلاب است. به دلیل کمبود اطلاعات در اکثر حوزه‌های آبخیز کشور، بسیاری از محققان برای مطالعه‌های هیدرولوژیکی و سیل‌گیری از تجزیه و تحلیل‌های مکانی استفاده می‌کنند. بر این اساس، شناسایی مهم‌ترین عوامل مؤثر بر ایجاد و تشدید وقوع سیل و هم‌چنین تهیه نقشه حساسیت‌پذیری آن می‌تواند یکی از مهم‌ترین راهکارها در راستای کاهش خطر سیل باشد. در پژوهش حاضر، حوزه آبخیز سیاه خور در شمال شرق شهر اسلام آباد غرب (استان کرمانشاه)، برای پیش‌بینی مناطق مستعد سیل و شناسایی عوامل مهم و مؤثر بر وقوع آن انتخاب شد.

روش بررسی: برای شناسایی مناطق مستعد سیل در این منطقه مورد مطالعه، الگوریتم یادگیری ماشین (ML) روش جنگل تصادفی در محیط پایتون مورد استفاده قرار گرفت. نقشه پراکنش سیلاب‌های گذشته به‌منظور پیش‌بینی سیلاب در آینده تهیه شد. یافته‌ها: ۵۳ رخدادهای سیلاب در تاریخ آذر ماه ۱۴۰۳ در منطقه ثبت شده بود و ۴۹ منطقه نیز به عنوان مناطق غیر سیلابی در نظر گرفته شد، ۷۰٪ به‌منظور مدل‌سازی و ۳۰٪ به‌منظور اعتبارسنجی استفاده شد. با بررسی مطالعات قبلی و پیمایش منطقه مورد مطالعه ۸ عامل

۱- دانشجوی دکتری، گروه مهندسی طبیعت، دانشکده منابع طبیعی و محیط زیست، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران.
۲- دانشیار، گروه مهندسی طبیعت، دانشکده منابع طبیعی و محیط زیست، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران.* (مسئول مکاتبات)
۳- دانشیار، گروه کشاورزی و منابع طبیعی، واحد تربت جام، دانشگاه آزاد اسلامی، تربت جام، ایران.
۴- استاد گروه احیای مناطق خشک و کوهستانی، دانشکده منابع طبیعی، دانشگاه تهران، کرج، ایران.

مؤثر (شیب، تراکم زهکشی، کاربری اراضی، فاصله از رودخانه، خصوصیات زمین شناسی، بارندگی، جهت شیب و ارتفاع) به منظور پهنه‌بندی سیلاب انتخاب و تهیه شد. به منظور اعتبارسنجی و ارزیابی کارایی مدل‌ها نیز از منحنی تشخیص عملکرد نسبی (ROC) استفاده شد.

بحث و نتیجه‌گیری: نتایج نشان داد که از بین ۸ عامل مؤثر بر پهنه‌های سیلابی، سه عامل بارندگی (۰/۳۵)، فاصله از رودخانه (۰/۲۷) و ارتفاع (۰/۲۱) به ترتیب بیش‌ترین تأثیر را در سیل‌گیری منطقه مورد مطالعه دارند. همچنین نتایج ارزیابی خروجی مدل‌ها نشان داد که مقدار AUC در مدل RF، ۰/۹۸ بود که نشان‌دهنده کارایی و دقت بالای مدل RF در تهیه نقشه حساسیت به وقوع سیل در منطقه مورد مطالعه می‌باشد. بیش‌ترین مساحت حساسیت به سیل در مدل RF مربوط به طبقه خیلی کم است. نتایج حاصل نشان داد که استفاده از الگوریتم یادگیری ماشین (ML) روش جنگل تصادفی می‌تواند به طور مؤثر در تجزیه و تحلیل خطر سیلاب کارآمد باشد.

واژه‌های کلیدی: سیل‌گیری، منحنی تشخیص عملکرد نسبی، مکان یابی، اسلام آباد غرب.

Preparation of flood-prone areas prediction map using random forest machine learning in the Siahkhor Watershed of Kermanshah Province

Ali Kiani¹

Baharak Motamedvaziri^{2*}

bmvaziri@gmail.com

MohammadReza khaleghi³

Hassan Ahmadi⁴

Date of Acceptance: January 8, 2025

Date of Submission: July 30, 2024

Abstract

Background and Objective: Flash floods are considered among the most dynamic natural disasters, necessitating measures to minimize economic damages, adverse effects, and their consequences by enhancing flood sensitivity. Therefore, mapping the sensitivity of areas prone to flooding is a critical step towards flood management. Due to the scarcity of information in most of the country's watershed areas, many researchers rely on spatial analyses for hydrological studies and flood mapping. Consequently, identifying the most significant factors influencing the creation and intensification of floods, as well as mapping their sensitivity, can be one of the most important strategies for reducing flood risk. In the current study, the Siah Khor watershed in the northeast of Islam Abad-e Gharb (Kermanshah province) was selected for predicting flood-prone areas and identifying key factors affecting their occurrence.

Material and Methodology: To identify flood-prone areas in this study region, the Random Forest machine learning (ML) algorithm was employed within a Python environment. A map depicting the distribution of past flood events was created to predict future flooding.

Findings: Fifty-three flood events were recorded in the area, and forty-nine regions were identified as non-flood zones; 70% of the data was used for modeling and 30% for validation purposes. After reviewing previous studies and surveying the study area, eight influential factors (slope, drainage density, land use, distance from the river, geological characteristics, rainfall, slope direction, and elevation) were selected for flood zoning. The Relative Operational Characteristic (ROC) curve was utilized for model validation and efficiency evaluation.

1- PhD Candidate, Department of Nature Engineering, Faculty of Natural Resources and Environment, Science and Research Branch, Islamic Azad University, Tehran, Iran.

2- Associate Professor, Department of Nature Engineering, Faculty of Natural Resources and Environment, Science and Research Branch, Islamic Azad University, Tehran, Iran. **(Corresponding Author)*

3- Associate Professor, Department of Agriculture and Natural Resources, Torbate- Jam Branch, Islamic Azad University, Torbate- Jam, Iran.

4- Professor, Department of Reclamation of Arid and Mountainous Regions, Faculty of Natural Resources, University of Tehran, Karaj, Iran.

Predicting Flood-Prone Areas Using Machine Learning Methods in the Siahkhor Watershed of Kermanshah Province

Discussion and Conclusion: The results indicated that among the eight factors affecting flood zones, rainfall (0.35), distance from the river (0.27), and elevation (0.21) have the most significant impact on flooding in the study area, respectively. Furthermore, the evaluation of model outputs revealed that the Area Under the Curve (AUC) value for the Random Forest (RF) model was 0.98, demonstrating the high efficiency and accuracy of the RF model in mapping flood sensitivity in the study area. The largest area of flood sensitivity in the RF model corresponds to the very low category. The findings suggest that employing the Random Forest machine learning (ML) algorithm can be effectively used in analyzing flood risk.

Keywords: Flood, Siahkhor Watershed, Site Selection, Machine Learning Algorithm.

مقدمه

پدیده سیل یکی از پدیده‌های پرتکرار دهه گذشته در ایران است که خسارت‌های مالی و تلفات جانی بسیاری را سبب شده است. یکی از سه مخاطره طبیعی و اصلی ایران، رخداد سیل است و بدون شک حداقل در سال در یک نقطه از این کشور سیلاب بزرگی رخ می‌دهد. سیل در ایران خسارت‌های بسیاری از نظر اقتصادی، نابودی محیط زیست، منابع طبیعی و مسکونی و تلفات جانی وارد می‌نماید. در سال‌های گذشته حدود ۷۰٪ اعتبارهای سالانه‌ی طرح کاهش اثر بلاهای طبیعی و ستاد حوادث غیرمترقبه صرف جبران خسارت‌های ناشی از سیل شده است (۱). یکی از زمینه‌های پژوهش برای کنترل مخاطره سیل، شناسایی نقاط و پهنه‌های بحرانی منطقه است. زیرا نبود شناخت کافی این نقاط، سبب خسارت‌های مالی و جانی سنگین می‌شود. بنابراین شناسایی مناطق حساس به وقوع سیلاب جهت فراهم کردن مدیریت لازم برای کاهش خسارت سیل ضروری می‌باشد (۲). اطلاعات مربوط به پیش‌بینی موقعیت سیلاب‌های آینده برای دولت و ادارات و سازمان‌های ذی ربط بسیار ضروری است. هیدرولوژیست‌ها مدل‌های زیادی را برای تهیه نقشه مناطق مستعد سیل انجام داده اند لیکن بسیاری از این مدل‌ها یا به داده‌های زیادی نیاز دارند و یا کالبره کردن آن‌ها سخت می‌باشد و بعضی از مدل‌ها نیز نیازمند فهم فرایندهای فیزیکی داخل حوزه آبخیز هستند. در سال‌های اخیر تکنیک‌های پیشرفته تری مانند روش‌های یادگیری ماشین و داده کاوی در تهیه نقشه مناطق مستعد سیل مورد استفاده قرار گرفته است (۳).

یوسفی و همکاران (۴) پتانسیل سیل خیزی را با استفاده از مدل‌های یادگیری ماشین CART^۲، GLM^۳ و GAM^۱ در حوزه کشکان تعیین کردند. دقت ۰/۹۱ مدل CART نشان-دهنده دقت عالی این مدل برای حوزه کشکان است. این مدل، مساحت بیشتری از حوزه را تحت شرایط پتانسیل بالا و متوسط

خطر سیل‌گیری نشان می‌دهد. چزگی و همکاران (۵) مناطق مستعد سیل خیزی را با استفاده از مدل‌های یادگیری ماشین (مدل جنگل تصادفی، درخت رگرسیون پیشرفته و ماشین بردار پشتیبان) در حوزه آبخیز شهرستانک شهرستان خوسف مشخص کردند. منحنی تشخیص عملکرد نسبی (ROC) ۰/۹۹، ۰/۹۸، ۰/۹۵ به ترتیب دارای دقت بالاتری هستند. رضایی و همکاران (۶) مدل‌سازی مکانی و تهیه نقشه پتانسیل سیل‌گیری را با استفاده از الگوریتم‌های یادگیری ماشین در استان بوشهر انجام دادند. شاخص سطح زیر منحنی نشان داد که مدل ماشین بردار پشتیبان (۰/۸۶)، مدل جمعی تعمیم‌یافته (۰/۸۵) و مدل جنگل تصادفی (۰/۸۸) دارای دقت خیلی خوب هستند با توجه به نتایج به‌دست‌آمده، مدل جنگل تصادفی دقت بیشتری را نسبت به سایر مدل‌ها دارد. چن^۷ و همکاران (۷) حساسیت به سیل را با استفاده از روش‌های داده محور درخت بیز ساده^۵، درخت تصمیم متناوب^۶ و روش‌های جنگل تصادفی^۴ در منطقه کوانان، چین مدل کردند. نتایج نشان داد که روش RF یک مدل کارآمد و قابل اعتماد در ارزیابی حساسیت سیل می‌باشد.

باند^۸ و همکاران (۸) برای شناسایی مناطق مستعد سیل در حوزه آبخیز کلوان در استان مرکزی، ایران، پنج الگوریتم یادگیری ماشین (ML^۱)، (درخت رگرسیون تقویت شده (BRT^۲))، جنگل تصادفی (RF)، جنگل تصادفی موازی (PRF)، جنگل تصادفی منظم (RRF^۳)، و درختان بسیار تصادفی (ERT^۴) مورد آزمایش قرار گرفتند. پانزده متغیر اقلیمی و ژئومحیطی به عنوان ورودی مدل‌های حساسیت به سیلاب استفاده شد. نتایج نشان داد که ERT بهینه ترین مدل با مقدار مساحت زیر منحنی (AUC) ۰/۸۲ بوده است. عیسی زاده و همکاران (۹) مناطق سیل خیز را با استفاده از تکنیک‌های یادگیری ماشین در استان سمنگان افغانستان شبیه سازی کردند. این مطالعه

8- Band

9- Machine Learning

10- Boosted Regression Tree

11- Random Forest

12- Parallel Random Forest

13- Regularized Random Forest

14- Extremely Randomized Trees

1- classification and regression trees

2- generalized linear models

3- Generalized General Model

4- Chen

5- naïve Bayes tree

6- alternating decision tree

7- random forest

نشان داد که ترکیب GIS و شبکه‌های عصبی مصنوعی منجر به عملکرد قابل قبولی برای شبیه‌سازی و مدل‌سازی مناطق مستعد سیل در مکان‌های جغرافیایی مختلف می‌شود.

مواد و روش‌ها

این پژوهش محدوده جغرافیایی حوضه سیاه خور با مساحت ۹۸۸۴ هکتار در شمال شرق شهر اسلام آباد غرب و در فاصله ۱۰ کیلومتری آن در استان کرمانشاه به خود اختصاص داده است. محدوده حوضه مورد مطالعه از $30^{\circ}46'$ تا $36^{\circ}46'$ طول شرقی و $49^{\circ}07'$ تا $48^{\circ}16'$ عرض شمالی می‌باشد. مهمترین آبراهه موجود در منطقه رودخانه چم برزه می‌باشد که حوضه مطالعاتی از سرشاخه‌های اصلی این رودخانه می‌باشد. شکل (۱) موقعیت منطقه مورد مطالعه را در ایران، استان کرمانشاه، شهر اسلام آباد را نشان می‌دهد.

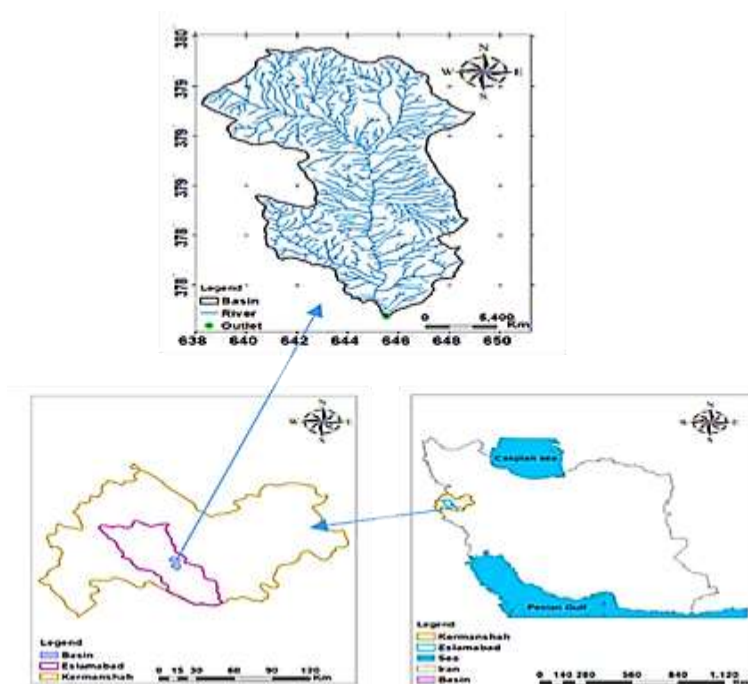
هدف از پژوهش حاضر، مدل‌سازی مکانی و تهیه نقشه حساسیت پذیری سیل و شناسایی عوامل مهم و مؤثر بر وقوع آن در حوضه سیاه خور در شمال شرق شهر اسلام آباد غرب (استان کرمانشاه) با استفاده از روش یادگیری ماشین جنگل تصادفی، می‌باشد. به این منظور ابتدا با استفاده از بازدیدهای گسترده میدانی و با استفاده از اطاعات مناطق سیل خیز دریافتی از سازمان آب منطقه ای استان کرمانشاه، مناطق سیل‌خیز مشخص گردید و متعاقباً نقشه پراکنش سیل‌خیزی تهیه شد. در مرحله بعد عوامل مؤثر بر وقوع سیل‌خیزی از قبیل ارتفاع، جهت جغرافیایی، فاصله از آبراهه، تراکم زهکشی، کاربری اراضی، زمین شناسی، شیب و بارندگی سالانه (میلی‌متر)، در محیط GIS تهیه گردید. برای ارزیابی مدل‌ها از منحنی تشخیص عملکرد نسبی (ROC)

استفاده شد. نقشه نهایی خطر سیل می‌تواند نقش به‌سزایی در تصمیم‌گیری‌های آتی، مدیریت سیل‌خیزی و آمایش سرزمین داشته باشد.

پارامترهای مؤثر در تهیه نقشه پتانسیل سیل

برای تهیه نقشه‌ای که نشان دهد چه مناطقی بیشتر در معرض خطر وقوع سیل هستند، یا برای ساخت مدلی که آسیب‌پذیری در برابر بلایای طبیعی را ارزیابی کند، باید مجموعه‌ای از عوامل مؤثر تعریف و شناسایی شوند. این عوامل برای درک بهتر خطرات بالقوه و میزان آسیب‌پذیری منطقه بسیار مهم هستند و به ایجاد مدلی جامع کمک می‌کنند که می‌تواند احتمال وقوع بلایایی مانند سیل را پیش‌بینی یا ارزیابی کند. این کار شامل تحلیل داده‌های جغرافیایی، محیطی و زیرساختی است تا مدل به‌طور دقیق شرایط واقعی را منعکس کند.

مجموعه عوامل مؤثر بر سیلاب شامل ۸ فاکتور شیب، تراکم زهکشی، کاربری اراضی، فاصله از رودخانه، خصوصیات زمین شناسی، بارندگی، جهت شیب و ارتفاع بعنوان متغیرهای مستقل در نظر گرفته شد (شکل ۲). در این تحقیق تعداد ۵۳ مورد مناطقی که قبلاً دچار سیل شده بود و در سازمان آب منطقه ای کرمانشاه ثبت شده است، انتخاب و ۴۸ نقطه نیز که فاقد سیل خیزی بود برای آموزش انتخاب شد. ۷۰ درصد به‌منظور مدل‌سازی و ۳۰ درصد به‌منظور اعتبارسنجی استفاده شد. نقاط غیر سیل با استفاده از بازدیدهای میدانی، نقشه توپوگرافی و گوگل ارث، با توجه به مناطقی نظیر تپه‌ها و کوه‌ها که سیلاب قادر به پیشروی در آنجا نیست، انتخاب گردید (شکل ۳).



شکل ۱- موقعیت شهر اسلام آباد در استان کرمانشاه

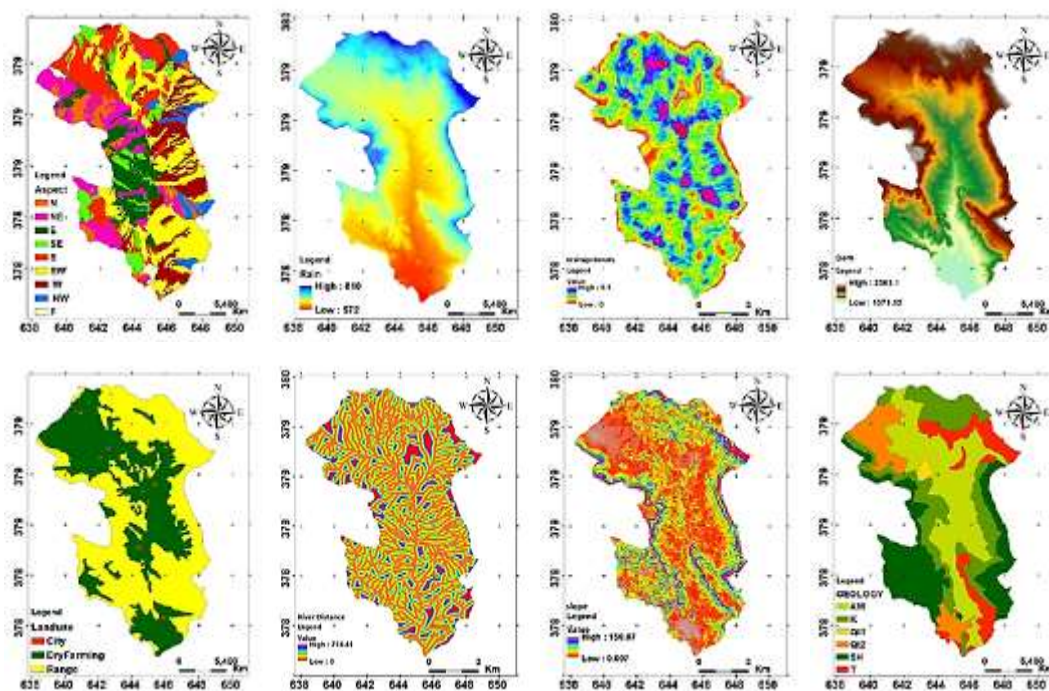
Figure 1. Location of the Islamabad city in Kermanshah province

خاک نسبتاً غیر قابل نفوذ هستند، منجر به ایجاد حجم بالایی از رواناب سطحی می‌شوند (۱۴).

عامل بارش به عنوان اصلی ترین عامل ایجاد سیل در یک منطقه ای تا بارشی نباشد سیلابی (بارشهای با شدت و مدت طولانی بیش‌تر مدنظر است) هم رخ نمی‌دهد (۱۵). با افزایش ارتفاع مقدار بارندگی نیز افزایش می‌یابد و با افزایش بارندگی در منطقه رواناب حاصل از آن نیز شدت می‌گیرد (۱۶). جهت شیب بر پتانسیل سیل‌خیزی منطقه نقش مهمی را ایفا می‌کند، جهات شیب در هدایت و پخش رواناب‌ها به بخش‌های مختلف نقش برجسته‌ای دارد (۱۵). دامنه‌های شمالی و غربی به دلیل دریافت بارش زیاد، ماندگاری طولانی مدت برف و داشتن رطوبت بیشتر نسبت به دامنه‌های جنوبی و شرقی پتانسیل سیل‌خیزی بالاتری دارند و دامنه‌های جنوبی و شرقی، پتانسیل سیل‌خیزی کم‌تری دارند (۱۷). تغییر کاربری اراضی، قطع بی‌رویه درختان و چرای بیش از حد دام‌ها باعث کاهش پوشش گیاهی زمین‌ها شده‌اند. این تغییرات منجر به کاهش قدرت نفوذ آب در خاک می‌شود و در نتیجه، احتمال وقوع سیلاب را افزایش می‌دهد. پوشش گیاهی

پارامتر شیب از عوامل موثر و مهم در بروز سیل در یک حوضه می‌باشد. با توجه به این امر که افزایش شیب بدون در نظر گرفتن سایر پارامترها باعث افزایش سرعت رواناب و زمان تمرکز میشود که در پهنه بندی سیل خیزی از اهمیت بالایی دارد (۱۰). فعالیت‌های انسانی مانند ساخت‌وسازهای پرتراکم در نزدیکی رودخانه‌ها و استخراج غیر اصولی و بی‌رویه مصالح از بستر رودخانه‌ها، باعث افزایش خطر وقوع سیل‌های ناشی از مداخله انسانی در آبراهه‌های حوزه آبخیز و رودخانه‌ها می‌شود. این مداخلات می‌توانند تعادل طبیعی رودخانه‌ها را بر هم بزنند و منجر به افزایش احتمال سیلاب‌های مخرب شوند (۱۱). آبراهه‌ها در حوضه، عمل تخلیه جریانهای سطحی را انجام می‌دهند و با فاصله گرفتن از آن، خطر ایجاد سیلاب کاهش می‌یابد (۱۲). تراکم زهکشی نقش اساسی در زمان تمرکز و اوج سیلاب دارد (۱۳). نوع سنگ و پوشش خاک تأثیر به سزایی بر ظرفیت نفوذپذیری دارد. خاک یا سنگ‌هایی که قابل نفوذ هستند، اجازه می‌دهند آب به داخل زمین نفوذ کند و روند تخلیه آن به سمت آبراهه اصلی را به تأخیر می‌اندازند، که در نتیجه، رواناب سطحی کاهش می‌یابد. در مقابل، حوضه‌هایی که دارای سنگ بستر یا

نقش مهمی در حفظ خاک و افزایش نفوذپذیری آن دارد و از این طریق، به کاهش خطر سیلاب کمک می‌کند (۱۱).



شکل ۲- عوامل مؤثر بر تهیه نقشه حساسیت به سیل

Figure 2. Factors Influencing Flood Susceptibility Map Preparation

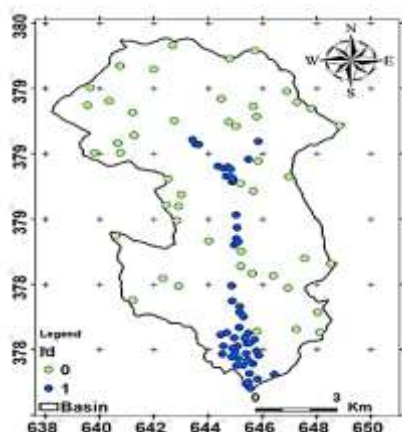
الگوریتم جنگل تصادفی

یکی از مهم‌ترین ویژگی‌های جنگل‌های تصادفی، توانایی بالای آن‌ها در ارزیابی اهمیت متغیرها است، به طوری که مشخص می‌کند هر متغیر چه نقشی در پیش‌بینی پاسخ دارد. این الگوریتم در واقع نوعی الحاقی از درخت‌های رگرسیون و طبقه‌بندی است که اولین بار توسط لئو بریمن در سال ۲۰۰۱ معرفی و توسعه داده شد. مدل پیش‌بینی کننده جنگل تصادفی بر پایه میانگین‌گیری از نتایج به‌دست‌آمده از تمامی درختان تصمیم موجود در جنگل استوار است و برای بسیاری از مجموعه داده‌ها طبقه‌بندی با دقت بالایی را ارائه می‌دهد (۲۱). در این فرآیند، بردار ورودی توسط هر درخت در جنگل پردازش می‌شود و خروجی به صورت برچسب‌های کلاسی است که بر اساس اکثریت آرا تعیین می‌شوند (۱۹، ۲۲). مزایای جنگل‌های تصادفی را می‌توان به صورت زیر خلاصه نمود: (۱) صحت؛ مدل‌های جنگل تصادفی اغلب نسبت به یک تک درخت دارای صحت قابل توجه بیشتری می‌باشند. صحت آن، اغلب قابل رقابت با بهترین روش‌های

الگوریتم جنگل‌های تصادفی به عنوان یکی از پیشرفته‌ترین روش‌های مبتنی بر درخت شناخته می‌شود که از مجموعه‌ای از درخت‌های تصمیم برای انجام وظایف دسته‌بندی و رگرسیون استفاده می‌کند. این الگوریتم به وسیله ترکیب نتایج حاصل از تعداد زیادی درخت تصمیم، دقت و پایداری مدل را افزایش می‌دهد و از آن برای کاهش واریانس و پیشگیری از بیش‌برازش (Overfitting) بهره می‌گیرد. به همین دلیل، جنگل‌های تصادفی به عنوان یکی از قدرتمندترین الگوریتم‌های یادگیری ماشین محسوب می‌شوند و در بسیاری از کاربردها، به ویژه در زمینه داده‌های پیچیده و بزرگ‌مقیاس، عملکرد بسیار خوبی از خود نشان می‌دهند (۱۸). روش‌های درخت پایه روش‌های آماری ناپارامتری (مدل آزاد) برای اجرای آنالیز کلاس بندی و آنالیز رگرسیون با استفاده از الگوریتم افرازهای بازگشتی می‌باشند (۱۹، ۲۰).

درختان با سرعت زیاد رویانده می‌شوند، چرا که تعداد کمی متغیر در هر زمان مورد استفاده قرار می‌گیرد. (۵) توانایی مدلسازی اثرات متقابل پیچیده میان متغیرهای پیش‌بینی کننده (۶) انعطاف پذیری برای انجام انواع مختلفی از تجزیه و تحلیل داده‌های آماری، از جمله رگرسیون و طبقه‌بندی (۲۳).

جایگزین می‌باشد. (۲) مقاومت به بیش برآزش؛ رویاندن تعداد زیاد درخت خطای بیش برآزش را ایجاد نمی‌کند. (۳) انتخاب پیشگوها از بین تعداد زیادی متغیرهای کاندید تحلیلگر، نیازی به انتخاب متغیر یا تقلیل داده ندارد؛ جنگل تصادفی به طور خودکار، در نهایت بهترین متغیرهای پیشگو را شناسایی می‌کند. (۴) سرعت؛



شکل ۳- نقشه مکان‌های سیلابی و غیرسیلابی برای آموزش و اعتبارسنجی مدل جنگل تصادفی

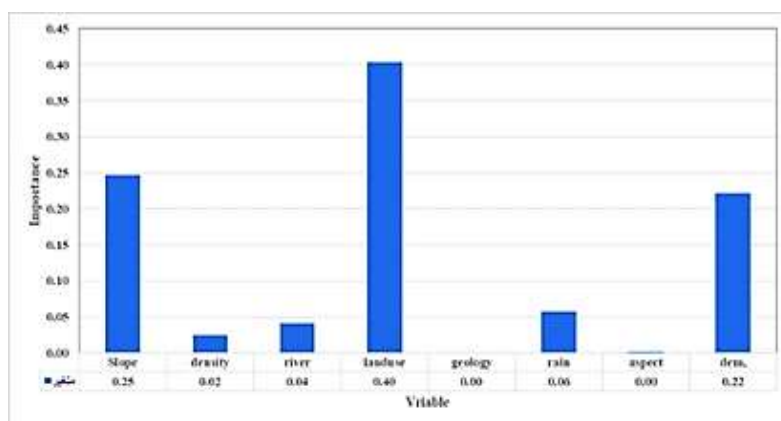
Figure 3. Spatial map of flood and non-flood locations for random forest model training and validation

نتایج

عوامل متعددی در یک حوزه آبخیز بر جریان آب در سطح حوضه اثرگذار می‌باشند که هرکدام از این عوامل پتانسیل متفاوتی در تغییر جریان و رخداد سیلاب دارند. تعیین مهم‌ترین و مؤثرترین فاکتورهای تأثیرگذار بر سیلاب جهت انجام مدل‌سازی بسیار ضروری می‌باشد (۲۲). در این پژوهش اولویت بندی فاکتورهای مؤثر بر سیلاب در حوزه آبخیز سیاه خور در شکل (۴) نشان داده شده است. مهم‌ترین متغیرها برای پیش‌بینی حساسیت به سیلاب، کاربری اراضی (۰/۳۵)، شیب (۰/۲۵)، ارتفاع (۰/۲۲) بود. سایر متغیرها در مدل جنگل تصادفی پیش‌بینی شده با اهمیت متوسط تا پایین‌تر برای ارزیابی حساسیت به سیل همراه بودند. شکل (۵) نقشه حساسیت به سیل تولید شده توسط مدل جنگل تصادفی را نشان می‌دهد. نقشه سیل خیزی به ۵ کلاس شامل: طبقه حساسیت خیلی کم، کم، متوسط، زیاد و خیلی زیاد تقسیم گردید (شکل ۶). حدود ۵۲ درصد منطقه در طبقه حساسیت کم و خیلی کم قرار می‌گیرد و حدود ۳۵ درصد در طبقه حساسیت خیلی زیاد تا زیاد قرار دارد.

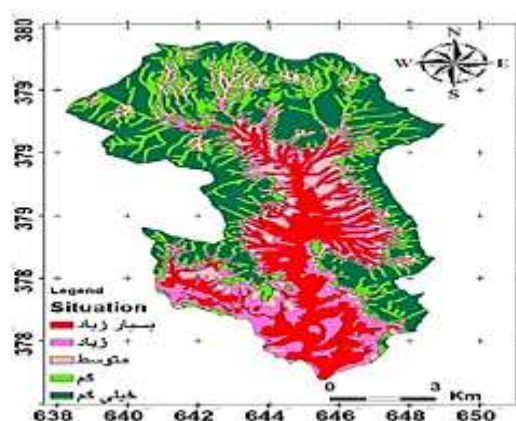
ارزیابی مدل با استفاده از منحنی ROC

برای ارزیابی نتایج و دقت مدل‌های مورد استفاده، از منحنی تشخیص عملکرد نسبی (ROC) و مساحت زیر این منحنی استفاده می‌شود. مساحت زیر منحنی ROC که به AUC معروف است، بیانگر توانایی مدل در پیش‌بینی صحیح وقوع و عدم وقوع رویدادها می‌باشد. در یک مدل ایده‌آل، بیشترین مساحت زیر منحنی وجود دارد و مقادیر AUC بین ۰/۵ تا ۱ متغیر است. اگر مدلی نتواند وقوع رویدادها را بهتر از حد احتمال تصادفی پیش‌بینی کند، مقدار AUC آن برابر ۰/۵ است. زمانی که منحنی ROC مساحت زیر منحنی برابر یک داشته باشد، نشان‌دهنده بهترین دقت نقشه پهنه‌بندی تهیه شده است. همبستگی کیفی- کمی بین سطح زیر منحنی و ارزیابی پیش‌بینی به صورت زیر طبقه‌بندی می‌شود: ۰/۹-۱/۰، عالی، ۰/۸ تا ۰/۹، بسیار خوب، ۰/۷ تا ۰/۸، خوب، ۰/۶ تا ۰/۷، متوسط و ۰/۵ تا ۰/۶، ضعیف می‌باشد (۲۴).



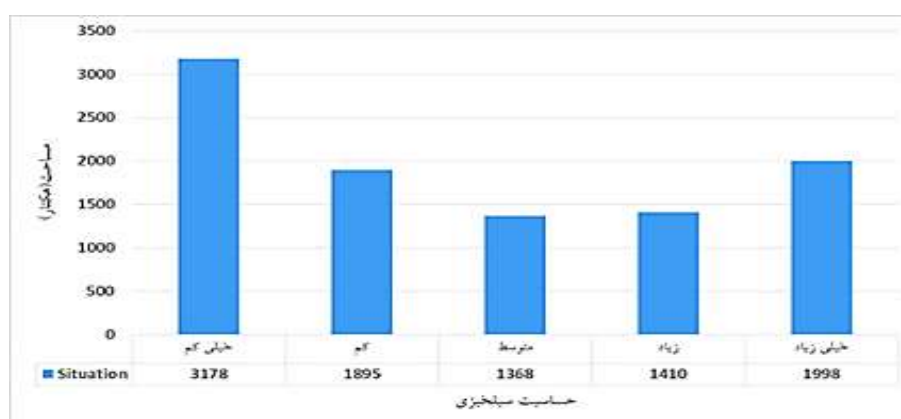
شکل ۴- نمودار اولویت‌بندی عوامل مؤثر بر خطر سیل با استفاده از روش جنگل تصادفی در حوزه آبخیز سیاه خور

Figure 4. Diagram of prioritizing factors affecting flood risk using the random forest method in the Siahkhor watershed



شکل ۵- نقشه حساسیت به سیل تولید شده توسط مدل جنگل تصادفی در حوزه آبخیز سیاه خور

Figure 5. Flood sensitivity map generated random forest model in the Siahkhor watershed



شکل ۶- نمودار طبقات خطر سیل‌گیری در حوزه آبخیز سیاه خور

Figure 6. Diagram of Flood risk classes in the Siahkhor watershed

ارزیابی مدل

اعتبارسنجی برابر با ۰/۶۷ بدست آمد. نتایج مربوط به اعتبارسنجی مدل جنگل تصادفی شکل (۷) نمایش داده شده است. نتایج اعتبارسنجی مدل استفاده شده با استفاده از مدل جنگل تصادفی نشان داد که سطح زیر منحنی (ROC) برابر با ۰/۹۵ می‌باشد که با نتایج تیموری و همکاران (۲۳)، چزگی و همکاران (۵)، کیانی اصل و همکاران (۲۶) همخوانی دارد.

Mean Absolute Error: 0.16

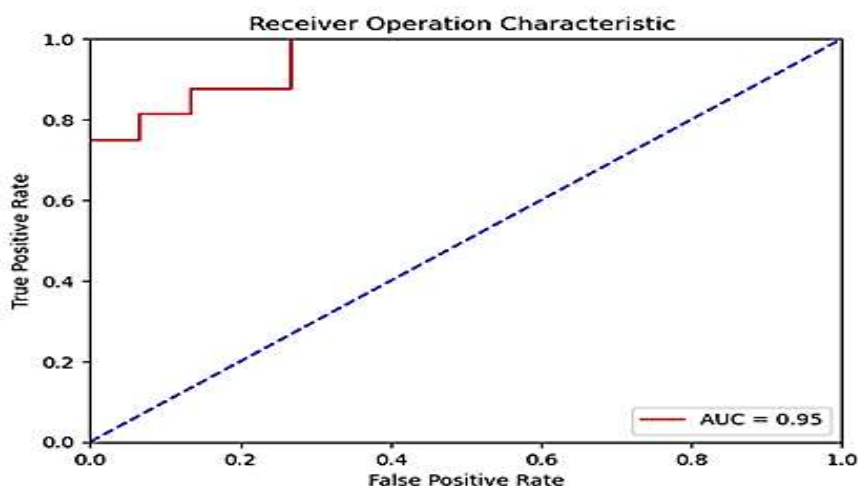
Mean Squared Error: 0.083

Root Mean Squared Error: 0.28

0.98 R^2 Training score:

R^2 Validation score: 0.67

به‌منظور ارزیابی مدل تهیه شده، از منحنی تشخیص عملکرد نسبی (ROC) استفاده شد. منحنی ROC یک نمایش گرافیکی است که تعادل بین نرخ خطای مثبت و منفی را برای هر مقدار احتمالی از نقاط برش نشان می‌دهد. مساحت زیر منحنی ROC (AUC)، توانایی سیستم در پیش‌بینی صحیح وقوع و عدم وقوع رویدادها، مانند وقوع یا عدم وقوع سیل، را بیان می‌کند. با ارائه دیدگاهی کمی از دقت مدل، کمک می‌کند تا عملکرد مدل در تمایز میان حالات مختلف به خوبی ارزیابی شود (۲۲). اعتبارسنجی نقشه‌های تهیه شده، گامی اساسی در توسعه و تعیین مناطق حساس و تعیین کیفیت آنها می‌باشد. میزان ضریب تبیین در دوره آموزش برابر با ۰/۹۸ و در دوره



شکل ۷- منحنی ROC و سطح زیر منحنی حساسیت سیل خیزی با استفاده از روش جنگل تصادفی در حوزه آبخیز سیاه

خور

Figure 7. ROC curve and AUC of flood sensitivity using random forest method in the Siahkhor watershed area

بحث و نتیجه‌گیری

ماشینی (جنگل تصادفی)، انجام گرفته است. در این تحقیق برای تعیین مناطق دارای پتانسیل خطر سیل در منطقه مورد مطالعه از ۸ پارامتر تأثیرگذار (شیب، تراکم زهکشی، کاربری اراضی، فاصله از رودخانه، خصوصیات زمین شناسی، بارندگی، جهت شیب و ارتفاع) و روش‌های جنگل تصادفی در محیط پایتون، استفاده شد. در این راستا ۵۳ مورد مخاطره منطقه سیل‌خیز ثبت شد. به‌طور تصادفی ۷۰ درصد برای فرآیند

سیل یکی از مخرب‌ترین مخاطرات طبیعی است که هر ساله موجب خسارت به زندگی و هم اموال می‌شود (۲۷). بنابراین، شناسایی زودهنگام حوضه‌های مستعد سیل ناگهانی را می‌توان با استفاده از مدل‌های پیشرفته یادگیری ماشین برای مدیریت بلایای سیل انجام داد (۲۵). این پژوهش باهدف تهیه نقشه پهنه-بندی پتانسیل سیل‌خیزی در حوزه آبخیز سیاه‌خور در اسلام‌آباد شهر کرمانشاه، با استفاده از الگوریتم‌های یادگیری

مدل سازی و ۳۰ درصد برای ارزیابی مدل های تهیه شده تقسیم بندی شد. اهمیت فاکتورهای نقشه های سیلاب پذیری محاسبه و در نهایت نقشه ها اعتبارسنجی شدند.

مهم ترین متغیرها برای پیش بینی حساسیت به سیلاب، کاربری اراضی (۰/۳۵)، شیب (۰/۲۵)، ارتفاع (۰/۲۲) بود. محققان مختلفی از جمله رستمی و همکاران (۲۵)، آوند و همکاران (۲۲) و ستارزاده و همکاران (۱)، رضایی و همکاران (۶)، عرب (۲۸) همین عوامل را به عنوان حساس ترین عوامل در مناطق تحت تأثیر سیل تشخیص دادند. میزان ضریب تبیین در دوره آموزش برابر با ۰/۹۸ و در دوره اعتبارسنجی برابر با ۰/۶۷ بدست آمد. نتایج اعتبارسنجی مدل استفاده شده با استفاده از مدل جنگل تصادفی نشان داد که سطح زیر منحنی (ROC) برابر با ۰/۹۵ می باشد. نقشه سیل خیزی به ۵ کلاس شامل: طبقه حساسیت خیلی کم، کم، متوسط، زیاد و خیلی زیاد تقسیم گردید. نتایج نشان داد که حدود ۵۲ درصد منطقه در طبقه حساسیت کم و خیلی کم قرار می گیرد و حدود ۳۵ درصد در طبقه حساسیت خیلی زیاد تا زیاد قرار دارد.

بیشتر مناطق مستعد سیل در این حوضه سازند زمین شناسی امیران و کشکان، ارتفاع پایین، فواصل نزدیک به رودخانه، عمده اراضی بصورت کشاورزی قرار دارد. از لحاظ زمین شناسی بیشتر بر روی سازندهای امیران که تناوب مارنهای سبز تا خاکستری متمایل به سیاه با میان لایه های آهک مارنی دارد و از نظر فرسایشی بسیار حساس با نفوذپذیری کم (رخساره های کم تراوا) می باشد که تراکم شبکه های زهکشی (الگوی آبراهه برروی آنها زیاد) بر روی آن موید این موضوع می باشد. قسمت کمی از عرصه مستعد سیل نیز بر روی نیز نهشته های تخریبی ناپیوسته با منشأ آبرفتی کواترنر (Qt2) می باشد که نسبتاً جوان و رخساره های این گروه به دلیل برخورداری از بافت متوسط تا درشت دانه و منفصل بودن ذرات، از قابلیت نفوذ پذیری بالائی برخوردار می باشند که از جنبه های مختلف در فعالیت های آبخیزداری از جایگاه و اهمیت ویژه ای برخوردارند. احتمال وقوع سیل گیری در مناطق پایین دست به دلیل ارتفاع و شیب کم تر، بیش تر از مناطق بالادست، است. کاهش فاصله با حریم اصلی رودخانه و در معرض قرار گرفتن املاک، اراضی کشاورزی از دیگر

عوامل مؤثر در افزایش خسارت در مواقع سیلابی در این مناطق بوده است. در حوزه آبخیز مورد مطالعه با افزایش ارتفاع برخلاف مناطق جنوبی مقدار بارندگی افزایش مییابد و در ارتفاعات پایین تر و نزدیک به خروجی حوضه، مقدار بارندگی کاهش می یابد. از طرفی احتمال وقوع سیل گیری در مناطق پایین دست نیز بیش تر می باشد، بنابراین یک رابطه مستقیم بین افزایش بارندگی و سیل گیری و خسارت در مناطق پایین دست این حوزه آبخیز نیز وجود دارد که باید مورد توجه قرار بگیرد. در فاصله کمی از خروجی حوضه نیز نیروگاه سیکل ترکیبی دالاهو در شهرستان اسلام آباد که رودخانه خروجی از جنب این نیروگاه عبور می کند.

تهیه نقشه خطر وقوع سیلاب و تعیین اولویت عوامل مؤثر در بروز سیلاب ابزاری مهم در شناسایی نقاط حساس و جلوگیری از تحریک این مناطق است و به آن توجه زیادی می شود. در میان مدل های مختلف، استفاده از مدل های نوین الگوریتم جنگل تصادفی و تلفیق آن با GIS اثر بسزایی در این تحقیقات دارد و می تواند موجب افزایش دقت نقشه های تهیه شده شود. نتایج این پژوهش نشان داد که منطقه تحقیق حدود ۳۵ درصد در طبقه خیلی زیاد و زیاد خطر وقوع سیلاب دارد و مدل جنگل تصادفی ابزار مناسبی برای کمی کردن ارتباط سیلاب و عامل های مؤثر بر آن است. بنابراین می توان گفت تهیه نقشه پهنه بندی سیلاب ابزار مهمی برای کمک به سیاست گذاران، طراحان و مهندسان برای کاهش خسارات ناشی از وقوع آن است. پیشنهاد می شود در پژوهش های آتی این مدل به همراه مدل های دیگر ارزیابی و مقایسه شود. در پایان شایان ذکر است که روش جنگل تصادفی برای حجم داده های زیاد به خوبی عمل می کند و از دقت زیادی برخوردار است. بنابراین استفاده از تکنیک الگوریتم جنگل تصادفی در ارزیابی خطر سیل گرفتگی، مفید و قابل اعتماد است.

ملاحظات اخلاقی پیروی از اصول اخلاق پژوهش

همکاری مشارکت کنندگان در تحقیق حاضر به صورت داوطلبانه و با رضایت آنان بوده است.

حامی مالی

هزینه تحقیق حاضر توسط نویسندگان مقاله تامین شده است.

- Area of Khosef City. jwmseir 2024; 17 (63) : 4, 38-50.
6. Rezaei F, Pourqasmi H, Shamsi F, Khosravi S, Rasoul R. Spatial modeling and preparation of flood potential map using machine learning algorithms (case study: Bushehr province),. Comprehensive Watershed Management. 2024.
 - Qanawati E, Safari A, Beheshti JE, Mansoorian I. Flood potential zoning using CN and AHP hydrological model integration in GIS environment, case study: Balkhlo river basin. Zagros Landscape Geographical Quarterly. 2013;7(25):67-80.
 7. Chen W., Li Y., Xue W., Shahabi H., Li Sh., Hong H., Wang X., Bian H., Zhang Sh, Pradhan B., Bin Ahmad B. Modeling flood susceptibility using data-driven approaches of naïve Bayes tree, alternating decision tree, and random forest methods, Science of The Total Environment, 2020; 701,134979, ISSN 0048-9697.
 8. Band, S.S., Janizadeh, S., Chandra Pal, S., Saha, A., Chakraborty, R.; Melesse, A.M.; Mosavi, A. Flash Flood Susceptibility Modeling Using New Approaches of Hybrid and Ensemble Tree-Based Machine Learning Algorithms. Remote Sens. 2020; 12, 3568.
 9. Isazade, V., Qasimi, A. B., Al Kafy, A., Dong, P., & Mohammadi, M. Simulation of flood-prone areas using machine learning and GIS techniques in Samangan Province, Afghanistan. Geodesy and Cartography, 2024; 50(1), 20–29.
 10. Ghanavati E. A., Safari A., Beheshti Javid E., Mansourian E. (2014). Flood Risk Zonation Using Compilation CN

مشارکت نویسندگان

طراحی و ایده‌پردازی: بهارک معتمدوزیری، علی کیانی
 روش‌شناسی و تحلیل داده‌ها: بهارک معتمدوزیری، علی کیانی،
 محمد رضا خالقی. حسن احمدی
 نظارت نهایی: بهارک معتمدوزیری
 نگارش: علی کیانی

تعارض منافع

بنابر اظهار نویسندگان، مقاله حاضر فاقد هرگونه تعارض منافع بوده است.

References

1. Sattarzadeh I, Amirpoua S, Haji Kandi H, Sadeghian MS. Spatial prediction of flood-prone areas in the Karkheh watershed of Lorestan province using the combined random forest model - weight of evidence. Watershed Research. 2023;36(2):87-103.
2. Fang J, Zhang C, Fang J, Liu M, Luan Y. Increasing exposure to floods in China revealed by nighttime light data and flood susceptibility mapping. Environmental Research Letters. 2021;16(10):104044.
3. Costache R, Arabameri A, Elkhachy I, Ghorbanzadeh O, Pham QB. Detection of areas prone to flood risk using state-of-the-art machine learning models. Geomatics, Natural Hazards and Risk. 2021;12(1):1488-507.
4. Yousefi, H., Yonesi, H. A., Davoudimoghadam, D., Arshia, A., & Shamsi, Z. Determination of Flood potential Using CART, GLM and GAM Machine learning Models. Irrigation and Water Engineering, 2022; 12(4), 84-105.
5. Chezgi J, Poyan S. Determining Flood-Prone Areas Using Machine Learning Models in the Shahrestank Watershed

18. Talebi A, Gudarzi S, Pourqasmi HR. Investigating the possibility of preparing a landslide risk map using random forest algorithm (study area: Sardarabad watershed, Lorestan province). *Natural Environment Hazards*. 2017;7(16):46-54.
19. Breiman L. Random forests. *Machine learning*. 2001;45:5-32.
20. Hastie T, Tibshirani R, Friedman JH, Friedman JH. *The elements of statistical learning: data mining, inference, and prediction*: Springer; 2009.
21. Lee S, Kim J-C, Jung H-S, Lee MJ, Lee S. Spatial prediction of flood susceptibility using random-forest and boosted-tree models in Seoul metropolitan city, Korea. *Geomatics, Natural Hazards and Risk*. 2017;8(2):1185-203.
22. Avand MT, Moradi HR, Ramzanzadeh M. Preparation of flood sensitivity map using two random forest machine learning models and Bayesian generalized linear model. *Environment and Water Engineering*. 2019;6(1):83-95.
23. Timuri M, Vakili Tejareh F, Mazin M, Ramezani M. Comparison of machine learning models in flood sensitivity zoning of Karaj Dam watershed. *Iranian Journal of Watershed Science and Engineering*. 2023;17(61):30-40.
24. Razavi Terme SV, Malek MR. Preparation of flood sensitivity map using the combination of intuitive belief model (EBF) and analysis hierarchy process (AHP) case study: Jahorm city. *Journal of Mapping and Geospatial Information Engineering*. 2016;8(3):1-15.
25. Rostami F, Tavakoli M, Rostami N, Ebrahimi H. Assessing the flood Model and AHP Via GIS (Case Study: River Basin Balekhlo), *Physical Geography Quarterly*, 25, 67-80.
11. Cheraghi Ghale-Seri A, Habibnejad R, Mahmoud SH. Preparation of flood sensitivity map using support vector machine (SVM) model and geographic information system. *Natural Environment Hazards*. 2019; 9(25):61-80.
12. Malekian A, Khozani O, Ashornjad G. Flood risk zoning of Akhtarabad watershed using fuzzy hierarchical analysis method. *Natural Geography Research*. 2013; 4(4):131-52.
13. Yamani M, Enayati M. The relationship between the geomorphological characteristics of basins and the flood potential. *Geographical researches*. 2015; 37(54):47-57.
14. Garde R. *River morphology*; 2006.
15. Mousavi SM, Saeed N, Rakhshani MH, Hosseinzadeh SM. Evaluation and zoning of flood risk using TOPSIS fuzzy logic in GIS environment, case study: Baghmolek watershed. *Natural Environment Hazards*. 2005;5(10):79-97.
16. Nasrin Nejad NE, Rangzen K, Nasraleh K, Saberi A. Zoning of flood potential of Baghan watershed using Fuzzy Hierarchy Analysis (FAHP), remote sensing and geographic information system in natural resources (application). *GIS in Natural Resources Sciences*. 2013; 5(4):15-34.
17. Parastar S. Zonation of flood hazard in the watershed Balkhly chay using by ANP model, M.Sc in Physical Geography (Hydro Geomorphology), University of Mohaghegh Ardabili, 2014, 617 p.

27. Islam ARMT, Talukdar S, Mahato S, Kundu S, Eibek KU, Pham QB, et al. Flood susceptibility modelling using advanced ensemble machine learning models. *Geoscience Frontiers*. 2021;12(3):101075.
28. Arab N, Mahini S, A M, Tabrizi A, Witte T. Flood risk analysis using random forest machine learning method (case study: Mashhad city). *Ecohydrology*. 2023;10(1):1-15.
- potential of watersheds using hierarchical analysis process (case study: Ilam city watershed). *Comprehensive watershed management*. 2021;1(1):1-16.
26. Kayani Asl MA, Motshefa B, Roshan SH. Evaluation of machine learning algorithms (RF and SVM) in producing flood sensitivity map of Maron watershed. *Iranian Journal of Watershed Science and Engineering*. 2023;17(61):41-51.