



Flood susceptibility mapping in the semi-arid region of Tabas using machine learning algorithms

Mahnaz Naemitabar¹ | Mohammadali Zangeneh Asadi² | Mahdi Boroughani³

1. Research Center of Social Science & Geographical Studies, Hakim Sabzevari University, Sabzevar, Iran. mahnaznaemi70@gmail.com

2. Department of Climatology and Geomorphology, Faculty of Geography and Environmental Science, Hakim Sabzevari University, Sabzevar, Iran. Ma.zanganehasadi@hsu.ac.ir

3. Corresponding Author, Research Center of Social Science & Geographical Studies, Hakim Sabzevari University, Sabzevar, Iran. m.boroughani@hsu.ac.ir

Article Information

Research Paper

Vol:	16
No:	61
P:	46-67
Received:	2025-03-01
Revised:	2025-04-14
Accepted:	2025-04-14
Published:	2025-11-01

Keywords:

- Flood
- Machine Learning
- Tabas City
- IGR Index
- Semi-arid Regions

Cite this Article:

Naemitabar, Mahnaz., Zangeneh Asadi, Mohammadali, & Boroughani, Mahdi. (2025). Flood susceptibility mapping in the semi-arid region of Tabas using machine learning algorithms. *Journal of Arid Regions Geographic Studies* 16(61): 46-67.

doi: 10.22034/jargs.2025.509539.1182

Publisher: Hakim Sabzevari University

Authors retain the copyright and full publishing rights.



Abstract

Aim: This study aims to identify flood-prone areas in the semi-arid region of Tabas using LT, K-SVM, and KNN machine learning algorithms and to examine the most critical factors affecting flooding in the region.

Materials & Methods: First, 288 flood points were collected from the Natural Resources and Watershed Management Organization of South Khorasan Province. In the second step, the data were divided into two parts: 70% for training and model development, and 30% for validation. Then, 12 key factors in assessing flood potential were examined, and the accuracy of the algorithms was evaluated using various statistical indicators.

Findings: The flood zoning maps indicate that, according to the LT algorithm, the northern and western regions face a high risk of flooding, estimated at 43%. The western and southern areas are at very high risk (36%). In contrast, the K-SVM algorithm indicates that the central, western, and southern parts of the region have a high risk of flooding (59%), whereas the eastern, northern, and central areas face a very high risk of flooding (41%). According to the KNN algorithm, the northern, central, and eastern regions show a high flood risk, estimated at 38%. The northwestern and western regions are classified as very high-risk zones, accounting for 27% of the total area.

Conclusion: Flood zoning maps indicate that high-risk areas are located near waterways, where subsurface water flow generates moisture fronts. Hydrological factors, such as flood-related indices, indicate that the western, central, northern, northwestern, and southeastern regions are the main flood-prone areas.

Innovation: The present study uses machine learning algorithms, which bring significant innovations. This approach not only increases the accuracy of predictions but also enables the use of multi-purpose data and the identification of complex patterns. The outcomes of this method will play a significant role in flood management and in preserving the health of aquatic ecosystems.

Extended Abstract

1. Introduction

In recent decades, flooding has become one of the most widespread natural disasters globally. In Iran, due to specific climatic and geographical conditions, ignoring the flood phenomenon can have irreparable consequences. Accurate estimation, assessment, and zoning of flood-prone areas are crucial for minimizing potential risks and damages. Arid regions, characterized by heavy rainfall, poor watershed management, limited vegetation cover, and impermeable soils, experience destructive and hazardous floods. Surface runoff mechanisms in these regions, due to reduced water penetration in the soil, lead to rapid accumulation of water on the ground surface. This causes sudden and severe floods that can have devastating effects on the environment and communities. These unique conditions in arid regions present significant management and environmental challenges that demand a comprehensive understanding and effective solutions for flood management. However, it is possible to reduce the damage and loss caused by floods by predicting the areas at risk. These methods involve locating potential flood areas and evaluating their impacts to mitigate destructive effects. New techniques in the field of GIS (Geographic Information Systems) and remote sensing have led to significant developments in flood prediction modeling.

2. Materials and Methods

In this study, 288 flood-prone locations were first collected from the Natural Resources and Watershed Management Organization of South Khorasan Province. Then, these locations were divided into two groups: 70% of the data for training and model implementation, and 30% for validation. Flood-prone areas, as well as locations that had no flood history, were randomly selected and divided into two groups. In the next step, the factors affecting the occurrence of flooding were examined. In the present study, several variables involved in the flood phenomenon were considered independent variables, and their relationship with flood zones was analyzed. Then, 12 key factors in flood potential assessment were examined. The IGR index was also used to evaluate the predictability of quantitatively effective factors. Then, the multilinear factor test was used to assess the linearity of the variables. Also, five different statistical methods have been used to validate and evaluate the algorithms, including sensitivity, specificity, positive and negative indices, and ROC-AUC analysis.

3. Results and Discussion

The results obtained from the IGR index show that land use, precipitation, slope, lithology, distance from the river, and the SPI index are the most critical factors influencing flooding in the study area. In the multiple collinearity test, the results show that all flood factors have VIF values less than 10 and a Tolerance greater than 0.2. This indicates the lack of interference and overlap of factors in this field. The performance evaluation results of the algorithms indicate that the LT algorithm achieves the highest SST, SPF, PPV, and NPV values in both datasets. The K-SVM algorithm is ranked second, with SST and PPV values of 0.77 and 0.71, respectively. The KNN algorithm demonstrates strong performance, achieving the highest SST (0.68) and PPV (0.65) values among the algorithms studied. These results indicate that all three algorithms show good accuracy and performance in analyzing the dataset. The results of the flood zoning maps suggest that areas with high and very high flood risk are located on slopes exceeding 20 degrees, which has a significant impact on the occurrence of flooding in the region. In addition to the slope of the main river, the slope of the land is also a key factor in the intensification of flooding in the region. In the study area, rivers and their tributaries follow their paths along the slopes. Flood-prone areas are located in low-lying areas. As a result, the rate of erosion and transport of alluvial and sedimentary materials in these areas increases sharply. Areas near the river network are at high risk due to persistent high water flow and spring floods. Studies show that parts of the region with geologically and lithologically resistant surface formations, low permeability, and sparse vegetation cover are prone to flooding.

4. Conclusions

According to the results of flood zoning maps, areas with high flood risk are located near the confluence of watercourses. This phenomenon is caused by the flow of subsurface water from rivers to adjacent slopes, resulting in the formation of moisture in the soil and weakening its resistance. This trend, together with the high density of drainage, exacerbates the risk of flooding in these areas. The western, central, northern, and southeastern regions are known to be prone to flooding due to topographic features such as altitude, steep slopes, high rainfall, and dense watercourse networks. Human activities, such as excessive grazing of livestock and the destruction of vegetation, have reduced soil permeability and increased runoff and flood intensity in these areas. In contrast, low-risk flood areas usually have dense vegetation, are far from significant watercourses, and have a less dense drainage network. These areas have different conditions in terms of soil, altitude, and slope, and, for this reason, are less exposed to flooding. The findings of this study contribute to the preservation of natural resources, soil, and vegetation, as well as to effective crisis management.

5. Acknowledgment & Funding

- The authors are thankful to all interview participants for supporting this research.
- The manuscript did not receive a grant from any organization.

6. Conflict of Interest

- The authors declare no conflict of interest.



پهنه‌بندی سیلاب در منطقه نیمه خشک طبس با استفاده از الگوریتم‌های یادگیری ماشین

مهناز ناعمی تبار^۱، محمدعلی زنگنه اسدی^۲، مهدی بروغنی^۳ 

۱- مرکز پژوهشی مطالعات جغرافیایی و علوم اجتماعی، دانشگاه حکیم سبزواری، سبزوار، ایران. mahnaznaemi70@gmail.com
۲- آب و هواشناسی و ژئومورفولوژی، دانشکده جغرافیا و علوم محیطی، دانشگاه حکیم سبزواری، سبزوار، ایران. Ma.zanganehasadi@hsu.ac.ir
۳- نویسنده مسئول، مرکز پژوهشی مطالعات جغرافیایی و علوم اجتماعی، دانشگاه حکیم سبزواری، سبزوار، ایران. m.boroughani@hsu.ac.ir

چکیده

هدف: هدف از پژوهش حاضر شناسایی مناطق مستعد سیل‌خیز در منطقه نیمه‌خشک طبس با استفاده از الگوریتم‌های یادگیری ماشین لجستیک درختی (LT)، (K-SVM) خطی و نزدیک‌ترین همسایه (KNN) است تا مهم‌ترین عوامل مؤثر در سیل‌خیزی منطقه شناسایی و مورد بررسی قرار گیرد.

روش و داده: ابتدا ۲۸۸ نقطه سیلابی از منابع سازمان منابع طبیعی و آبخیزداری استان خراسان جنوبی گردآوری شده است. در گام دوم داده‌ها به دو بخش تقسیم شدند: ۷۰ درصد جهت آموزش و توسعه مدل و ۳۰ درصد جهت اعتبارسنجی. سپس ۱۲ عامل کلیدی در ارزیابی پتانسیل سیلاب مورد بررسی قرار گرفت و دقت الگوریتم‌ها با استفاده از شاخص‌های آماری مختلف ارزیابی شد.

یافته‌ها: نقشه‌های پهنه‌بندی سیل نشان می‌دهند که در الگوریتم LT، مناطق شمالی و غربی با خطر سیل خیزی زیاد (۴۳٪) و نواحی غربی و جنوبی با خطر خیلی زیاد (۳۶٪) مواجه هستند. در مقابل، الگوریتم K-SVM نشان می‌دهد که قسمت‌های مرکزی، غربی و جنوبی منطقه در معرض سیل خیزی زیاد (۵۹٪) قرار دارند، در حالی که نواحی شرقی، شمالی و مرکزی خطر سیل خیزی خیلی زیاد (۴۱٪) را نشان می‌دهند. در الگوریتم KNN نیز نواحی شمالی، شرقی و مرکزی با خطر سیل خیزی زیاد (۳۸٪) و مناطق شمال غربی، غربی و مرکزی با خطر سیل خیزی خیلی زیاد (۲۷٪) قرار گرفته است که ناشی از تغییرات کاربری اراضی، به ویژه در مناطقی است که فاقد پوشش گیاهی هستند که باعث افزایش رواناب می‌شود و خطر سیلاب را تشدید می‌کند.

نتیجه‌گیری: نقشه‌های پهنه‌بندی سیلاب نشان می‌دهند که مناطق پرخطر در نزدیکی آبراهه‌ها قرار دارند جایی که جریان آب‌های زیرسطحی باعث تشکیل جبهه‌های رطوبتی می‌شود. این امر مقاومت خاک را کاهش داده و احتمال وقوع سیل خیزی را افزایش می‌دهد. عوامل هیدرولوژیکی مانند شاخص‌های مرتبط با سیل‌خیزی، مناطق غربی، مرکزی، شمال و شمال غرب و جنوب شرق را به عنوان مناطق سیل‌خیز اصلی معرفی می‌کنند.

نوآوری، کاربرد نتایج: استفاده از الگوریتم‌های یادگیری ماشین در پژوهش حاضر نوآوری‌های قابل توجهی را به همراه دارد. این رویکرد نه تنها دقت پیش‌بینی‌ها را افزایش می‌دهد، بلکه امکان استفاده از داده‌های چندمنظوره و شناسایی الگوهای پیچیده را فراهم می‌کند. نتایج حاصل از این رویکرد، در مدیریت سیلاب و حفظ سلامت اکوسیستم‌های آبی تأثیرگذار خواهد بود.

اطلاعات مقاله

مقاله پژوهشی

شماره: ۱۶

دوره: ۶۱

صفحه: ۴۶-۶۷

تاریخ دریافت: ۱۴۰۳/۱۲/۱۱

تاریخ ویرایش: ۱۴۰۴/۰۱/۲۵

تاریخ پذیرش: ۱۴۰۴/۰۱/۲۵

تاریخ انتشار: ۱۴۰۴/۰۸/۱۰

کلیدواژه‌ها:

- سیلاب
- یادگیری ماشین
- شهر طبس
- شاخص IGR
- مناطق نیمه خشک

نحوه ارجاع به این مقاله:

ناعمی تبار، مهناز، زنگنه اسدی، محمدعلی و بروغنی، مهدی. (۱۴۰۴). پهنه‌بندی سیلاب در منطقه نیمه‌خشک طبس با استفاده از الگوریتم‌های یادگیری ماشین. *مطالعات جغرافیایی مناطق خشک*، ۱۶(۶۱): ۴۶-۶۷

doi: 10.22034/jargs.2025.509539.1182

ناشر: دانشگاه حکیم سبزواری



نویسندگان حق نشر و حقوق کامل انتشار را حفظ می‌کنند.

۱- مقدمه

سیل به‌عنوان یکی از رایج‌ترین و گسترده‌ترین بلایای طبیعی در جهان در چند دهه اخیر شناخته شده است (Kumar et al., 2025). سیل پدیده‌ای است که در آن جریان آب از ظرفیت طبیعی رودخانه فراتر می‌رود و از کناره‌های آن سرریز می‌شود. این امر منجر به غرقاب شدن اراضی کم ارتفاع اطراف رودخانه می‌گردد. پدیده سیلاب می‌تواند منجر به خسارات جدی در مناطق کم ارتفاع مجاور رودخانه شود (Chadee et al., 2014; Mirchooli et al., 2023). در کشور ایران، با توجه به شرایط اقلیمی و جغرافیایی خاص، عدم توجه به پدیده سیلاب می‌تواند پیامدهای جبران‌ناپذیری به همراه داشته باشد. برآورد و ارزیابی دقیق سیلاب و همچنین پهنه‌بندی مناطق مستعد سیل، از اهمیت بالایی برخوردار است تا بتوان خطرات و خسارات احتمالی را به حداقل رساند (Zangeneh Asadi et al., 2021). طی سال‌های اخیر، سیلاب به دلیل توسعه شهری سریع، همراه با زیرساخت‌های نامناسب و برنامه‌ریزی محیطی ناکارآمد، خطر سیلاب را تشدید کرده است. تغییرات در کاربری زمین، از جمله گسترش شهرها و تخریب جنگل‌ها، در کنار تغییرات اقلیمی، منجر به افزایش رواناب و در نتیجه، وقوع سیلاب‌های مکرر می‌شود (Chidiebere et al., 2024). مناطق خشک و نیمه‌خشک به طور معمول با کمبود آب و خشکسالی طولانی‌مدت مشخص می‌شوند و اغلب تصور می‌شود که به دلیل بارندگی کم سالانه با مشکلات سیلابی ناچیز روبه‌رو هستند. با این حال، حوادث سیل در دهه‌های اخیر نشان می‌دهد که این مناطق به شدت در معرض سیل قرار دارند (Nabinejad & Schüttrumpf, 2023). در مناطق خشک و نیمه‌خشک که بخش قابل‌توجهی از سرزمین ایران را در بر می‌گیرد، خشکسالی‌های مکرر و سیلاب‌های ویرانگر از چالش‌های عمده محسوب می‌شوند (Soltani, 2002). مناطق خشک، با ویژگی‌هایی نظیر بارش‌های سنگین، مدیریت نامناسب حوضه آبریز، پوشش گیاهی ناچیز و خاک‌های غیرقابل نفوذ، شاهد سیلاب‌های مخرب و خطرناکی هستند. مکانیسم‌های رواناب سطحی در این مناطق به دلیل کاهش نفوذ آب در خاک، منجر به تجمع سریع آب در سطح زمین می‌شود. این امر موجب تشکیل سیلاب‌های ناگهانی و شدید می‌گردد که می‌تواند اثرات مخربی بر محیط زیست و جوامع داشته باشد. این شرایط خاص در مناطق خشک، چالش‌های مدیریتی و زیست‌محیطی منحصره‌فردی را ایجاد می‌کند که نیازمند درک عمیق و راهکارهای مناسب برای مدیریت سیلاب‌ها است (Rezaei-Sadr & Eslamian, 2022). تغییرات آب و هوایی، افزایش جمعیت، توسعه شهرنشینی و برنامه‌ریزی نامناسب شهری، عواملی هستند که موجب بروز سیلاب‌های ویرانگر و خسارات جبران‌ناپذیر در طول سال‌ها شده‌اند. تکرار این رخدادهای لزوم تهیه نقشه‌های جامع و دقیق از مناطق در معرض خطر سیل را ضروری می‌سازد تا بدین وسیله آمادگی لازم برای مقابله با بلایای طبیعی افزایش یافته و استراتژی‌های مؤثرتری برای کاهش خسارات احتمالی به‌کار گرفته شود (Adedoyin Benson & Akinola Adesuji, 2025).

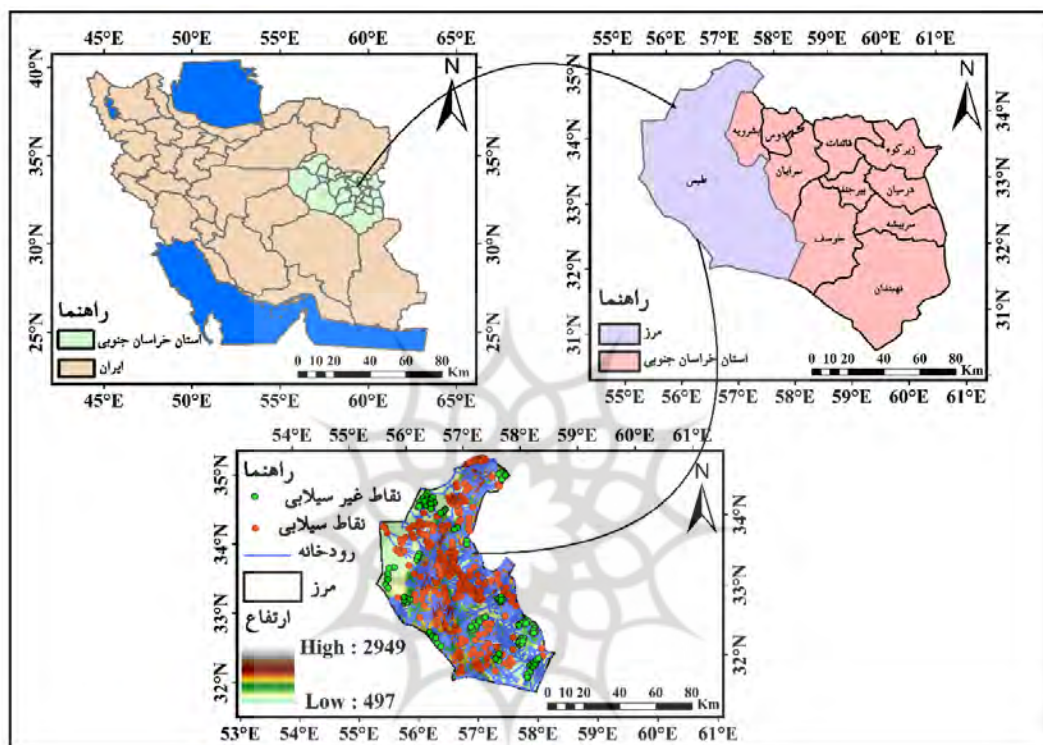
با این حال، می‌توان با پیش‌بینی مناطق در معرض خطر، میزان خسارات و تلفات ناشی از آن را کاهش داد. این روش‌ها شامل شناسایی مکان‌های احتمالی وقوع سیل و ارزیابی عواقب آن است تا بتوان از اثرات مخرب آن کاست (Ali et al., 2020). تکنیک‌های نوین در حوزه GIS (سیستم اطلاعات جغرافیایی) و سنجش از دور منجر به تحولات چشمگیری در مدل‌سازی پیش‌بینی سیل شده است (Tien Bui et al., 2019). برای رفع چالش‌های موجود در این زمینه، مدل‌های یادگیری ماشین به دلیل مزایای قابل توجهی همچون دقت بالا، عملکرد قوی و قابلیت پیش‌بینی، به‌طور گسترده‌ای در توسعه و بهبود مدل‌های پیش‌بینی سیل مورد استفاده قرار می‌گیرند (Ahmadlou et al., 2019). پیشرفت و محبوبیت الگوریتم‌های یادگیری ماشین به دلیل توانایی منحصره‌فرد این الگوریتم‌ها در پیش‌بینی غیرخطی بودن سیل است. این الگوریتم‌ها می‌توانند با استفاده از داده‌های تاریخی سیل، الگوها و رفتار سیل را شناسایی کنند، بدون این‌که نیاز به درک عمیق معادلات ریاضی پیچیده‌ای داشته باشند که فرایندهای فیزیکی و رفتار حوضه‌های آبریز را توصیف می‌کنند. این رویکرد، پیش‌بینی وقایع هیدرولوژیکی را ساده‌تر و در دسترس‌تر می‌کند و به دانشمندان و متخصصان این حوزه کمک می‌کند تا بدون نیاز به دانش تخصصی عمیق در ریاضیات، به درک بهتری از پدیده‌های مرتبط با سیل دست یابند (Mosavi et al., 2018). مدیریت سیلاب در مقیاس منطقه‌ای، نیازمند توجه فوری به مدل‌سازی سیلاب در حوضه‌های آبریز است. این رویکرد فضایی می‌تواند در درک بهتر پدیده سیلاب و اتخاذ راهکارهای مؤثر برای سازگاری و کاهش خسارات ناشی از آن نقش کلیدی ایفا کند؛ لذا ضروری است که مطالعات جامع در این زمینه انجام شود تا بتوان تدابیر لازم را برای مقابله با اثرات مخرب سیلاب‌ها اتخاذ نمود (Cao et al., 2019; Su et al., 2019; Costache & Bui, 2020; Pourhashemi & Aliabadi., 2025). در زمینه پهنه‌بندی خطر سیلاب تاکنون مطالعات زیادی در نقاط مختلف جهان و ایران صورت گرفته است. رحیم پور و رضائی مقدم

در پژوهشی جهت مدل سازی پتانسیل خطر وقوع سیلاب در حوضه آبریز آجی چای با استفاده از الگوریتم های داده کاوی به این نتایج دست یافتند که مدل جنگل تصادفی با سطح زیر منحنی ۰/۹۴ نسبت به سایر الگوریتم ها برتری دارد. جالب توجه است که الگوی توزیع فضایی مناطق در معرض خطر در همه مدل ها یکسان است که نشان دهنده یک الگوی ثابت در این منطقه است (Rahimpour & Rezaei Moghaddam, 2024). ابراهیمی و ایلانلو به پهنه بندی وقوع سیلاب در حوضه زهکشی شهرستانک با استفاده از مدل WMS و GIS پرداختند. این مطالعه نشان داد که ادغام مدل WMS با سیستم اطلاعات جغرافیایی رویکرد مؤثری برای پهنه بندی سیلاب و شناسایی مناطق پرخطر است. نتایج حاکی از آن است که این روش ترکیبی می تواند در ارزیابی و مدیریت مخاطرات سیلاب در مناطق مشابه مورد استفاده قرار گیرد (Ebrahimi & Ilanlou, 2024). قنواتی و همکاران جهت بررسی حساسیت سیل در استان خوزستان از الگوریتم های مقایسه ای یادگیری ماشین استفاده نمودند. نتایج پژوهش نشان داد که مساحت زیر منحنی برای مدل های GLM و FDA بالاتر از ۰/۷ به دست آمده است؛ اما مدل RF با کسب مساحت زیر منحنی ۸/۹۸ درصد عملکرد مناسب تری نسبت به سایر مدل ها دارد (Qanavati et al., 2024). هیتوری و همکاران در مطالعه ای، از تکنیک های پیشرفته یادگیری ماشین برای توسعه نقشه های خطر سیل در حوضه آبریز متلیلی در شمال شرق مراکش بهره بردند. ارزیابی مدل ها حاکی از برتری مدل RF در پیش بینی دقیق مناطق در معرض خطر سیل است. این یافته ها نشان دهنده پتانسیل یادگیری ماشین در بهبود مدیریت منابع آب و کاهش آسیب پذیری در برابر سیلاب ها است (Hitouri et al., 2024). وگانس و همکاران در پژوهشی نوآورانه، از مدل جنگل تصادفی به منظور پیش بینی خطر سیلاب در چهار ایالت غربی ایالات متحده، شامل آریزونا، نیومکزیکو، کلرادو و یوتا بهره بردند. نتایج این مطالعه نشان داد که علی رغم چالش هایی مانند دقت داده های موجود و پیچیدگی درک تعاملات متغیرها، این رویکرد می تواند به عنوان یک چارچوب قدرتمند برای ارزیابی خطر سیلاب در مناطق مختلف مورد استفاده قرار گیرد و این روش شناسی چارچوبی قوی برای ارزیابی خطر سیلاب در مناطق دیگر ارائه می دهد (Venegas et al., 2024). در پژوهشی غلامی و همکاران به بررسی خطر سیلاب در منطقه جنوبی ایران پرداختند. آن ها با به کارگیری مدل پیشرفته bLSTM و تکنیک COPRAS در زمینه یادگیری عمیق و تصمیم گیری چندمعیاره (MCDM)، رویکردی جامع را در منطقه هرمزگان ارائه کردند. نتایج این مطالعه نشان داد که این روش شناسی می تواند به عنوان ابزاری قدرتمند در ارزیابی و نقشه برداری از ریسک سیلاب و سایر بلایای طبیعی در مناطق مختلف جهان مورد استفاده قرار گیرد (Gholami et al., 2025). به طور کلی نتایج پژوهش ها در زمینه مدل سازی سیلاب با الگوریتم های مختلف، بر پتانسیل الگوریتم ها جهت افزایش پیش بینی دقت سیل، ارزیابی خطر و ارزیابی خسارات ناشی از سیلاب تأکید می کنند. پژوهش های انجام شده به طور کلی تطبیق پذیری الگوریتم های ML را در پرداختن به جنبه های مختلف مدیریت و پهنه بندی سیلاب نشان می دهد. در این پژوهش، هدف شناسایی مناطق مستعد سیل خیز در منطقه طبس در استان خراسان جنوبی با به کارگیری الگوریتم های یادگیری ماشین است. این منطقه از نظر جغرافیایی، دارای ناهمواری های توپوگرافی و شیب های تند است که در هنگام بارندگی های شدید، خطر سیلاب های ناگهانی را افزایش می دهد. اهمیت بررسی سیلاب در این منطقه، به دلیل نیاز مبرم به مدیریت ریسک سیل، تضمین ایمنی ساکنین و برنامه ریزی های پایدار در مواجهه با تغییرات اقلیمی است. پژوهش حاضر با پر کردن خلاءهای تحقیقاتی موجود، به بررسی سیل خیزی در منطقه می پردازد. سیل، پدیده ای طبیعی است که به دلیل رواناب شدید و ناگهانی در مناطق خشک رخ می دهد. این رخداد می تواند ناشی از عوامل متعددی از جمله بارش های سنگین، ذوب سریع برف در بهار و مدیریت نامناسب کاربری زمین باشد. سیل خیزی، به شرایطی اشاره دارد که احتمال وقوع سیل را در یک منطقه افزایش می دهد و این عوامل شامل ویژگی های ژئومورفولوژیکی، نوع خاک، پوشش گیاهی و تغییرات آب و هوایی است. هدف از مطالعه سیل خیزی در منطقه، شناسایی مناطق در معرض خطر و اتخاذ تدابیر پیشگیرانه جهت کاهش خسارات احتمالی است. این پژوهش تلاش می کند تا با بررسی دقیق عوامل، به پیش بینی و مدیریت بهتر سیل خیزی در منطقه کمک نماید. استفاده از الگوریتم های یادگیری ماشین در پژوهش حاضر نوآوری های قابل توجهی را به همراه دارد. این رویکرد نه تنها دقت پیش بینی ها را افزایش می دهد، بلکه امکان استفاده از داده های چندمنظوره و شناسایی الگوهای پیچیده را فراهم می کند. با توجه به تغییرات اقلیمی و نیاز به مدیریت بهینه منابع آب در ایران، این پژوهش می تواند به عنوان یک مدل برای سایر نواحی با ویژگی های مشابه مورد استفاده قرار گیرد. این مطالعه می تواند تأثیر قابل توجهی بر مدیریت بلایای طبیعی و توسعه پایدار منطقه داشته باشد. نتایج حاصل، راهکارهایی را برای کاهش خسارات سیلاب ارائه می دهد و همچنین در راستای بهبود سیاست های زیست محیطی و مدیریت منابع آب، گام های مؤثری را ترسیم می نماید. این پژوهش، با تمرکز بر جنبه های مختلف مدیریت منابع طبیعی، به عنوان ابزاری ارزشمند برای حمایت از محیط زیست و تضمین توسعه پایدار عمل می کند.

۲- مواد و روش

۲-۱- منطقه مورد مطالعه

طبس در حوضه آبخیز کویر لوت در استان خراسان جنوبی ایران واقع شده است (Kaffash et al., 2024). منطقه مورد مطالعه در این پژوهش، مساحتی حدود ۱۵۷۰۰۰ کیلومتر مربع را در بر می‌گیرد (شکل ۱). الگوی بارش در این منطقه به طور عمده در فصول بهار و تابستان مشاهده می‌شود و طبق نمودار آمبروترمیک، حداقل نیمی از سال را در شرایط خشکی سپری می‌کند. طبق گزارش‌های شرکت آب منطقه‌ای، این استان فاقد رودخانه‌های دائمی و بزرگ است و تأمین آب بخش‌های کشاورزی، شرب و صنعت به طور عمده از منابع آب‌های زیرزمینی صورت می‌گیرد (Kavian et al., 2021).



شکل ۱. موقعیت منطقه مورد مطالعه

۲-۲- روش پژوهش

در این مطالعه، ابتدا ۲۸۸ موقعیت سیل‌خیز از سازمان منابع طبیعی و آبخیزداری استان خراسان جنوبی جمع‌آوری شده است. سپس این نقاط به دو گروه تقسیم شدند: ۷۰ درصد داده‌ها برای آموزش و اجرای مدل و ۳۰ درصد جهت اعتبارسنجی. مناطق مستعد سیلاب و همچنین مکان‌هایی که سابقه سیلاب نداشتند، به صورت تصادفی انتخاب و به دو دسته تقسیم‌بندی شدند. در فرایند آموزش مدل، تقسیم داده‌ها به نسبت ۷۰-۳۰ درصد، تعادل مناسبی بین یادگیری و ارزیابی ایجاد می‌کند. این روش تضمین می‌کند که مدل نه تنها الگوها را از داده‌های آموزشی فرا می‌گیرد، بلکه از یادگیری بیش از حد جزئیات نیز جلوگیری می‌کند. با اختصاص ۷۰ درصد داده‌ها به آموزش، مدل می‌تواند روابط و الگوها را شناسایی کند، در حالی که ۳۰ درصد باقی‌مانده به ارزیابی عملکرد مدل اختصاص می‌یابد. این تقسیم‌بندی هوشمندانه، خطر برازش بیش از حد^۱ را کاهش می‌دهد و اطمینان حاصل می‌شود که مدل به جای تمرکز بر جزئیات، بر درک کلی الگوها متمرکز است.

علاوه بر این، انتخاب تصادفی داده‌ها از مناطق سیل‌خیز و غیر سیل‌خیز، دقت مدل‌سازی را افزایش می‌دهد. تنوع در داده‌ها به الگوریتم‌ها کمک می‌کند تا الگوها و ویژگی‌های مختلف را شناسایی کند و قابلیت تعمیم آن را تقویت نماید. در نتیجه، الگوریتم‌ها می‌توانند پیش‌بینی‌های دقیق‌تری ارائه دهند و عملکرد بهتری در مواجهه با داده‌های جدید داشته باشند.

1. Overfitting

در مرحله بعد، عوامل تأثیرگذار در وقوع سیل خیزی مورد بررسی قرار گرفتند. در پژوهش حاضر، متغیرهای متعددی را که در پدیده سیلاب دخیل هستند، به عنوان متغیرهای مستقل در نظر گرفته شدند و ارتباط آن‌ها با پهنه‌های سیلابی مورد تحلیل قرار گرفت. سپس، ۱۲ عامل کلیدی در پتانسیل سنجی سیل خیزی مورد بررسی قرار گرفتند. این پارامترها شامل کاربری اراضی، طبقات ارتفاعی، لیتولوژی، شیب و جهت شیب، فاصله از آبراهه، بارش، خاک، انحنا سطح، تراکم زهکشی، شاخص رطوبت توپوگرافی^۱ (TWI) و شاخص توان آبراهه^۲ (SPI) هستند. داده‌های مورد نیاز از نقشه‌های توپوگرافی ۱:۲۵۰۰۰ و تصاویر ماهواره‌ای لندست (ASTER، ETM⁺، OLI، TM) سال ۲۰۲۴ و جهت تهیه لایه ارتفاع از لایه رقومی ارتفاعی (DEM)^۳ با قدرت تفکیک مکانی ۱۲/۵ متر در مقیاس ۱:۱۰۰۰ اداره محیط زیست و منابع طبیعی استفاده شده است. این لایه به عنوان پایه‌ای برای استخراج لایه‌های شیب و جهت شیب استفاده شد. در گام بعد، لایه تراکم زهکشی نیز از روی لایه آبراهه‌ها محاسبه گردید. جهت ایجاد لایه بارش نیز از داده‌های بارش ۳۰ ساله (۱۳۷۰-۱۴۰۰) ایستگاه‌های هواشناسی (بشرویه، نیاز طیس، نینندان، نیستان، دیهوک، طیس، پلنگ آباد، خروان طیس، رقع، دستگرد عشق آباد و معدن پرورده) با روش درون‌یابی کریجینگ استفاده شد. سپس اطلاعات لیتولوژی از نقشه زمین‌شناسی ۱:۱۰۰۰۰۰ منطقه مطالعاتی استخراج گردید. لایه رواناب نیز با استفاده از داده‌های بارش ۳۰ ساله منطقه استخراج گردید. سپس داده‌های بارندگی از ایستگاه‌های هواشناسی به عنوان منبع اصلی برای استخراج رواناب مورد استفاده قرار گرفتند. داده‌ها به صورت سری‌های زمانی روزانه و ساعتی جمع‌آوری و سازمان‌دهی شدند. سپس، با استفاده از مدل HEC-HMS، مدل‌سازی رواناب انجام شد. این روش، ابزاری قدرتمند در مدیریت منابع آب و پیش‌بینی سیلاب‌ها محسوب می‌شود. با تنظیم پارامترهای مدل و داده‌های بارندگی، شبیه‌سازی رواناب اجرا گردید. جهت تهیه نقشه فاصله از آبراهه از لایه آبراهه (رودخانه) استفاده شده است. برای تعیین فاصله، از ابزار Buffer در سیستم اطلاعات جغرافیایی GIS استفاده گردید تا فاصله هر نقطه در منطقه به دست آید. با استفاده از ابزار Distance to Nearest Hub نیز فاصله از رودخانه محاسبه گردید و فاصله‌های به دست آمده در قالب یک لایه رستری ایجاد شد که در آن هر پیکسل نمایانگر فاصله از آبراهه است. سپس، با به‌کارگیری ابزار بازطبقه‌بندی^۴ در GIS، فاصله‌های محاسبه شده به پنج طبقه مختلف تقسیم‌بندی شدند. لایه خاک با ترکیب اطلاعات واحدهای اراضی و نقشه خاک‌شناسی تهیه شد و لایه کاربری اراضی از تصاویر ماهواره‌ای لندست ۲۰۲۴ و نرم‌افزار ENVI استخراج گردید. جهت برنامه‌نویسی و اجرای الگوریتم‌ها، از نرم‌افزار R Studio 2021.09.2 جهت تهیه نقشه‌های مؤثر بر سیل خیزی و پهنه‌بندی از نرم‌افزار ArcGIS 10.4 و همچنین جهت تحلیل آماری داده‌ها از نرم‌افزار SPSS 2016 استفاده شده است. در جدول ۱ لایه‌های استفاده شده در پژوهش ارائه شده است.

جدول ۱. مجموعه داده‌های مورد استفاده در پژوهش به همراه منابع

داده‌ها	فرمت	لایه‌ها
تصاویر ماهواره‌ای OLI، TM، ASTER، ETM ⁺ لایه رقومی ارتفاعی DEM	رستر (Raster)	ارتفاع
		شیب
		جهت شیب
		انحنا سطح
		TWI
		SPI
تصاویر ماهواره‌ای لندست سال ۲۰۲۴	وکتور (Vector)	کاربری اراضی
داده‌های بارندگی سازمان هواشناسی استان خراسان جنوبی	داده‌های عددی (Numerical data)	بارش
نقشه توپوگرافی ۱:۲۵۰۰۰	وکتور (Vector)	تراکم زهکشی و فاصله از رودخانه
زمین‌شناسی	وکتور (Vector)	نقشه زمین‌شناسی ۱:۱۰۰۰۰۰
خاک	وکتور (Vector)	نقشه خاک منابع طبیعی و آبخیزداری استان خراسان رضوی

1. Topographic wetness index (TWI)
2. Stream power index (SPI)
3. Digital Elevation Model (DEM)
4. Reclassify

شاخص رطوبت توپوگرافی و شاخص توان آبراهه از جدول ۲ محاسبه می‌شود.

جدول ۲. شاخص‌های مورد استفاده در پژوهش

منبع	روابط	شاخص
Kirkby & Beven, 1979	$TWI = \ln \left(\frac{AS}{\tan \beta} \right)$	شاخص رطوبت توپوگرافی (TWI)
Moore et al., 1991	$SPI = \left(\frac{AS}{\tan \beta} \right)$	شاخص توان آبراهه (SPI)

AS مساحت حوضه آبخیز و β شیب به (درجه) است.

AS مساحت حوضه آبخیز و $\tan \beta$ گرادیان شیب برحسب درجه است.

• الگوریتم لجستیک درختی^۱

مدل درخت لجستیک، روشی نوین در طبقه‌بندی داده‌ها است که از تلفیق رگرسیون لجستیک و الگوریتم‌های یادگیری درخت تصمیم بهره می‌برد. این مدل، با ترکیب دو تکنیک قدرتمند، توانمندی‌های منحصر به فردی در طبقه‌بندی و پیش‌بینی ارائه می‌دهد. در این رویکرد، رگرسیون لجستیک به عنوان پایه‌ای برای ساخت درخت تصمیم مورد استفاده قرار می‌گیرد که منجر به ایجاد یک مدل طبقه‌بندی قوی و انعطاف‌پذیر می‌شود (Landwehr et al., 2005; Tien Bui et al., 2015). مدل‌های LT با ویژگی‌های برجسته‌ای همراه هستند که شامل سرعت ساخت و سادگی تفسیر آن‌ها می‌شود. در این مدل‌ها، الگوریتم LogitBoost در نسخه لجستیک به منظور توسعه یک مدل LR در هر گره از درخت به کار گرفته می‌شود. الگوریتم CART نیز برای مدیریت و هرس درخت مورد استفاده قرار می‌گیرد. اطلاعات حاصل از این فرایند برای تقسیم‌بندی و طبقه‌بندی داده‌ها مورد استفاده قرار می‌گیرد (Breiman et al., 2017).

• الگوریتم ماشین بردار پشتیبان خطی^۲

ماشین بردار پشتیبان (SVM)^۲ به عنوان یک روش قدرتمند در یادگیری ماشین شناخته می‌شود و مدل k-SVM بر اساس این مفهوم توسعه یافته است. این مدل، یک رویکرد پیشرفته در حل مسائل طبقه‌بندی و رگرسیون است که از اصول SVM بهره می‌برد. k-SVM با استفاده از تکنیک‌های رگرسیون و طبقه‌بندی، توانایی حل طیف وسیعی از مشکلات را در حوزه‌های مختلف دارد. این مدل، گسترش یافته مدل SVM است و با ارائه یک چارچوب جامع، به بهبود عملکرد در مسائل پیچیده کمک می‌کند (Usman et al., 2021; Chuma et al., 2021). روش رگرسیون، رویکردی قدرتمند برای ساده‌سازی فرایندهای پیچیده است. این تکنیک با تجزیه و تحلیل روابط متقابل بین عوامل پیش‌بینی‌کننده و نتایج، پیچیدگی‌های موجود در یادگیری را آشکار می‌سازد. از این طریق، رگرسیون قادر است با درک این تعاملات، پیچیدگی‌های یک فرایند را کاهش داده و آن را به صورت ساده‌تری ارائه دهد (Agrawal et al., 2020; Yaseen et al., 2018).

• الگوریتم نزدیکترین همسایه^۴ (KNN)

الگوریتم نزدیکترین همسایه، الگوریتمی مبتنی بر یادگیری نظارت شده، در زمینه محاسبات نرم غیر پارامتریک، کاربردهای متنوعی دارد. این الگوریتم در مسائل طبقه‌بندی و رگرسیون مورد استفاده قرار می‌گیرد و قادر است داده‌ها را در فضای n بعدی به طور مؤثر طبقه‌بندی نماید. KNN با ارائه یک رویکرد انعطاف‌پذیر، به تحلیل و پیش‌بینی الگوها در داده‌های پیچیده کمک می‌کند (Kombo et al., 2020). این نظریه بر پایه اصل همسایگی جغرافیایی بنا نهاده شده است که بیان می‌کند اشیاء یا عناصر دارای مختصات مکانی مشابه، در نهایت خصوصیات یکسانی را به خود می‌گیرند. الگوریتم مبتنی بر این مفهوم، در زمینه‌های پیش‌بینی پدیده‌هایی مانند سیلاب، از طریق فرآیندی مشابه رأی‌گیری برای اشیاء فضایی، نتایج امیدوارکننده‌ای به دست آورده است. همان‌طور که در منابع مختلف ذکر شده، است فاکتورهایمانند فاصله در این الگوریتم نقش کلیدی دارند (Costache et al., 2019;)

1. Logistic Model Tree
2. Kernel SVM Algorithms
3. Support Vector Machine (SVM)
4. K-Nearest Neighbor

مورد استفاده در پژوهش، توانایی خودکار این الگوریتم‌ها در شناسایی و مدل‌سازی ویژگی‌های مهم است. این الگوریتم‌ها با طراحی هوشمندانه‌ای توسعه یافته‌اند که به آن‌ها امکان بهره‌برداری مؤثر از داده‌های اضافی را می‌دهد و به همین دلیل، انتخاب ویژگی به‌عنوان مرحله‌ای ضروری در فرایند یادگیری در نظر گرفته نشده است. همچنین جهت فرایند استانداردسازی داده‌ها جهت مدل‌سازی، از روش Robust استفاده شده است. این روش بر اساس میانه و دامنه داده‌ها عمل می‌کند و به‌ویژه برای داده‌هایی که دارای مقادیر پرت هستند، بسیار مناسب است.

• شاخص (IGR)^۲

در این مطالعه از شاخص IGR، که اولین بار توسط کوویلین^۳ در سال ۱۹۹۳ معرفی شده است، جهت ارزیابی قابلیت پیش‌بینی عوامل کمی مؤثر بهره گرفته شده است. مقادیر بالاتر IGR، نشان‌دهنده قدرت پیش‌بینی قوی‌تر این عوامل در مدل‌سازی است. این شاخص، ابزاری ارزشمند در تعیین عوامل کلیدی در فرایند مدل‌سازی و پیش‌بینی است (Shirzadi et al., 2017). این شاخص از جدول ۳ محاسبه خواهد شد:

جدول ۳. روابط مورد استفاده در شاخص IGR

روابط	شرح روابط
$Info(S) = - \sum_{i=1}^2 \frac{n(L_i, S)}{ S } \log_2 \frac{n(L_i, S)}{ S }$	S داده‌های آموزشی با n نمونه ورودی و $n(L_i, S)$ تعداد نمونه‌ها در داده‌های آموزشی S و متعلق به کلاس L_i است.
$Info(S, A) = \sum_{j=1}^m \frac{S_j}{ S } \log_2 Info(S)$	مقدار اطلاعات مورد نیاز برای تقسیم S به مجموعه $(S_1, S_2, S_3, \dots, S_m)$
$IGR(S, A) = \frac{Info(S) - Info(S, A)}{SplitInfo(S, A)}$	نحوه محاسبه شاخص IGR برای یک عامل مؤثر خاص
$SplitInfo(S, A) = - \sum_{j=1}^m \frac{S_j}{ S } \log_2 \frac{S_j}{ S }$	Split Info نشان‌دهنده اطلاعات تولید شده توسط تقسیم S داده‌های آموزشی به m زیر مجموعه Split Info است

• تست چندخطی عوامل مؤثر^۴

پرادهان معیاری را معرفی کرد که به طور گسترده‌ای برای ارزیابی خطی بودن متغیرها در مدل‌های احتمالی به کار می‌رود (Pradhan, 2010). این معیار، مفهوم چندخطی بودن را مطرح می‌کند که در آن پیش‌بینی می‌شود متغیری را می‌توان با دقت بالایی در یک مدل رگرسیون چندگانه، به صورت خطی بر اساس سایر متغیرها پیش‌بینی کرد. این روش، متغیرهای مستقل مهم را که می‌توانند در مدل به طور مؤثرتری مورد استفاده قرار گیرند، با استفاده از عامل تورم واریانس^۵ مشخص می‌کند. بررسی تست هم‌خطی در متغیرها، گامی ضروری در فرایند مدل‌سازی محسوب می‌شود (Arabameri et al., 2020). در این پژوهش، دوازده متغیر جهت تحلیل حساسیت سیل‌خیزی مورد استفاده قرار گرفته است. این شاخص طبق جدول ۴ محاسبه می‌شود.

جدول ۴. روابط مورد استفاده در تست چندخطی

روابط	شرح روابط
$VIF = \frac{1}{Tolerance}$	VIF (ضریب تورم واریانس) جهت تشخیص هم‌خطی در رگرسیون چندگانه و Tolerance عامل تحمل است.

• اعتبارسنجی و ارزیابی الگوریتم‌ها

در این پژوهش، پنج روش آماری مختلف برای ارزیابی و اعتبارسنجی نتایج مدل‌سازی به کار گرفته شده است. این روش‌ها شامل حساسیت، ویژگی، شاخص‌های مثبت و منفی و تحلیل ROC-AUC هستند. این تحلیل‌ها بر اساس چهار دسته موارد مثبت واقعی،

1. Feature Selection Methods
2. Information Gain Ratio (IGR)
3. Quinlan
4. Multicollinearity Test
5. Variance Inflation Factor (VIF)

منفی واقعی، مثبت کاذب و منفی کاذب طبقه‌بندی انجام شده‌اند (Shirzadi et al., 2018). شاخص‌های نام برده بر اساس جدول ۵ محاسبه می‌شود.

جدول ۵. شاخص‌های ارزیابی الگوریتم‌های مورد استفاده در پژوهش

شاخص	شرح روابط
$SST = \frac{TP}{TP+FN}$	TP شاخص مثبت واقعی و FN، مکان‌هایی را نشان می‌دهند که به درستی و نادرست به عنوان نقاط سیل خیز شناسایی شده‌اند.
$SPF = \frac{TN}{TN+FP}$	TN شاخص منفی واقعی و FP شاخص مثبت کاذب
$PPV = \frac{TP}{FP+TP}$	TP شاخص مثبت واقعی و FP شاخص مثبت کاذب
$NPV = \frac{TP}{TP+FN}$	TP شاخص مثبت واقعی و FN، مکان‌هایی را نشان می‌دهند که به درستی و نادرست به عنوان نقاط سیل خیز شناسایی شده‌اند.

• منحنی ROC- AUC

ROC- AUC به عنوان یک روش معتبر در زمینه اعتبارسنجی شناخته می‌شود. این تکنیک با ارائه یک منحنی به طور دقیق و بصری، توانایی تمایز بین رویدادها و غیررویدادها را نشان می‌دهد. منحنی ROC، ابزاری گرافیکی است که صحت طبقه‌بندی را به نمایش می‌گذارد (Fratini et al., 2010; Chen et al., 2021). این شاخص بر اساس جدول ۶ محاسبه می‌شود.

جدول ۶. روابط مورد استفاده در منحنی ROC- AUC

شاخص	شرح روابط
$AUC = \frac{\sum TP + \sum TN}{P+N}$	TP شاخص مثبت واقعی و TN شاخص منفی واقعی

۳- یافته‌ها

۳-۱- نتایج به دست آمده از شاخص IGR

با توجه به جدول ۷، پارامترهای کاربری اراضی، بارش، شیب، لیتولوژی، فاصله از رودخانه و شاخص SPI از مهم‌ترین عوامل تأثیرگذار بر سیل‌خیزی منطقه مطالعاتی هستند. IGR صفر به عنوان یک معیار جهت شناسایی عوامل غیرضروری عمل می‌کند و کنار گذاشتن این عوامل، منجر به ساده‌سازی مدل و افزایش قابلیت تفسیر آن می‌گردد. این امر همچنین به اعتبار الگوریتم‌ها می‌افزاید؛ زیرا الگوریتم‌ها بدون در نظر گرفتن عوامل کم‌تأثیر، به واقعیت نزدیک‌تر می‌شود و نتایج دقیق‌تری ارائه می‌دهد. عواملی که مقدار IGR آن‌ها صفر به دست آمده است، ارتباطی با متغیرهای اصلی ندارند و به همین دلیل، حذف آن‌ها از مدل‌سازی ضروری است. این عوامل به دلیل همبستگی منفی با متغیر وابسته، تأثیر متقابلی بر یکدیگر دارند. حذف این عوامل، به بهبود دقت پیش‌بینی کمک می‌کند؛ زیرا الگوریتم‌ها می‌توانند با تمرکز بر داده‌های مرتبط، عملکرد بهینه‌تری داشته باشند.

جدول ۷. تعیین مهم‌ترین عوامل مؤثر بر سیل‌خیزی با استفاده از شاخص IGR

عوامل مؤثر	میانگین IGR	عوامل مؤثر	میانگین IGR
کاربری اراضی	۰/۵۸۸	شیب	۰/۴۸۹
جهت شیب	۰/۰۰۰	لیتولوژی	۰/۳۳۳
انحنا سطح	۰/۰۰۰	فاصله از رودخانه	۰/۴۰۱
بارش	۰/۴۲۷	تراکم زهکشی	۰/۰۰۰
ارتفاع	۰/۰۰۰	شاخص TWI	۰/۰۰۰
خاک	۰/۰۰۰	شاخص SPI	۰/۱۶۴

۳-۲- تست چندخطی عوامل مؤثر

در این مطالعه، آستانه ۷ جهت محاسبه VIF به کار گرفته شده است تا پیش‌بینی‌کننده‌های مستقل با اطمینان بالا شناسایی شوند. نتایج جدول ۸ در آزمایش هم‌خطی چندگانه، حاکی از آن است که تمامی عوامل سیل، مقادیر VIF کمتر از ۱۰ و تلوئانس بیشتر از ۰/۲ را کسب نموده‌اند که عدم تداخل و همپوشانی عوامل را در این زمینه نشان می‌دهد.

جدول ۸. آزمون تشخیص چندخطی عوامل مؤثر بر سیل‌خیزی

عوامل مؤثر	VIF	عوامل مؤثر	VIF
کاربری اراضی	۱/۱۱۴	شیب	۱/۷۲۸
جهت شیب	۴/۲۲۲	لیتولوژی	۱/۶۱۲
انحنای سطح	۳/۰۱۴	فاصله از رودخانه	۲/۷۷۱
بارش	۱/۵۱۲	تراکم زهکشی	۱/۵۰۵
ارتفاع	۱/۰۷۵	شاخص TWI	۲/۰۴۵
خاک	۱/۴۳۵	شاخص SPI	۲/۱۱۸

۳-۳- تکنیک‌های اعتبارسنجی الگوریتم‌ها

نتایج ارزیابی و عملکرد الگوریتم‌ها در جدول ۹ نشان می‌دهد که الگوریتم‌های مورد استفاده در مجموعه داده‌های تعلیمی و صحت‌سنجی، شاخص‌های آماری مشابهی را به نمایش می‌گذارند. در این میان، الگوریتم LT بالاترین مقادیر شاخص‌های SST، SPF، PPV و NPV را در هر دو مجموعه داده به خود اختصاص داده است. الگوریتم K-SVM با بالاترین مقادیر SST (۰/۷۷) و PPV (۰/۷۱) در رتبه دوم قرار می‌گیرد. الگوریتم KNN نیز عملکرد قابل توجهی دارد و بالاترین مقادیر SST (۰/۶۸) و PPV (۰/۶۵) را در میان الگوریتم‌های مورد بررسی به دست آورده است. این نتایج حاکی از آن است که هر سه الگوریتم، دقت و عملکرد خوبی را در تجزیه و تحلیل مجموعه داده‌ها نشان می‌دهد. در جدول ۱۰ نیز شاخص‌های مربوط به ماتریس متقاطع مورد استفاده در پژوهش ارائه شده است.

جدول ۹. شاخص‌های اعتبارسنجی الگوریتم‌ها

شاخص	داده‌های تعلیمی			داده‌های صحت‌سنجی		
	LT	K-SVM	KNN	LT	K-SVM	KNN
SST	۰/۸۰	۰/۷۷	۰/۶۸	۰/۷۸	۰/۶۳	۰/۶۳
SPF	۰/۸۶	۰/۶۷	۰/۵۹	۰/۶۹	۰/۶۱	۰/۵۹
PPV	۰/۷۹	۰/۷۱	۰/۶۵	۰/۵۸	۰/۵۶	۰/۵۶
NPV	۰/۷۳	۰/۶۲	۰/۵۴	۰/۵۱	۰/۵۲	۰/۵۰

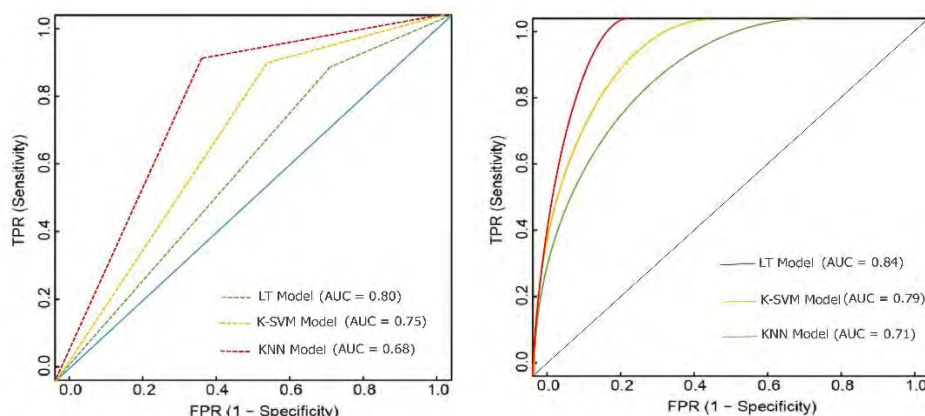
جدول ۱۰. شاخص‌های ماتریس متقاطع داده‌های آموزشی و آزمایشی

معیار	داده‌های آموزشی	داده‌های آزمایشی	داده‌های آموزشی	داده‌های آزمایشی
TP	۱۵۰	۱۲۰	۱۴۰	۱۱۰
FP	۳۰	۲۵	۲۰	۱۵
TN	۲۰۰	۱۸۰	۲۱۰	۱۹۰
FN	۲۰	۱۵	۳۰	۲۵

۳-۴- نتایج منحنی ROC-AUC

در پژوهش حاضر صحت الگوریتم‌های مورد استفاده با منحنی ROC مورد بررسی قرار گرفته است. طبق نتایج به دست آمده مقدار AUC الگوریتم LT برابر با ۰/۸۴، الگوریتم K-SVM برابر با ۰/۷۹ و الگوریتم KNN برابر با ۰/۷۱ به دست آمده است (شکل ۱). مقادیر AUC به دست آمده حاکی از آن است که الگوریتم‌های مورد استفاده در پژوهش از عملکرد مناسبی برخوردارند. همچنین با

توجه به شکل ۲ نیز مجموعه داده‌های آموزشی در الگوریتم LT برابر با ۰/۸۰، الگوریتم K-SVM برابر با ۰/۷۵ و الگوریتم KNN برابر با ۰/۶۸ به دست آمده است که نشان‌دهنده عملکرد بهینه الگوریتم‌ها در مجموعه داده‌های آموزشی نیز است.

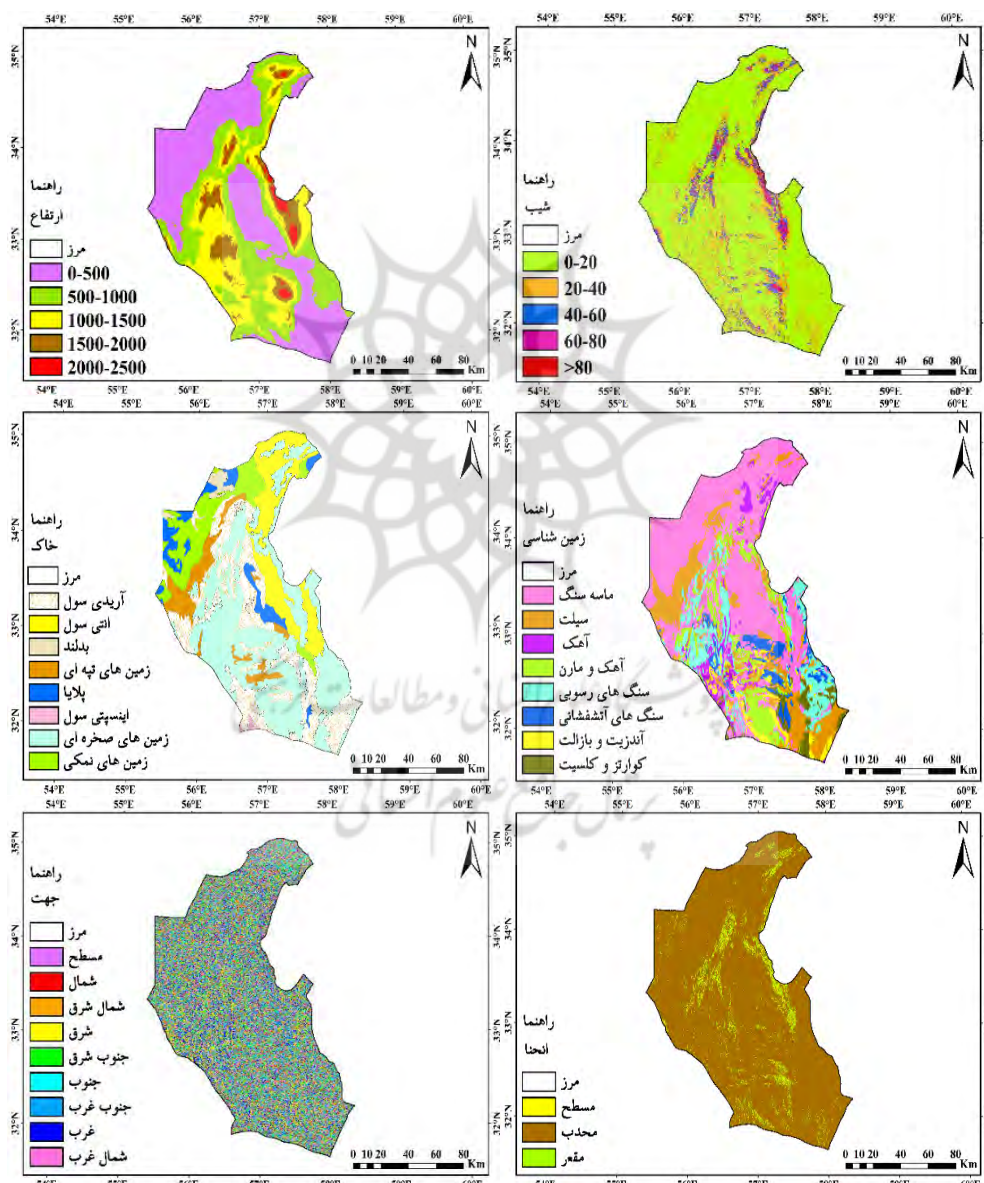


شکل ۱. نتایج صحت‌سنجی منحنی ROC-AUC شکل ۲. نتایج داده‌های آموزشی در منحنی ROC-AUC

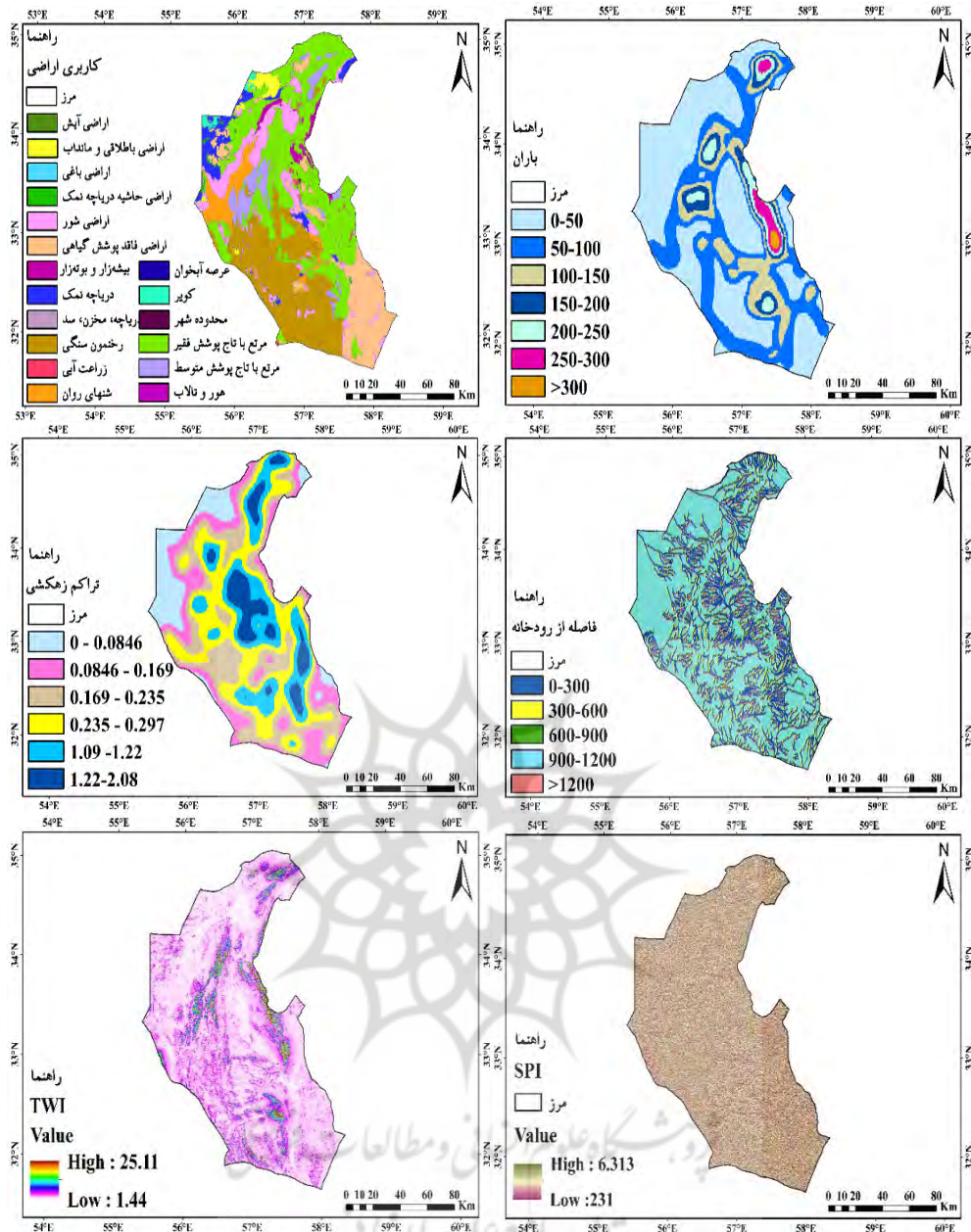
• بحث

در پژوهش حاضر، دوازده شاخص که به طور مستقیم بر روی پدیده سیل‌خیزی تأثیرگذارند، به منظور توسعه و تجزیه و تحلیل الگوریتم‌ها انتخاب شدند. شیب زمین نقش قابل توجهی در بروز سیل در حوزه‌های آبخیز ایفا می‌کند. با افزایش شیب، نفوذ رواناب کاهش می‌یابد و در نتیجه رواناب با سرعت و قدرت بیشتری جاری می‌شود (Shahabi, 2021). در منطقه مورد مطالعه در شیب بیش از ۲۰ درجه که ۶۸/۴۴ درصد از سطح منطقه را شامل می‌شود، رواناب ناشی از بارندگی افزایش می‌یابد که نقش بسیار مهمی در وقوع جریان‌های سیلابی ایفا می‌کند. توپوگرافی از دیگر عوامل مهمی است که با تأثیر بر شیب دامنه‌ها و رطوبت خاک، به طور مستقیم بر رواناب و در نتیجه بر پتانسیل وقوع سیلاب‌ها نقش دارد (Rezaei Moghaddam et al., 2021). اختلاف ارتفاع قابل توجه در منطقه مطالعاتی، که از ۵۰۰ متر تا ۲۰۰۰ متر متغیر است، تأثیر به سزایی بر هیدرولوژی منطقه دارد. ناهمواری‌های موجود، پتانسیل وقوع سیل و خطرات مرتبط با آن را در منطقه آشکار می‌سازد. در رابطه با عامل زمین‌شناسی، میزان نفوذپذیری سازندها عاملی کلیدی در ایجاد سیل‌خیزی محسوب می‌شود، به‌ویژه در منطقه مورد بررسی که عمدتاً از سازندهای آتشفشانی تشکیل شده است. به طور کلی منطقه مطالعاتی دارای سازندهایی با نفوذپذیری کم است که این ویژگی در ارتباط با سیل‌خیزی، حائز اهمیت است. سنگ‌های بازالتی به دلیل نفوذناپذیری یا نفوذپذیری بسیار کم، نقش مهمی در تولید رواناب دارند، در حالی که سازندهای آهکی و ماسه‌ای به دلیل نفوذپذیری بالا، از اهمیت کمتری در این زمینه برخوردارند. خاک‌های آریدی‌سول در منطقه، به دلیل قلیایی بودن و خشکی در جذب رواناب ضعیف هستند. در شرایط بارندگی شدید و ناگهانی، محدودیت در نفوذپذیری خاک‌های منطقه، می‌تواند منجر به تجمع رواناب روی سطح زمین و در نتیجه، افزایش احتمال وقوع سیل‌خیزی در منطقه شود. همچنین، ساختار غیرقابل نفوذ پلایا، مانع از جذب سریع رواناب می‌شود و در مواقع بارندگی شدید، سطح رواناب‌ها به سرعت افزایش می‌یابد و خطر سیل‌خیزی را تشدید می‌کند. همچنین در قسمت‌هایی از منطقه با تحدب بالا، رواناب‌های حاصل از بارندگی به سرعت تخلیه می‌شوند و به همین دلیل، تجمع آب در این نواحی کمتر است. این امر به کاهش چشمگیر خطر سیل‌خیزی در این مناطق منجر می‌شود. همچنین، در مناطقی با تحدب بسیار کم یا سطوح مقعر، رواناب حاصل از باران تجمع می‌یابد، که می‌تواند باعث افزایش احتمال سیل‌خیزی و آب‌گرفتگی در این نواحی شود. الگو کاربری اراضی نیز از مهم‌ترین عوامل کلیدی در بروز سیلاب‌ها محسوب می‌شود. نوع کاربری اراضی تأثیر به سزایی بر جریان آب‌های سطحی و فرسایش خاک دارد که در نهایت منجر به تغییر در میزان وقوع سیلاب‌ها می‌شود (Benito et al., 2010). تبدیل زمین‌های طبیعی به مزارع کشاورزی، اثرات قابل توجهی بر هیدرولوژی منطقه دارد. شیوه‌های کشاورزی خاص، مانند فشردگی خاک، می‌تواند منجر به کاهش نفوذ آب و افزایش رواناب سطحی شود. این امر به ویژه در مناطقی که فاقد پوشش گیاهی و دارای زمین‌های صخره‌ای هستند، باعث افزایش فرسایش خاک می‌گردد. در نتیجه، سیلاب‌ها در پایین‌دست منطقه بیشتر رخ می‌دهد. مناطقی که نزدیک‌ترین فاصله از رودخانه دارند، به ویژه در دشت‌های سیلابی، با خطر سیلاب شدیدی در منطقه مواجه هستند. نزدیکی به رودخانه، احتمال وقوع سیل را افزایش می‌دهد و به همین دلیل، این مناطق

مستعد آسیب‌های ناشی از سیلاب هستند. در مناطقی با تراکم زهکشی بالا، رودخانه‌ها مستعد وقوع سیلاب‌های شدیدتر و پرخطرتر هستند. این امر نشان‌دهنده ارتباط مستقیم بین تراکم زهکشی و پتانسیل سیل‌خیزی رودخانه‌ها است. در مقایسه با رودخانه‌های با تراکم پایین، مناطق با تراکم زهکشی بیشتر، شاهد افزایش احتمال وقوع سیلاب‌های مخرب‌تر خواهند بود (Chapi et al., 2017). در منطقه مطالعاتی نواحی نزدیک به رودخانه پتانسیل بالایی در وقوع سیلاب دارند. عامل مهم دیگری که نقش به سزایی در سیل‌خیزی دارد، بارندگی است (Sharma et al., 2017; Das & Scaringi 2021; Breinl et al., 2021). اقلیم نیمه‌خشک منطقه مطالعاتی، عامل اصلی نوسانات بارندگی در طول فصول مختلف است. بارندگی‌های فصلی، به ویژه در مواقع همزمانی با ذوب برف، می‌تواند به بروز سیلاب‌های ناگهانی و غیرقابل پیش‌بینی منجر شوند. همچنین در منطقه مورد مطالعه، شاخص‌های SPI و TWI با مقادیر بالا، مناطقی را مشخص می‌کنند که مستعد اشباع و تجمع رواناب هستند. این مناطق به دلیل شرایط هیدرولوژیکی، در معرض خطر سیلاب بیشتری قرار دارند. این شاخص‌ها نشان می‌دهند که احتمال وقوع سیل در این نواحی بیشتر است و نیاز به مدیریت و برنامه‌ریزی دقیق برای کاهش اثرات آن وجود دارد (شکل ۳).



شکل ۳. عوامل مؤثر بر سیلاب در منطقه مورد مطالعه



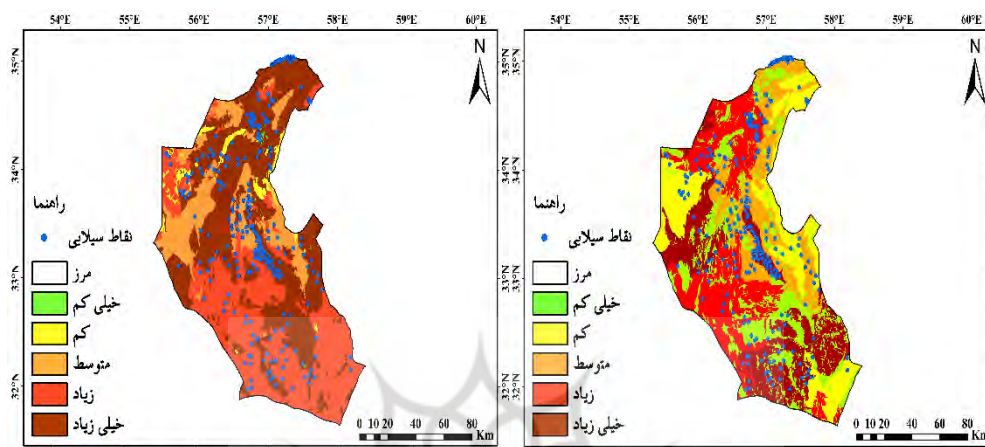
ادامه شکل ۳. عوامل مؤثر بر سیلاب در منطقه مورد مطالعه

• تهیه نقشه‌های پهنه‌بندی سیلاب

در این پژوهش، سه نقشه پتانسیل خطر سیل‌خیزی با استفاده از الگوریتم‌های یادگیری ماشین و در نظر گرفتن ۱۲ پارامتر تأثیرگذار بر سیلاب، تولید شده است. داده‌های مربوط به منطقه مورد مطالعه، پس از پردازش و آموزش در الگوریتم‌ها، برای محاسبه و پیش‌بینی سیلاب مورد استفاده قرار گرفتند. در این فرآیند، تمامی پیکسل‌های منطقه به الگوریتم‌ها آموزش داده شد تا قابلیت پیش‌بینی دقیق‌تر را فراهم کند. در مرحله نهایی، نقشه‌های پتانسیل سیل در محیط Arc GIS 10.4 ایجاد شد و با استفاده از آستانه‌های مشخص، منطقه به پنج طبقه مختلف از نظر شدت سیل با روش شکست طبیعی^۱ تقسیم‌بندی گردید. این طبقات شامل خیلی کم، کم، متوسط، زیاد، و خیلی زیاد است که نشان‌دهنده توزیع خطر سیل در منطقه مورد مطالعه است (شکل ۴ تا ۶). مطابق با نتایج حاصل از الگوریتم‌های مختلف، مناطق مستعد سیل در این منطقه به طور قابل توجهی متفاوت است. در الگوریتم LT، مناطق شمالی و غربی

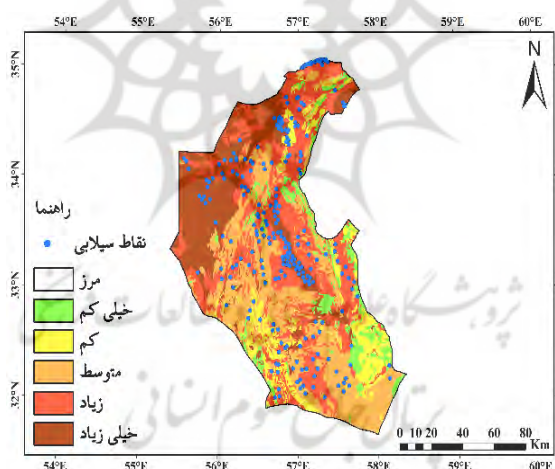
1. Natural Break

با خطر سیل خیزی زیاد (۴۳٪) و نواحی غربی و جنوبی با خطر خیلی زیاد (۳۶٪) مواجه هستند. در مقابل، الگوریتم K-SVM نشان می‌دهد که قسمت‌های مرکزی، غربی و جنوبی منطقه در معرض سیل خیزی زیاد (۵۹٪) قرار دارند، در حالی که نواحی شرقی، شمالی و مرکزی خطر سیل خیزی خیلی زیاد (۴۱٪) را نشان می‌دهند. جالب توجه است که الگوریتم KNN پهنه‌بندی متفاوتی را ارائه می‌کند، جایی که نواحی شمالی، شرقی و مرکزی با خطر سیل خیزی زیاد (۳۸٪) و مناطق شمال غربی، غربی و مرکزی با خطر سیل خیزی خیلی زیاد (۲۷٪) شناسایی شده‌اند. تفاوت در نتایج الگوریتم KNN در مقایسه با سایر الگوریتم‌ها، به عواملی از جمله حساسیت الگوریتم به مقیاس داده‌ها و همچنین تکنیک‌های تحلیل مکانی مورد استفاده در پردازش داده‌ها است. این عوامل به دلیل ماهیت خاص داده‌ها، مقیاس‌بندی و پیش‌پردازش دقیق و مناسب این الگوریتم است.



شکل ۵. نقشه پهنه‌بندی سیل با مدل K-SVM

شکل ۴. نقشه پهنه‌بندی سیل با مدل LT



شکل ۶. نقشه پهنه‌بندی سیل با مدل KNN

با توجه به نقشه‌های پهنه‌بندی سیل، نواحی با سیل‌خیزی زیاد و خیلی زیاد در شیب بیش از ۲۰ درجه واقع شده است که این ویژگی تأثیر قابل توجهی بر وقوع سیل‌خیزی در منطقه دارد. علاوه بر شیب رودخانه اصلی، شیب اراضی نیز عاملی کلیدی در تشدید سیل‌خیزی منطقه است. در منطقه مطالعاتی رودها و شاخه‌های فرعی آن مسیر خود را در امتداد دامنه‌ها طی می‌کنند. مناطق مستعد سیل در قسمت‌های کم ارتفاع واقع شده‌اند. در نتیجه، میزان فرسایش و حمل مواد آبرفتی و رسوبی در این نواحی به شدت افزایش می‌یابد. از لحاظ فاصله از شبکه آبراهه، مناطقی که در نزدیکی آبراهه قرار دارند، به دلیل جریان دائمی و پرآب و همچنین طغیان‌های رودخانه در فصل بهار، در معرض خطر بالایی قرار دارند. مطالعات نشان می‌دهد که قسمت‌هایی از منطقه با سازندهای سطحی مقاوم از نظر زمین‌شناسی و لیتولوژی، نفوذپذیری کم و پوشش گیاهی اندک، مستعد وقوع سیل‌خیزی هستند. به طور کلی، با توجه به نوع سازندها، سیلاب‌ها باعث ورود حجم قابل توجهی از رسوبات به بستر رودخانه می‌شوند که با نتایج پژوهش‌های اسلام و همکاران و

آوند و همکاران مبنی بر سازندهای زمین‌شناسی و خاک‌هایی با نفوذپذیری کم، شیب تند و فقدان یا ضعف پوشش گیاهی مطابقت دارد (Towfiqul Islam et al., 2021; Avand et al., 2021). در شکل ۷ تصاویری از سیل‌های رخ داده در منطقه ارائه شده است.



شکل ۷. نمایی از وضعیت سیل‌خیز بستر رودخانه طبس

۴- نتیجه‌گیری

مناطق نیمه خشک اغلب با الگوهای بارندگی نامنظم و رویدادهای بارندگی نادر اما شدید مواجه است. هدف از پژوهش، ارزیابی و پهنه‌بندی خطر سیل‌خیزی در منطقه نیمه‌خشک طبس با استفاده از الگوریتم‌های یادگیری است. سپس ۱۲ عامل کلیدی مؤثر در وقوع سیلاب بر اساس شاخص IGR مورد بررسی قرار گرفتند. طبق نتایج به دست آمده، پارامترهای کاربری اراضی، بارش، شیب، لیتولوژی، فاصله از رودخانه و شاخص SPI از مهم‌ترین عوامل تأثیرگذار بر سیل هستند. سپس عملکرد الگوریتم‌ها با استفاده از شاخص‌های آماری مورد ارزیابی قرار گرفت. نتایج نشان داد که الگوریتم‌های مختلف در مجموعه داده‌های تعلیمی و صحت‌سنجی، روند آماری مشابهی را نشان می‌دهند. در این میان، الگوریتم LT بالاترین مقادیر را در هر دو مجموعه داده به خود اختصاص داده است. الگوریتم K-SVM با بالاترین مقادیر $SST (0/77)$ و $PPV (0/71)$ در رتبه بعدی قرار می‌گیرد و نشان‌دهنده آن است که هر سه الگوریتم، دقت و عملکرد مناسبی در پهنه‌بندی خطر سیل‌خیزی دارند. همچنین مقادیر به دست آمده AUC در الگوریتم LT برابر با ۰/۸۴، الگوریتم K-SVM برابر با ۰/۷۹ و الگوریتم KNN برابر با ۰/۷۱ به دست آمد که بیانگر عملکرد مناسب الگوریتم‌ها است. در فرآیند پهنه‌بندی سیل‌خیزی، ضروری است که محدودیت‌های الگوریتم‌ها را در نظر گرفت. الگوریتم LT بر اساس فرضیات خطی بین متغیرها بنا شده است؛ اما این فرض در مورد سیل که یک پدیده پیچیده و غیرخطی است، ممکن است دقیق نباشد. در الگوریتم KNN نیز با افزایش حجم داده‌ها، زمان محاسبات افزایش می‌یابد که می‌تواند در کاربرد دنیای واقعی محدودیت‌هایی ایجاد کند. علاوه بر این در الگوریتم K-SVM به دلیل پیچیدگی محاسباتی، به ویژه در مواجهه با مجموعه داده‌های بزرگ، نیازمند منابع و زمان محاسباتی قابل توجهی است. با به کارگیری روش‌های پیشرفته نوین‌سازی و بهینه‌سازی داده‌ها، می‌توان محدودیت‌های موجود در مدل‌سازی را کاهش داد و دقت الگوریتم‌ها را بهبود بخشید. تکنیک‌های نوین یادگیری عمیق به عنوان ابزارهای قدرتمندی در شناسایی و حذف نویز از داده‌ها عمل می‌کنند. این تکنیک‌ها به عنوان فیلترهایی عمل می‌کنند که انحرافات را تشخیص داده و تصحیح می‌کنند. علاوه بر این، توسعه الگوریتم‌های مبتنی بر یادگیری تقویتی، رویکردی مؤثر در تنظیم خودکار پارامترهای مدل ارائه می‌دهد. این امر نه تنها دقت مدل را افزایش می‌دهد، بلکه کارایی کلی الگوریتم‌ها را نیز بهبود می‌بخشد. ارزیابی تأثیر این روش‌ها بر دقت و عملکرد الگوریتم‌ها، گام مهمی در جهت بهینه‌سازی فرآیند مدل‌سازی است. به طور کلی استفاده از هوش مصنوعی در مدل‌سازی سیلاب این امکان را فراهم می‌کند تا با دقت بیشتری رفتار سیلاب‌ها و تغییرات سطح زمین را پیش‌بینی نمود. همچنین استفاده از الگوریتم‌های هوش مصنوعی جهت ارزیابی ریسک سیلاب و ارائه پیشنهادات جهت مدیریت منابع آب و زمین کارآمد است. طبق نتایج به دست آمده از نقشه‌های پهنه‌بندی سیل‌خیزی، مناطق با خطر سیلاب زیاد در نزدیکی محل تلاقی آبراهه‌ها قرار دارند. این پدیده ناشی از جریان آب‌های زیرسطحی از آبراهه‌ها به سمت دامنه‌های مجاور است که منجر به ایجاد رطوبت در خاک و تضعیف مقاومت آن می‌شود. این روند به همراه تراکم بالای زهکشی، خطر سیلاب را در این نواحی تشدید می‌کند. مناطق غربی، مرکزی، شمالی و جنوب‌شرقی، به دلیل ویژگی‌های توپوگرافی مانند

ارتفاع، شیب تند، بارش زیاد و تراکم شبکه آبراهه‌ها، مستعد سیلاب شناخته می‌شوند. فعالیت‌های انسانی مانند چرای بی‌رویه دام و تخریب پوشش گیاهی، نفوذپذیری خاک را کاهش داده و رواناب و شدت سیلاب را در این مناطق افزایش داده است. در مقابل، مناطق کم‌خطر سیلاب، معمولاً دارای پوشش گیاهی متراکم، دور از آبراهه‌های اصلی و با شبکه زهکشی کم‌تراکم هستند. این مناطق از نظر خاک، ارتفاع و شیب شرایط متفاوتی دارند و به همین دلیل، کمتر در معرض سیل‌خیزی قرار می‌گیرند. جهت کاهش خطر سیلاب در این منطقه، ضروری است اقدامات و راهکارهایی اجرا شود: از جمله احداث مخازن ذخیره آب در مناطق مرکزی طبرستان با هدف حفظ منابع آبی و کاهش آسیب‌پذیری در برابر سیلاب، حفاظت از رودخانه‌های طبرستان و نایبند، به‌ویژه در دهستان نایبند و آسمان آباد، از طریق مدیریت تراز آب و کنترل جریان سیلاب، ساخت دیوارهای محافظتی در نواحی حاشیه‌ای مانند سرخ پوست و گراب، جهت جلوگیری از ورود آب به این مناطق در مواقع سیل، طراحی و اجرای سیستم‌های زهکشی کارآمد در مناطقی مانند روستای مهرآباد، به منظور تسهیل در تخلیه آب‌های سطحی، احیای اکوسیستم رودخانه‌ای در دره سبز و نهر چشمه کمر، با کاشت گونه‌های گیاهی بومی که به‌عنوان سدهای طبیعی عمل می‌کنند. این اقدامات نه تنها به حفظ منابع آبی کمک می‌کند، بلکه از آسیب‌های جدی ناشی از سیلاب در این مناطق جلوگیری خواهد کرد. نتایج به دست آمده از پژوهش حاضر جهت حفظ منابع طبیعی، خاک، پوشش گیاهی و مدیریت بحران مفید است.

۵- سپاس‌گزاری

نویسندگان از حمایت‌های معنوی دانشگاه حکیم سبزواری تشکر می‌کنند.

۶- منابع

- ابراهیمی، لایلا و ایلانلو، مریم. (۱۴۰۳). پهنه‌بندی وقوع سیلاب حوضه زهکشی شهرستانک با استفاده از مدل هیدرولوژیکی WMS و تلفیق GIS. مدیریت مخاطرات محیطی، (۱) ۱۱، ۱۵-۲۹.
- رحیم پور، توحید و رضائی‌مقدم، محمدحسین. (۱۴۰۳). مدل‌سازی پتانسیل خطر وقوع سیلاب در حوضه آبریز آجی چای با استفاده از الگوریتم‌های داده‌کاوی. پژوهش‌های فرسایش محیطی، (۴) ۱۴، ۱۹-۳۸. رضائی‌مقدم، محمدحسین، مختاری، داوود و شفیعی - مهر، مجید. (۱۴۰۰). پهنه‌بندی خطر سیلاب در حوضه آبریز شهر چای میانه با استفاده از مدل ویکور. هیدروژئومورفولوژی، ۲۸، ۱۹-۳۷.
- زنگنه‌اسدی، محمدعلی، امیراحمدی، ابوالقاسم و ناعمی‌تبار، مهناز. (۱۴۰۰). ارزیابی کارایی مدل‌های ویکور، L-THIA و شبکه عصبی مصنوعی در تحلیل منطقه‌ای سیلاب (مطالعه موردی: استان خراسان رضوی). اکوهیدرولوژی، (۱) ۸، ۸۹-۱۰۸. ۱۰۸-۸۹.
- سلطانی، محمود. (۱۳۸۱). ارزیابی اراضی مناطق مناسب انتخاب سیلاب در GIS. پایان‌نامه کارشناسی ارشد، گروه مهندسی عمران، دانشگاه طوسی، تهران، ایران.
- قناتی، رویت، سالجقه، علی، پورقاسمی، حمیدرضا، خلیقی سیگارودی، شهرام و کشت‌کار، حمیدرضا. (۱۴۰۴). ارزیابی روش‌های یادگیری ماشین (RF و FDA، GLM، SVM) در تهیه نقشه حساسیت سیل بخشی از استان خوزستان. مدل‌سازی و مدیریت آب و خاک، (۱) ۵، ۲۳۱-۲۴۶.
- کاویان، عطاله، شاهدی، کاکا، چمنی، رضا و دهقانی، مرتضی. (۱۴۰۰). ارزیابی روند تغییرات زمانی- مکانی متغیرهای اقلیمی استان خراسان جنوبی با استفاده از GIS و تکنیک‌های آماری. ترویج و توسعه آبخیزداری، (۳۳) ۹، ۵۴-۶۷.
- یمانی، مجتبی و عنایتی، مریم. (۱۳۸۴). ارتباط ویژگی‌های ژئومورفولوژیک حوضه‌ها و قابلیت سیل‌خیزی (تجزیه و تحلیل داده‌های سیل از طریق مقایسه ژئومورفولوژیک حوضه‌های فشنند و بهجت آباد). پژوهش‌های جغرافیایی، ۵۴، ۴۷-۵۷.

References

- Adedoyin Benson, A., & Akinola Adesuji, K. (2025). Flood hazard zones prediction using mashin-learning-based geospatial approach in lower niger river basin, Nigeria. Natural Hazards Research, <https://doi.org/10.1016/j.nhres.2025.01.002>.
- Agrawal, P., Ganesh, T., & Mohamed, A.W. (2020). A Novel Binary Gaining-Sharing Knowledge-Based Optimization Algorithm for Feature Selection. Neural. Comput. Appl. 33, 5989-6008. <https://doi.org/10.1007/s00521-020-05375>.

- Ahmadlou, M., Karimi, M., Alizadeh, S., Shirzadi, A., Parvinnejhad, D., Shahabi, H., & Panahi, M. (2019). Flood susceptibility assessment using integration of adaptive networkbased fuzzy inference system (ANFIS) and biogeography-based optimization (BBO) and BAT algorithms (BA). *Geocarto Int*, 34 (11), 1252–1272. <https://doi.org/10.1080/10106049.2018.1474276>.
- Al-Areeq, A.M., Abba, S.I., Yassin, M.A., Benaafi, M., Ghaleb, M., & Aljundi, I.H. (2022). Computational Machine Learning Approach for Flood Susceptibility Assessment Integrated with Remote Sensing and GIS Techniques from Jeddah, Saudi Arabia. *Remote Sens.* 14, 5515. <https://doi.org/10.3390/rs142155153>.
- Ali, S. A., Parvin, F., Pham, Q. B., Vojtek, M., Vojteková, J., Costache, R., & Ghorbani, M. A. (2020). GIS-based comparative assessment of flood susceptibility mapping using hybrid multi-criteria decision-making approach, naïve Bayes tree, bivariate statistics and logistic regression: a case of Topľa basin, Slovakia. *Ecological Indicators*, 117, 106620. <https://doi.org/10.1016/j.ecolind.2020.106620>.
- Arabameri, A., Chen, W., Loche, M., Zhao, X., Li, Y., Lombardo, L., Cerda, A., Pradhan, B., & Bui, D.T. (2020). Comparison of machine learning models for gully erosion susceptibility mapping. *Geosci. Front.* 11, 1609–1620. <https://doi.org/10.1016/j.gsf.2019.11.009>.
- Arabameri, A., Nalivan, O.A., Saha, S., Roy, J., Pradhan, B., Tiefenbacher, J.P., & Ngo, P.T.T. (2020). Novel Ensemble Approaches of Machine Learning Techniques in Modeling the Gully Erosion Susceptibility. *Remote Sens.* 12, 1890. doi:10.3390/rs12111890.
- Avand, M., Moradi, H., & Lasboyee, M.R. (2021). Using Machine Learning Models, Remote Sensing, and GIS to Investigate the Effects of Changing Climates and Land Uses on Flood Probability. *J. Hydrol. (Amst.)*, 595, 125663. <https://doi.org/10.1016/j.jhydrol.2020.125663>.
- Benito, G., Rico, M., Sánchez-Moya, Y., Sopena, A., Thorndycraft, V.R., & Barriendos, M. (2010). The impact of late Holocene climatic variability and land use change on the flood hydrology of the Guadalentín River. southeast Spain. *Glob. Planet. Chang.* 70 (1–4), 53–63. <https://doi.org/10.1016/j.gloplacha.2009.11.007>.
- Breiman, L., Friedman, J.H., Olshen, R.A., & Stone, C.J. (2017). *Classification and Regression Trees*; Routledge: London, UK, ISBN 1315139472.
- Breinl, K., Lun, D., Müller-Thomy, H., & Blöschl, G. (2021). Understanding the relationship between rainfall and flood probabilities through combined intensity-duration-frequency analysis. *J Hydrol* 602:126759. <https://doi.org/10.1016/j.jhydrol.2021.126759>.
- Cao, Q., Mehran, A., Ralph, F.M., & Lettenmaier, D.P. (2019). The role of hydrological initial conditions on Atmospheric River floods in the Russian River basin. *J. Hydrometeorol*, 1667–1686. <https://doi.org/10.1175/JHM-D-19-0030.1>.
- Ceola, S., Domeneghetti, A., & Schumann, GJP. (2022). Unraveling Long-Term Flood Risk Dynamics Across the Murray-Darling Basin Using a Large-Scale Hydraulic Model and Satellite Data. *Front. Water*, 3:797259. doi: 10.3389/frwa.2021.797259.
- Chadee, D. D., Sutherland, J.M. & Agard, J. (2014). *Flooding and climate change, sectorial impacts and adaptation strategies for the Caribbean region*. Nova Science Publishers, Inc. <https://www.ebay.co.uk/itm/196957031357>.
- Chapi, K., Singh, V.P., Shirzadi, A., Shahabi, H., Bui, D.T., Pham, B.T., & Khosravi, K. (2017). A novel hybrid artificial intelligence approach for flood susceptibility assessment. *Environ. Model. Softw.* 95, 229–245. <https://doi.org/10.1016/j.envsoft.2017.06.012>.
- Chen, W., Lei, X., Chakraborty, R., Pal, S.C., Sahana, M., & Janizadeh, S. (2021). Evaluation of different boosting ensemble machine learning models and novel deep learning and boosting framework for head-cut gully erosion susceptibility. *J. Environ. Manag.* 284, 112015. <https://doi.org/10.1016/j.jenvman.2021.112015>.
- Chidiebere, N.C., Godswill, O.C., Anthony, E.C., & Ikechukwu, U. (2024). Effects of Climatic Variations and Changing Land Use/Land Cover on Flooding in Southern Nigeria. *Indonesian Journal of Environmental Management and Sustainability*, 8(1), 16-24. 10.26554/ijems.2024.8.1.16-24.
- Choubin, B., Moradi, E., & Golshan, M. (2019). An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. *Sci Total Environ*, 651, 2087–2096. <https://doi.org/10.1016/j.scitotenv.2018.10.064>.
- Chuma, G.B., Bora, F.S., Ndeko, A.B., Mugumaarhahama, Y., Cirezi, N.C., Mondo, J.M., Bagula, E.M., Karume, K., Mushagalusa, G.N., & Schimtz, S. (2021). Estimation of Soil Erosion Using RUSLE Modeling and Geospatial Tools in a Tea Production Watershed (Chisheke in Walungu), Eastern Democratic Republic of Congo. *Model. Earth Syst. Environ.* 8, 1273–1289. DOI:10.1007/s40808-021-01134-3.

- Costache, R., & Bui, D.T. (2020). Identification of areas prone to flash-flood phenomena using multiple-criteria decision-making, bivariate statistics, machine learning and their ensembles. *Sci. Total Environ.* 136492. <https://doi.org/10.1016/j.scitotenv.2019.136492>.
- Costache, R., Pham, Q.B., Sharifi, E., Linh, N.T.T., Abba, S.I., Vojtek, M., Vojteková, J., Nhi, P.T.T., & Khoi, D.N. (2019). Flash-Flood Susceptibility Assessment Using Multi-Criteria Decision Making and Machine Learning Supported by Remote Sensing and GIS Techniques. *Remote Sens.* 12, 106. <https://doi.org/10.3390/rs12010106>.
- Das, S., & Scaringi, G. (2021). River flooding in a changing climate: rainfall-discharge trends, controlling factors, and susceptibility mapping for the Mahi catchment, Western India. *Nat Hazards.* <https://doi.org/10.1007/s11069-021-04927-y>.
- Frattini, P., Crosta, G., & Carrara, A. (2010). Techniques for evaluating the performance of landslide susceptibility models. *Eng. Geol.* 111, 62–72. <https://doi.org/10.1016/j.enggeo.2009.12.004>.
- Garde, R. J. (2006). *River Morphology*, Published by New Age International (P) Ltd. Publishers, New Delhi.
- Garrote, J. (2022). Free GlobalDEMs and Flood Modelling-AComparison Analysis for the January 2015 Flooding Event in Mocuba City(Mozambique). *Water*, 14, 176. <https://doi.org/10.3390/w14020176>.
- Gholami, H., Mohammadifar, A.A., Golzari, Sh., Torkamandi, R., Moayedi, E., Zare Reshkooyeh, M., Song, Y., & Zeeden, Ch. (2025). Mapping flood risk using a workflow including deep learning and MCDM– Application to southern Iran. *Urban Climate*, 59, 102272. <https://doi.org/10.1016/j.uclim.2024.102272>.
- Heidari, E., Mahmoudzadeh, A., & Mansouri Daneshvar, M.R. (2021). Urban flood susceptibility evaluation and prediction during 2010–2030 in the southern watersheds of Mashhad city, Iran. *Environmental Systems Research*, 10:41. <https://doi.org/10.1186/s40068-021-00245-1>.
- Hitouri, S., Mohajane, M., Lahsaini, M., Ali, S.A., Setargie, T.A., Tripathi, G., D'Antonio, P., Singh, S.K., & Varasano, A. (2024). Flood Susceptibility Mapping Using SAR Data and Machine Learning Algorithms in a Small Watershed in Northwestern Morocco. *Remote Sens.* 16, 858, 1–21. <https://doi.org/10.3390/rs16050858>.
- Kaffash, M., Yousefi, E., & Shariati, M. (2024). Spatial Analysis of Flood Risk in Tabas Watershed Using Satellite Images and Geographic Information System. *Journal of Drought and Climate change Research (JDCR)*, 2(6), 105-118. [10.22077/jdcr.2024.7489.1066](https://doi.org/10.22077/jdcr.2024.7489.1066).
- Kavian, A., Shahedi, K., Chamani, R., & Dehghani, M. (2021). Analysis of Spatial-temporal Trend in Climatology Variable by Using GIS and Statistical Technic (A case study: Southern-Khorasan Province). *Extension and Development of Watershed Management*, 9(33), 54-67. DOR: 20.1001.1.26454777.1400.9.33.7.0. **[In Persian]**
- Kirkby, M.J., & Beven, K.J. A. (1979). physically based, variable contributing area model of basin hydrology. *Hydrol. Sci. Bull.* 24, 43–69. <https://doi.org/10.1080/02626667909491834>.
- Kombo, O., Kumaran, S., Sheikh, Y., Bovim, A., & Jayavel, K. (2020). Long-Term Groundwater Level Prediction Model Based on Hybrid KNN-RF Technique. *Hydrology*, 7, 59. <https://doi.org/10.3390/hydrology7030059>.
- Kumar, V., Vaibhav Sharma, K., Mangukiya, N., Kumar Tiwari, D., Vijay Ramkar, P., & Rathnayake, U. (2025). Machine learning applications in flood forecasting and predictions, challenges, and way-out in the perspective of changing environment. *AIMS Environmental Science*, 12(1), 72–105. [10.3934/environsci.2025004](https://doi.org/10.3934/environsci.2025004).
- Landwehr, N., Hall, M., & Frank, E. (2005). Logistic Model Trees. *Mach. Learn.* 59, 161–205. <https://doi.org/10.1007/s10994-005-0466>.
- Mirchooli, F., Gholami, I. and Boroughani, M., 2023. Flood Susceptibility Mapping of the Famnat Watershed, Gilan Province. *Journal of Water and Soil*, 37(6), 841-853. (In Persian). <https://doi.org/10.22067/jsw.2023.84146.1328>
- Moore, ID., Grayson, R. B., & Ladson, A. R. (1991). Digital terrain modeling: a review of hydrological, geomorphological and biological applications. *Hydrol Pro.* 5:3–30. <https://doi.org/10.1002/hyp.3360050103>.
- Mosavi, A., Ozturk, P., & Chau, K. W. (2018). Flood prediction using machine learning models: literature review. *Water*, 10 (11), 1536. <https://doi.org/10.3390/w10111536>.
- Nabinejad, S., & Schütttrumpf, H. (2023). Flood Risk Management in Arid and Semi-Arid Areas: A Comprehensive Review of Challenges, Needs, and Opportunities. *Water*, 15, 3113. <https://doi.org/10.3390/w15173113>.
- Pavesi, L., Volpi, E., & Fiori, A. (2024). Flood risk assessment through large-scale modeling under uncertainty. *Nat. Hazards Earth Syst. Sci.*, 24, 4507–4522. <https://doi.org/10.5194/nhess-24-4507-2024>.

- Pourhashemi, S., & Aliabadi, K. (2025). Flood Susceptibility Mapping in Ghare Ghom watershed. *Arid Regions Geographic studies*. In press. 10.22034/jargs.2024.478013.1142. **[In Persian]**
- Pradhan, B. (2010). Flood susceptible mapping and risk area estimation using logistic regression, GIS and remote sensing. *J. Spat. Hydrol.* 9, 1–18. <https://scholarsarchive.byu.edu/josh/vol9/iss2/4>.
- Qanavati, R., Selajgheh, A., Pourghasemi, H. R., Khalighi-Sigarudi, Sh., & Keshtkar, H. R. (2024). Evaluation of machine learning methods (SVM, GLM, FDA and RF) in preparing a flood susceptibility map of a part of Khuzestan province. *Soil and Water Modeling and Management*, 10.22098/mmws.2024.15566.1485. **[In Persian]**
- Rahimpour, T., & Rezaei Moghaddam, M.H. (2024). Modeling the Flood Hazard Potential in the Aji Chai basin using Data Mining Algorithms. *E.E.R.* 14 (4), 19-38. <http://magazine.hormozgan.ac.ir/article-1-862-fa.html>. **[In Persian]**
- Rezaei Moghaddam, M.H., Mokhtari, D., & Shafieimehr, M. (2021). Zoning flood risk in the Shahr Chai Watershed in Miyaneh using Vikor model. *Hydrogeomorphology*, 28, 19-37. 20.1001.1.23833254.1400.8.28.2.5. **[In Persian]**
- Rezaei-Sadr, H., Eslamian, S. (2022). *Arid Zone Flooding* (Vol. 1). Flood Handbook.
- Sakizadeh, M., & Mirzaei, R. A. (2016). Comparative Study of Performance of K-Nearest Neighbors and Support Vector Machines for Classification of Groundwater. *J. Min. Environ.* 7, 149–164. <https://doi.org/10.22044/jme.2016.480>.
- Sami, N.A., & Ibrahim, D.S. (2021). Forecasting Multiphase Flowing Bottom-Hole Pressure of Vertical Oil Wells Using Three Machine Learning Techniques. *Pet. Res.* 6, 417–422. <https://doi.org/10.1016/j.ptlrs.2021.05.004>.
- Shahabi, H. (2021). Flood susceptibility mapping in northern regions of Iran using advanced data mining algorithms (Case study: Haraz watershed), *Journal of Regional Planning*, 11(41), Marvdasht, 165- 184. DOI: 10.30495/jzpm.2021.4246.
- Sharma, V.K., Mishra, N., & Shukla, A.K. (2017). Satellite data planning for flood mapping activities based on high rainfall events generated using TRMM, GEFS and disaster news. *Ann GIS* 23:131–140. <https://doi.org/10.1080/19475683.2017.1304449>.
- Shirzadi, A., Soliamani, K., Habibnejhad, M., Kavian, A., Chapi, K., Shahabi, H., Chen, W., Khosravi, K., Thai Pham, B., Pradhan, B., Ahmad, A., Bin Ahmad, B., & Tien Bui, D. (2018). Novel GIS based machine learning algorithms for shallow landslide susceptibility mapping. *Sensors*, 18, 3777. <https://doi.org/10.3390/s18113777>.
- Shirzadi, A.A., Soleimani, K., Habibnejad Roshanbaha, M., Kaviyan, A.A., & Chapi, K. (2017). A Novel Ensemble Algorithm Based Model for Shallow Landslide Susceptibility Assessment Around the Bijar City. *Geography and Development*, 43, 225-246. 10.22111/GDIJ.2017.3076
- Soltani, M. (2002). Land evaluation to suitable areas selection of the flood spreading in the GIS. M.Sc. thesis, Department of Civil Engineering. University of K.N. Toosi, Tehran, Iran. **[In Persian]**
- Su, J., Lü, H., Zhu, Y., Cui, Y., & Wang, X. (2019). Evaluating the hydrological utility of latest IMERG products over the Upper Huaihe River Basin, China. *Atmos. Res.* 225, 17–29. <https://doi.org/10.1016/j.atmosres.2019.03.025>.
- Tien Bui, D., Khosravi, K., Shahabi, H., Daggupati, P., Adamowski, J.F., Melesse, M., Thai Pham, B., Pourghasemi, H.R., Mahmoudi, M., & Bahrani, S. (2019)^a. Flood spatial modeling in northern Iran using remote sensing and gis: a comparison between evidential belief functions and its ensemble with a multivariate logistic regression model. *Remote Sens.* 11(13), 1589. <https://doi.org/10.3390/rs11131589>.
- Tien Bui, D., Ngo, P. T.T., Pham, T.D., Jaafari, A., Minh, N.Q., Hoa, P.V., & Samui, P. (2019)^b. A novel hybrid approach based on a swarm intelligence optimized extreme learning machine for flash flood susceptibility mapping. *Catena*, 179, 184–196. <https://doi.org/10.1016/j.catena.2019.04.009>.
- Tien Bui, D., Tuan, T.A., Klempe, H., Pradhan, B., & Revhaug, I. (2015). Spatial Prediction Models for Shallow Landslide Hazards: A Comparative Assessment of the Efficacy of Support Vector Machines, Artificial Neural Networks, Kernel Logistic Regression, and Logistic Model Tree. *Landslides*, 13, 361–378. DOI:10.1007/s10346-015-0557-6.
- Towfiqul Islam, A.R.M., Talukdar, S., Mahato, S., Kundu, S., Uddin Eibek, K., Bao Pham, Q., Kuriqi, A., & Thi Thuy Linh, N. (2021). Flood susceptibility modelling using advanced ensemble machine learning models. *Geoscience Frontiers*, 12, 101075. <https://doi.org/10.1016/j.gsf.2020.09.006>.
- Usman, A.G., I,sik, S., & Abba, S.I. (2021). Hybrid Data-Intelligence Algorithms for the Simulation of Thymoquinone in HPLC Method Development. *J. Iran. Chem. Soc.* 18, 1537–1549. DOI:10.1007/s13738-020-02124-5.

- Venegas-Quiñones, H.L., García-Chevesich, P., Valdés-Pineda, R., Ferré, T.P.A., Gupta, H., Groenendyk, D., Valdés, J.B., McCray, J.E., & Bakkensen, L. (2024). Creating Sustainable Flood Maps Using Machine Learning and Free Remote Sensing Data in Unmapped Areas. *Sustainability*, 16, 8918. <https://doi.org/10.3390/su16208918>.
- Yamani, M., & Enayati, M. (2005). The relationship between geomorphological characteristics of basins and flood potential (Analysis of flood data through geomorphological comparison of Fashand and Bahjatabad basins. *Geographical Research*, 54, 47-57. **[In Persian]**
- Yaseen, Z.M., Deo, R.C., Hilal, A., Abd, A.M., Bueno, L.C., Salcedo-Sanz, S., & Nehdi, M.L. (2018). Predicting Compressive Strength of Lightweight Foamed Concrete Using Extreme Learning Machine Model. *Adv. Eng. Softw.* 115, 112–125. <https://doi.org/10.1016/j.advengsoft.2017.09.004>.
- Zangeneh Asadi, M. A., Amirahmadi, A., & Naemitabar, M. (2021). Evaluation of the efficiency of VIKOR, L-THIA and artificial neural network models in regional flood analysis (case study: Khorasan Razavi Province). *Ecohydrology*, 8(1), 89-108. 10.22059/IJE.2021.305168.1346. **[In Persian]**

