

Evaluation of machine learning techniques (SVM, GLM, FDA, RF) in preparing flood susceptibility map of a part of Khuzestan province

Royat Ghanavati ¹, Ali Salajegheh ^{2*}, Hamid Reza Pourghasemi ³, Shahram Khalighi Sigaroodi ⁴,
Hamidreza Keshtkar ⁵

¹ Ph.D. Student, Department of Reclamation of Arid and Mountainous Regions, Faculty of Natural resources, University of Tehran, Tehran, Iran

² Professor, Department of Reclamation of Arid and Mountainous Regions, Faculty of Natural resources, University of Tehran, Tehran, Iran

³ Professor, Department of Soil Science, College of Agriculture, Shiraz University, Shiraz, Iran

⁴ Associate Professor, Department of Reclamation of Arid and Mountainous Regions, Faculty of Natural resources, University of Tehran, Tehran, Iran

⁵ Assistant Professor, Department of Reclamation of Arid and Mountainous Regions, Faculty of Natural resources, University of Tehran, Tehran, Iran

Extended Abstract

Introduction

Developing countries are particularly vulnerable to floods due to inadequate infrastructure, limited financial resources, and lack of advanced technology for mitigating flood impacts. Therefore, there is a critical need to develop high-performance flood forecasting models to delineate flood-sensitive areas. The frequency, lethality, and economic impact of floods have spurred the scientific community to create sophisticated algorithms and models to manage the inherent complexity of these natural events. Data mining algorithms have revolutionized scientific research by extracting patterns from vast, unstructured datasets and predicting future trends in complex natural phenomena. Machine learning techniques, a vital subset of data mining methods, excel in making accurate predictions by addressing data limitations and preventing overfitting with proper configuration. Previous studies have demonstrated that machine learning algorithms significantly improve the speed and accuracy of mapping potential flood risks. Consequently, this study aims to develop a flood susceptibility map for a region in Khuzestan province using advanced machine learning algorithms. This region has experienced frequent floods, leading to substantial human and financial losses. Notably, during the floods of 2018, villages near the Dez and Karkheh dams encountered severe challenges.

Materials and Methods

The preparation of the flood risk map is based on two key hypotheses: (1) the past is indicative of the future, implying that future hazards will occur under conditions similar to those of past events, and (2) flood risk conditioning factors are spatially related and can be utilized in forecasting models. To test these hypotheses, the locations of past floods were obtained from relevant authorities and verified through field visits. These locations were randomly divided into two groups: a training group (70%) and a validation group (30%). Data on flood risk conditioning factors, including topography, hydroclimatic conditions, and geological information, were collected and used to create raster maps of these predictive factors. The locations of flood points were treated as dependent variables. Machine learning algorithms, specifically Support Vector Machine (SVM), Generalized Linear Model (GLM), Flexible Discriminant Analysis (FDA), and Random Forest (RF), were applied to generate the flood risk map. The performance of the models was assessed using the area under the receiver operating characteristic curve (ROC) with the validation group data (30% of the flood points), and the best-performing model was selected. The final flood risk map was then produced based on this optimal model.

Results and Discussion

According to the collinearity analysis of the 13 factors influencing floods, all factors had tolerance thresholds greater than 0.1 and variance inflation factors less than 5. Therefore, collinearity was not an issue, and no factors needed to be removed. Flood susceptibility modeling was conducted using four models: SVM, GLM, FDA, and RF. The resulting flood hazard maps from these models were classified into five risk categories: very low, low, medium, high,

and very high. The results indicated that all four models identified flat lands and surface runoff margins as areas with higher flood susceptibility. In all models, more than half of the study area was classified as having low and very low flood risk. Specifically, the SVM, GLM, FDA, and RF models identified 73.9%, 69%, 72.6%, and 63.9% of the area, respectively, as low and very low risk, with the remainder falling into medium to very high risk categories. Additionally, the RF and GLM models indicated a larger portion of the region was at high to very high risk, with 4.7% and 3.9% of the area classified as high risk, respectively. Among the four models, the RF model demonstrated the highest performance, with an area under the curve (AUC) value of 98.8%.

Conclusion

Predicting high-risk areas is crucial for guiding decisions and implementing preventive measures. This study evaluated the performance of four machine learning models—SVM, GLM, FDA, and RF—in preparing a flood hazard map for a part of Khuzestan province, using the area under the ROC curve as the evaluation metric. The results revealed that the RF model achieved the highest accuracy, with an area under the curve of 98.8%, and was identified as the most suitable model for predicting flood risk areas. According to this model, the areas classified as very low, low, medium, high, and very high risk accounted for 34.2%, 29.7%, 18.9%, 12.4%, and 4.7% of the region, respectively. Additionally, the GLM and FDA models demonstrated acceptable accuracy, with AUC values of 76.3% and 75.2%, respectively. These results underscore the efficacy of machine learning models in predicting flood risk areas. Given the increasing population, urban development, and infrastructure expansion in mountainous areas and floodplains, it is essential to develop various hazard susceptibility maps and multi-hazard maps for sustainable development. Future research should focus on evaluating different machine learning models and creating hazard maps for other potential hazards in the region, ultimately leading to the development of comprehensive multi-hazard maps. The findings of this research will assist decision-makers and policymakers in making informed planning decisions for both current and future land use and infrastructure development.

Keywords: Data mining Models, Flood hazard, Random forest model, Khuzestan province.

Article Type: Research Article

Acknowledgement

We would like to express our sincere gratitude to the University of Tehran and the University of Shiraz for the financial and logistical support, which significantly contributed to this research project.

Conflicts of interest

The authors declare no conflicts of interest.

Data availability statement

The data presented in this study are available upon corroborated request from the corresponding author. The meteorological data used in this study are confidential and not publicly available.

Authors' contribution

Royat Ghanavati: Writing - original draft preparation, manuscript editing, visualization; **Ali Salajegheh:** Resources, control of the results, supervision; **Hamid Reza Pourghasemi:** Performing software/statistical analysis, methodology, visualization; **Shahram Khalighi Sigaroodi:** Conceptualization, control of the results; **Hamidreza Keshtkar:** manuscript review and editing, control of the results.

*Corresponding Author, E-mail: salajegh@ut.ac.ir

Citation: Ghanavati, R., Salajegheh, A., Pourghasemi, H. R., Khalighi Sigaroodi, Sh., & Keshtkar, H. R. (2025). Evaluation of machine learning techniques (SVM, GLM, FDA, RF) in preparing flood susceptibility map of a part of Khuzestan province. *Water and Soil Management and Modeling*, 5(1), 231-246.
doi:10.22098/mmws.2024.15566.1485

Received: 04 August 2024, Received in revised form: 31 August 2024, Accepted: 31 August 2024, Published online: 21 March 2025.

Water and Soil Management and Modeling, Year 2025, Vol. 5, No. 1, pp. 231-246

Publisher: University of Mohaghegh Ardabili

© Author(s)





ارزیابی روش‌های یادگیری ماشین (RF، FDA، GLM، SVM) در تهیه نقشه حساسیت سیل بخشی از استان خوزستان

رویت قنوتی^۱، علی سلاجقه^{۲*}، حمیدرضا پورقاسمی^۳، شهرام خلیقی سیگارودی^۴، حمیدرضا کشت کار^۵

^۱ دانشجوی دکتری، گروه احیا مناطق خشک و کوهستانی، دانشکده منابع طبیعی، دانشگاه تهران، تهران، ایران

^۲ استاد، گروه احیاء مناطق خشک و کوهستانی، دانشکده منابع طبیعی، دانشگاه تهران، تهران، ایران

^۳ استاد، بخش علوم خاک، دانشکده کشاورزی، دانشگاه شیراز، شیراز، ایران

^۴ دانشیار، گروه احیا مناطق خشک و کوهستانی، دانشکده منابع طبیعی، دانشگاه تهران، تهران، ایران

^۵ استادیار، گروه احیا مناطق خشک و کوهستانی، دانشکده منابع طبیعی، دانشگاه تهران، تهران، ایران

چکیده

سیل یکی از مخرب‌ترین بلایای طبیعی است که خسارات جدی به منابع طبیعی و زیرساخت‌ها وارد کرده و تلفات انسانی بسیاری به همراه دارد. مدل‌های یادگیری ماشین به منظور شناسایی و مدیریت مناطق در معرض خطر سیل به‌طور گسترده‌ای مورد توجه بوده است. هدف از این تحقیق ارزیابی عملکرد چهار مدل ماشین بردار پشتیبان (SVM)، خطی تعمیم‌یافته (GLM)، تحلیل تفکیکی انعطاف‌پذیر (FDA) و جنگل تصادفی (RF) در مدل‌سازی پراکنش خطر وقوع سیل بخشی از استان خوزستان بود. برای این منظور ۱۳ عامل مؤثر بر سیل شامل عوامل توپوگرافی، هیدرواقليمی، سنگ‌شناسی و انسانی تعیین شد. سپس موقعیت ۳۳۴ نقطه محل وقوع و عدم وقوع سیلاب براساس بازدهی میدانی و گزارش‌های موجود مشخص شد؛ که ۷۰ درصد از این نقاط برای آموزش و ۳۰ درصد باقیمانده جهت اعتبارسنجی مدل‌ها، به صورت تصادفی در نظر گرفته شد. نتایج ارزیابی عملکرد مدل‌های مورد بررسی براساس شاخص مساحت زیر منحنی مشخصه عامل گیرنده (ROC) برای مدل‌های RF، GLM و FDA بالاتر از ۰/۷۵ به دست آمد؛ که مدل RF با سطح زیر منحنی ۹۸/۸ درصد از صحت بالاتری نسبت به سایر مدل‌ها برخوردار بود. براساس نقشه حساسیت خطر سیل حاصل از این مدل به ترتیب در ۴/۷ درصد و ۱۲/۴ درصد از سطح منطقه احتمال وقوع سیل خیلی زیاد و زیاد است. نتایج این تحقیق به مدیران در کاهش تهدیدهای مرتبط با سیل و اجرای راهکارهای مدیریتی مؤثر در جهت کاهش خسارات آن کمک می‌کند.

واژه‌های کلیدی: استان خوزستان، خطر سیل، مدل‌های داده‌کاوی، مدل جنگل تصادفی.

نوع مقاله: پژوهشی

*مسئول مکاتبات، پست الکترونیکی: salajegh@ut.ac.ir

استناد: قنوتی، رویت، سلاجقه، علی، پورقاسمی، حمیدرضا، خلیقی سیگارودی، شهرام، کشت کار، حمیدرضا (۱۴۰۴). ارزیابی روش‌های یادگیری ماشین (RF، FDA، GLM، SVM) در تهیه نقشه حساسیت سیل بخشی از استان خوزستان. *مدل‌سازی و مدیریت آب و خاک*، ۵(۱)، ۲۳۱-۲۴۶
doi:10.22098/mmws.2024.15566.1485



تاریخ دریافت: ۱۴۰۳/۰۵/۱۴، تاریخ بازنگری: ۱۴۰۳/۰۶/۱۰، تاریخ پذیرش: ۱۴۰۳/۰۶/۱۰، تاریخ انتشار: ۱۴۰۴/۰۱/۰۱
مدل‌سازی و مدیریت آب و خاک، سال ۱۴۰۴، دوره ۵، شماره ۱، صفحه ۲۳۱ تا ۲۴۶
ناشر: دانشگاه محقق اردبیلی © نویسندگان

۱- مقدمه

سیل یکی از فاجعه‌بارترین بلاهای طبیعی است که خرابی و خسارات انسانی گسترده‌ای را در سراسر جهان به همراه دارد (Saharia et al., 2021). این مخاطره ۴۴ درصد از بلایای جهان را به خود اختصاص داده که ۴۱ درصد از سیلاب‌های جهان در آسیا رخ می‌دهد و از سال ۲۰۰۰ تا ۲۰۱۹ حدود ۱/۵ میلیارد نفر را تحت تأثیر قرار داده است (Gupta and Dixit, 2022). از طرفی کشورهای در حال توسعه به دلیل نداشتن زیرساخت‌های مناسب، امکانات مالی و فناوری پیشرفته (AIQahtany and Abubakar, 2020) به‌منظور کاهش پیامدهای سیل، در مقایسه با کشورهای توسعه‌یافته از آسیب‌پذیری بالاتری برخوردار می‌باشند. بنابراین، ایجاد یک مدل پیش‌بینی سیل با کارایی بالا برای تهیه نقشه مناطق حساس به سیل دارای اهمیت بالایی است (Towfiqul Islam et al., 2021). شناسایی مناطق مستعد سیل به‌عنوان یک رویکرد حیاتی برای ارزیابی آسیب‌پذیری در نظر گرفته می‌شود و برای جلوگیری از وقوع سیلاب در آینده بسیار حائز اهمیت است (Vojtek and Vojteková, 2019). همچنین، سیاست‌گذاران و مسئولین جهت تهیه و تدوین برنامه‌های مدیریت بحران ضروری است (Rahman et al., 2019). اگرچه وقوع سیلاب ناشی از آب و هوا را نمی‌توان متوقف کرد، اما با شناسایی صحیح مناطقی که دارای پتانسیل وقوع سیلاب هستند، می‌توان تلفات و آسیب‌ها را تا حد امکان کنترل و به حداقل رساند (Sahoo and Sreeja, 2017).

مناطق خشک و نیمه‌خشک انواع مختلفی از خطرات و با شدت‌های متنوع، وقایع کم اثر تا شدید، تأثیر می‌پذیرند (Rutgersson et al., 2022). به‌طور مثال سیل‌های گسترده بخش‌های شمالی، جنوبی و جنوب‌غربی ایران در فروردین ماه سال ۱۳۹۸، تقریباً ۲۶ استان از ۳۱ استان را درگیر و دست کم ۷۰ کشته در سراسر کشور برجای گذاشت. به‌طوری‌که استان‌های شمالی (گلستان و مازندران) ۷۰ درصد از میانگین بارش سالانه را در یک روز تجربه کردند (Bozorgmehr, 2019). همچنین، استان‌های فارس، خوزستان و لرستان از دیگر استان‌های آسیب‌دیده بودند.

از آن‌جایی که سیلاب یکی از رایج‌ترین، مرگبارترین و پرهزینه‌ترین مخاطرات طبیعی است، جامعه علمی را بر آن داشته تا مجموعه‌ای از الگوریتم‌ها و مدل‌های مختلف را به‌منظور مدیریت سطح بالای پیچیدگی پدیده‌های طبیعی دنبال کنند (Satarzadeh et al., 2022). الگوریتم‌های داده‌کاوی با ارائه یک استراتژی کارآمد برای استخراج الگوی داده‌های بزرگ

سازمان‌نیافته و پیش‌بینی روند و الگوهای پیچیده طبیعی برای آینده، افق جدیدی را به روی همه شاخه‌های علمی باز کرده‌اند (Goodfellow et al., 2016). تکنیک‌های یادگیری ماشین به‌عنوان زیرمجموعه‌ای از روش‌های داده‌کاوی، می‌توانند با مقابله با محدودیت داده و امکان شناسایی الگو و روابطی که ممکن است برای انسان آشکار نباشد، یک پیش‌بینی کارآمد ارائه کنند (Yu et al., 2024). بررسی مطالعات انجام شده نشان می‌دهد که مدل‌های یادگیری ماشین بخصوص در زمینه پیش‌بینی مکانی مخاطرات گوناگون بسیار مورد توجه بوده‌اند. به‌عنوان مثال، بسیاری از مطالعات ارزیابی حساسیت زمین‌لغزش (Segue et al., 2024)، پیش‌بینی مکانی فرسایش خاک (Youssef et al., 2023)، ارائه نقشه حساسیت فرونشست (Wen et al., 2022)، سیل (Pourghasemi et al., 2023) و آتش‌سوزی (Pourghasemi et al., 2020) با به‌کارگیری این الگوریتم‌ها انجام گرفته است. (Cheraghi Ghalehsari et al., 2020). شناسایی مناطق حساس به سیل با استفاده از ماشین بردار پشتیبان (SVM) در حوزه آبخیز نکارود پرداختند. برای این منظور ۱۲ پارامتر ژئومورفولوژیکی، هیدرولوژیکی و فیزیوگرافی در محیط‌های ArcGIS، SAGA GIS و ENVI تهیه شدند. کارایی مدل نیز براساس سطح زیر منحنی ROC بررسی شد. نتایج ارزیابی نشان داد که مدل SVM دقت مناسبی در شناسایی پهنه‌های حساس سیلاب در منطقه مورد مطالعه دارد. (Vafakhah et al., 2020). عملکرد مدل‌های FR^1 ، $ANFIS^2$ و RF^3 برای تهیه نقشه حساسیت سیل در استان گیلان مورد بررسی قرار دادند. در این مطالعه از ۲۲۰ نقطه محل سیل و ۱۱ عامل مؤثر استفاده شد. در نهایت منحنی مشخصه عملکرد سیستم (ROC^4) و سطح زیر منحنی (AUC^5) برای ارزیابی و مقایسه عملکرد مدل‌ها به‌کار گرفته شد. نتایج اعتبارسنجی مدل‌ها نشان داد که مدل‌های FR ، $ANFIS$ و RF به‌ترتیب ۶۰/۶۰، ۶۳/۹۰ و ۷۱/۳۰ درصد دقت داشت و بالاترین دقت برای مدل جنگل تصادفی به‌دست آمد. (Rahman et al., 2021). نقشه خطر سیل را برای منطقه‌ای آسیب‌پذیر به سیلاب در شمال بنگلادش تهیه کردند. در ابتدا نقشه خطر سیل با استفاده از مدل هیدرودینامیکی (FLO 2D) همراه با الگوریتم یادگیری ماشین (SCG-NN) تهیه شد. عملکرد الگوریتم یادگیری ماشین با استفاده از مجموعه داده‌های اعتبارسنجی و شاخص‌های آماری مانند خطای میانگین مربعات ($MSE: 0/08$)، ریشه خطای میانگین مربعات ($RMSE: 0/28$) و ضریب تعیین ($R^2: 0/83$) انجام شد. طبق نقشه سیل تولید شده، اغلب سطح

⁴ Receiver operating characteristic

⁵ Area under the curve

¹ Frequency Ratio

² Adaptive neuro-fuzzy inference system

³ Random forest

به طوری که در سیلاب‌های سال ۱۳۹۸ روستاهای حاشیه سدهای دز و کرخه با مشکلات زیادی مواجه شدند (Rajabizadeh et al., 2020). نتایج تحقیق حاضر می‌تواند توسط مدیران و برنامه‌ریزان مربوطه مورد استفاده قرار گیرد.

۲- مواد و روش‌ها

۲-۱- منطقه مورد مطالعه

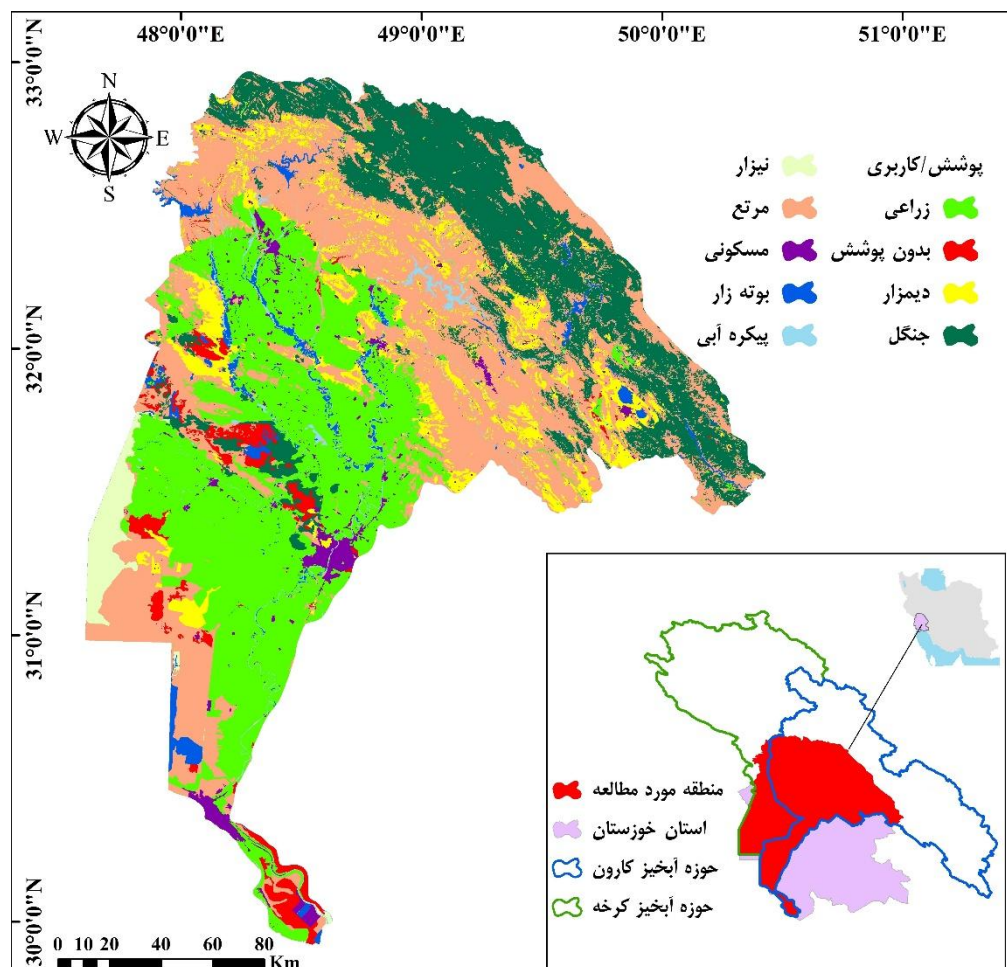
استان خوزستان واقع در جنوب غرب ایران با مساحت تقریبی ۶۴۰۰۰ کیلومترمربع، چهاردرصد از سطح کل کشور را در بر می‌گیرد. با توجه به وسعت این استان، در این تحقیق بخشی از آن که در حوزه آبخیز کارون بزرگ و کرخه قرار گرفته است، مورد بررسی قرار گرفت. شکل ۱ موقعیت منطقه مورد مطالعه را در استان و حوزه‌های مذکور نشان می‌دهد. این منطقه با مساحتی حدود ۳۷۴۷۶ کیلومترمربع بین عرض‌های شمالی ۲۹ درجه و ۵۳ دقیقه تا ۳۳ درجه و طول‌های شرقی ۴۷ درجه و ۳۹ دقیقه تا ۵۰ درجه و ۲۶ دقیقه واقع شده است؛ که از بخش کوهستانی منطقه شمال و شمال شرق، تا دشت پست و هموار جنوب غرب این استان را در بر می‌گیرد. میانگین بارش منطقه مورد مطالعه حدود ۲۸۰ میلی‌متر و میانگین دمای بیشینه و کمینه به ترتیب ۳۳ و ۱۸/۶۰ درجه سانتی‌گراد است و در برخی روزهای تابستان دما در دشت‌های پست به ۵۰ درجه سانتی‌گراد نیز می‌رسد. (Masoudi and Elhaesahar, 2016). نقشه کاربری/پوشش سال ۱۳۹۹ با مقیاس ۱:۵۰۰۰۰ از سازمان منابع طبیعی و آبخیزداری تهیه شد. بر این اساس پوشش غالب منطقه مورد مطالعه را مراتع (۳۳ درصد)، اراضی کشاورزی (۳۰ درصد) و جنگل (۱۹ درصد) تشکیل داده‌اند و درصد کمی از آن شامل دیمزار (۸ درصد) اراضی بدون پوشش (۳ درصد)، نیزار (۲ درصد)، مسکونی (۲ درصد)، پوشش آبی (۲ درصد) و بوته‌زار (۱ درصد) است.

منطقه مطالعاتی متعلق به طبقه کم (۴۷/۸۵ درصد) و متوسط (۲۷/۴۷ درصد) بود و بخش کمی از منطقه در طبقه بالا (۲۰/۶۴ درصد) و بسیار بالا (۴/۰۴ درصد) قرار گرفت.

در مطالعه‌ای صورت گرفته توسط Hitouri et al. (2024) استفاده از محصولات سنجش از دور با کیفیت بالا و رادار دیگرام مصنوعی (SAR) نقاط سیل را مشخص کردند، سپس با به‌کارگیری چهار مدل یادگیری ماشین (RF، CART، SVM و XGBOOST) نقشه حساسیت سیل حوزه آبخیز Metlili واقع در مراکش را تهیه کردند. ایشان از ۱۲ متغیر مستقل به‌عنوان عوامل شرطی‌سازی استفاده نمودند و ۷۰ درصد نقاط سیل برای آموزش و ۳۰ درصد برای اعتبارسنجی در نظر گرفته شد. نتایج ارزیابی مدل‌ها براساس ناحیه زیر منحنی (AUC) نشان داد که مدل‌های RF¹، CART¹، SVM و XGBOOST² حساسیت سیل را به ترتیب با دقت ۰/۸۱، ۰/۷۸، ۰/۷۵ و ۰/۷۳ پیش‌بینی نمودند. براین اساس مدل RF بهترین عملکرد را داشت و طبق نتایج این مدل به ترتیب ۱۸/۱۰ و ۳۱/۷۰ درصد از سطح حوزه مطالعاتی در طبقات با خطر بالا و خیلی بالا قرار داشت. این مطالعه نشان داد که ادغام مدل‌های یادگیری ماشین با داده‌های راداری می‌تواند نتایج امیدوارکننده‌ای در پیش‌بینی حساسیت سیل منطقه مورد مطالعه و سایر محیط‌های مشابه داشته باشد. نتایج تحقیقات پیشین نشان داد که الگوریتم‌های یادگیری ماشین سرعت تحلیل و دقت بالاتری در تحلیل و تهیه نقشه پتانسیل خطر دارند. بنابراین هدف این تحقیق ارزیابی عملکرد برخی مدل‌های پرکاربرد یادگیری ماشین در زمینه مدل‌سازی مخاطرات طبیعی، از جمله: ماشین بردار پشتیبان (SVM)، خطی تعمیم یافته (GLM)، تحلیل تفکیکی انعطاف‌پذیر (FDA) و جنگل تصادفی (RF) به‌منظور ارائه نقشه حساسیت سیل بخشی از استان خوزستان بود. این منطقه در معرض وقوع مکرر سیل و خسارات جانی و مالی فراوان بوده است؛

² Extrem gradient boosting

¹ Classification and regression trees



شکل ۱- پوشش/کاربری اراضی و موقعیت منطقه مورد مطالعه در استان خوزستان، حوزه‌های کرخه و کارون، ایران
Figure 1- Land Use/Land Cover (LULC) map and the location of study area in Khuzestan province, Karkheh and Karoon Basins, Iran

تهیه شد. همچنین، موقعیت نقاط خطر به‌عنوان عوامل وابسته در نظر گرفته شد. برای تولید نقشه خطر سیل الگوریتم‌های یادگیری ماشین^۱ (SVM^۱، GLM^۲، FDA^۳ و RF^۴) اعمال و ارزیابی کارایی آن‌ها با استفاده از داده‌های گروه اعتبارسنجی (۳۰ درصد نقاط خطر) صورت گرفت و بهترین مدل تعیین شد. در نهایت نقشه پراکنش مکانی خطر وقوع سیل براساس بهترین مدل ارائه شد.

۲-۲-۱- تهیه نقشه پراکنش نقاط وقوع سیل
موقعیت نقاط وقوع سیلاب در منطقه از سازمان منابع طبیعی و آبخیزداری تهیه شد. از آنجایی که این مرحله یک مرحله کلیدی و مهم در تهیه نقشه حساسیت خطر است (Conoscenti et al., 2013)، ارزیابی دقت موقعیت برخی از این نقاط و برداشت نقاط جدید خطر از طریق بازدیدهای میدانی نیز انجام شد. شکل ۲ نقشه پراکنش نقاط وقوع سیل را نشان می‌دهد، این نقشه شامل ۳۳۴

۲-۲- روش تحقیق

تهیه نقشه خطر سیل براساس دو فرضیه انجام می‌شود، یک: گذشته کلید آینده است، به این معنی که مخاطرات آتی تحت شرایط مشابه با آنچه در گذشته رخ داده است، اتفاق می‌افتند. و دو: عوامل مؤثر بر وقوع خطر از نظر مکانی بهم مرتبط بوده و می‌توانند در عملکردهای پیش‌بینی استفاده شوند (Reichenbach et al., 2018). بنابراین محل وقوع سیلاب‌های گذشته از سازمان منابع طبیعی و آبخیزداری کشور و براساس بازدیدهای میدانی تعیین و به‌صورت تصادفی به دو دسته گروه آموزش (۷۰ درصد) و گروه اعتبارسنجی (۳۰ درصد) تقسیم شد. سپس اطلاعات مربوط به عوامل مؤثر بر وقوع سیل از جمله اطلاعات توپوگرافی، هیدرواقليمی، زمین‌شناسی و غیره جمع‌آوری و نقشه رستری پارامترهای مربوط آن‌ها به‌عنوان عوامل پیش‌بینی کننده (مستقل)

³ Functional discriminate analysis

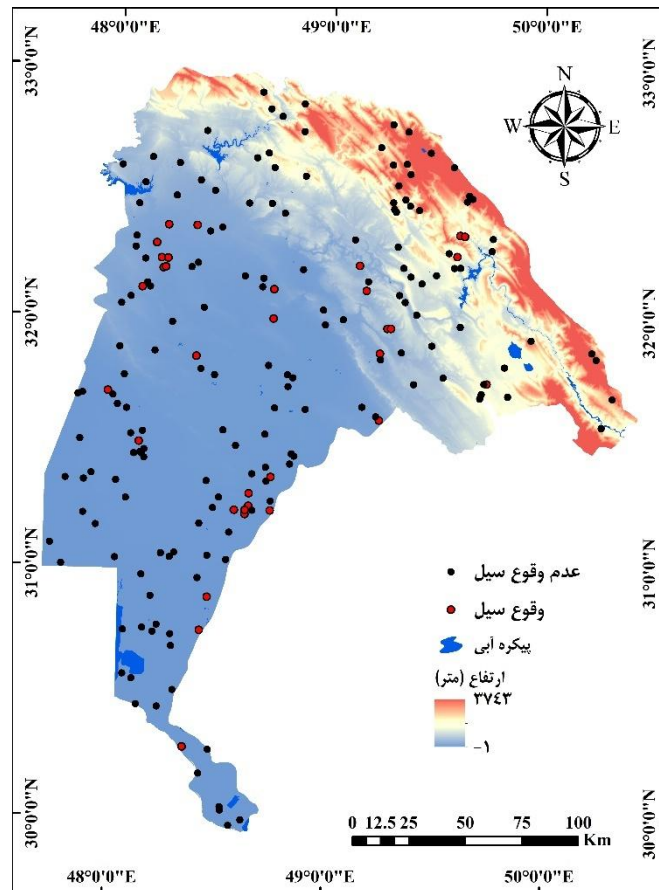
⁴ Random forest

¹ Support vector machin

² Generalized linear model

شده مورد استفاده قرار گرفت (Nachappa et al., 2020). لازم به ذکر است که در برخی موقعیت‌های سیل در شکل ۲، واقعه سیل چندین بار در یک موقعیت اتفاق افتاده است.

موقعیت وقوع و عدم‌وقوع سیل است؛ که ۷۰ درصد از این نقاط به‌صورت تصادفی به‌عنوان متغیر وابسته در فرآیند مدل‌سازی به‌کار گرفته شد و ۳۰ درصد باقی‌مانده برای ارزیابی صحت نقشه تهیه



شکل ۲- پراکنش نقاط وقوع و عدم‌وقوع سیلاب در منطقه مورد مطالعه
Figure 2- Distribution of flood occurrence and non-occurrence points in the study area

فاکتور توسط نرم‌افزارهای ArcGIS و SAGA با دقت ۳۰ متر × ۳۰ متر تهیه شد.

مقادیر بارش که از عوامل تعیین‌کننده وقوع و شدت سیل است (Dehnavi et al., 2015; Segond et al., 2007) براساس داده‌های اندازه‌گیری شده ۱۹ ایستگاه سینوپتیک مربوط به منطقه مورد بررسی از سازمان هواشناسی کشور^۵ اخذ شد. مدل رقومی ارتفاع یا DEM با دقت ۳۰ متر × ۳۰ متر از سایت USGS^۶ و داده‌های SRTM^۷ تهیه شد. بسیاری از عوامل توپوگرافی از قبیل شیب، جهت شیب، شاخص رطوبت توپوگرافی، انحنا طولی و عرضی دامنه براساس این نقشه در GIS تهیه شد. نقشه سنگ‌شناسی با مقیاس ۱:۱۰۰۰۰۰ از سازمان زمین‌شناسی کشور

۲-۲-۲- تهیه نقشه عوامل پیش‌بینی‌کننده سیلاب

نقشه عوامل پیش‌بینی‌کننده از طریق بررسی مطالعات پیشین (Pourghasemi et al., 2023; Pourghasemi and Rahmati, 2018; Zhou et al., 2018) و با در نظر گرفتن امکان دسترسی و به‌کارگیری آن در مدل‌سازی تعیین شد. براین اساس نقشه ۱۳ عامل مؤثر بر سیل شامل بارش، ارتفاع، شیب، جهت شیب، شاخص رطوبت توپوگرافی^۱ (TWI) انحنا طولی^۲، انحنا عرضی^۳، سنگ‌شناسی، تراکم زهکش، فاصله از آبراهه، کاربری/پوشش اراضی، شاخص نرمال شده تفاوت پوشش گیاهی^۴ (NDVI) و فاصله از جاده تهیه شد. لایه‌های رستری مربوط به هر

⁵ www.irimo.ir

⁶ www.earthexplorer.usgs.gov

⁷ Shuttle radar topography mission

¹ Topographic wetness index

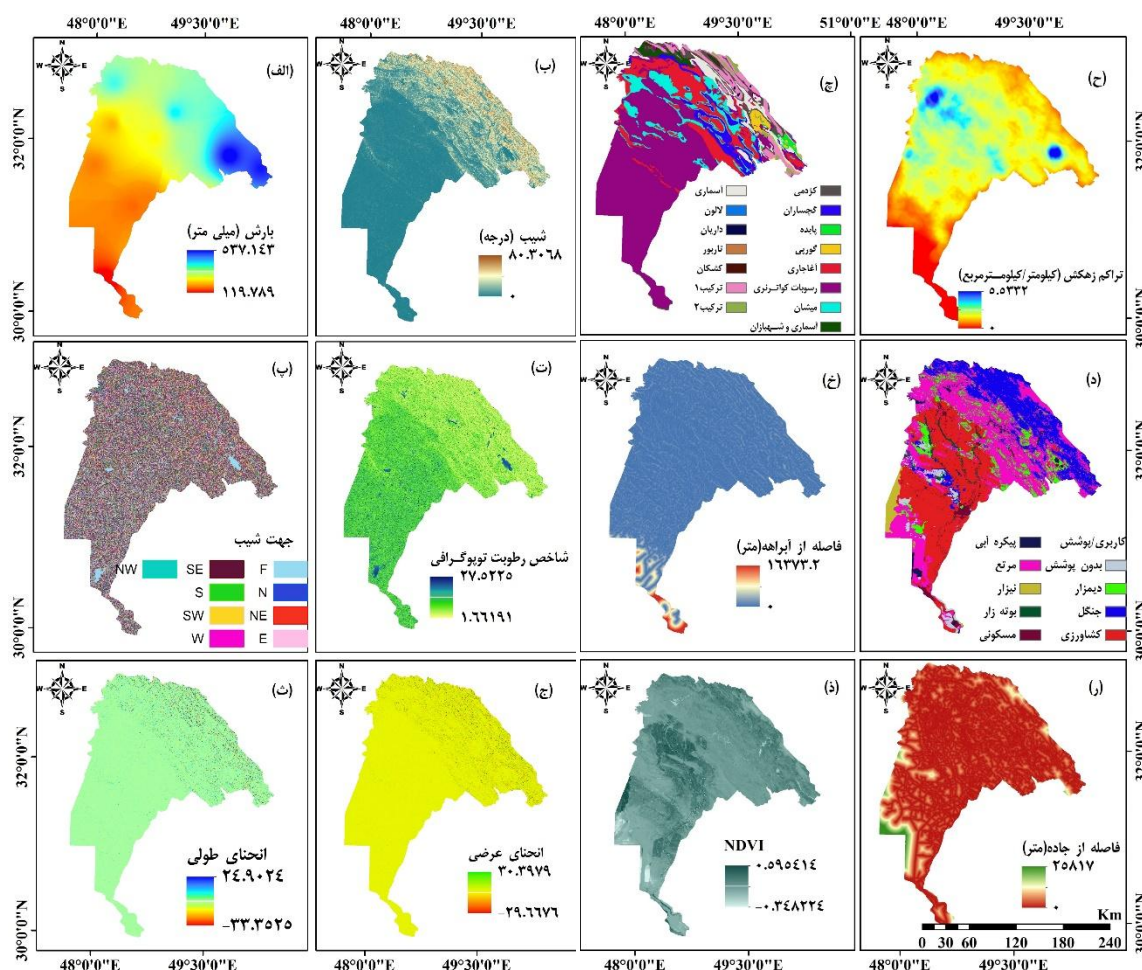
² Plan curvature

³ Profile curvature

⁴ Normalized difference vegetation index

تهیه شد. اطلاعات پوشش/کاربری منطقه مورد مطالعه مربوط به سال ۲۰۲۰ از سازمان منابع طبیعی و آبخیزداری کشور تهیه شد. شاخص NDVI براساس باندهای چهار (قرمز) و پنج (مادون قرمز نزدیک) تصاویر ماهواره‌ای لندست ۸ سنجنده OLI^۱ و در محیط

تهیه شد. اطلاعات پوشش/کاربری منطقه مورد مطالعه مربوط به سال ۲۰۲۰ از سازمان منابع طبیعی و آبخیزداری کشور تهیه شد. شاخص NDVI براساس باندهای چهار (قرمز) و پنج (مادون قرمز نزدیک) تصاویر ماهواره‌ای لندست ۸ سنجنده OLI^۱ و در محیط



شکل ۳- عوامل مؤثر بر وقوع سیل، الف) بارش (میلی‌متر)، ب) شیب (درجه)، پ) جهت شیب، ت) شاخص رطوبت توپوگرافی، ث) انحنا طولی دامنه، ج) انحنا عرضی دامنه، چ) واحدهای سنگی، ح) تراکم زهکش (کیلومتر/کیلومترمربع)، خ) فاصله از آبراهه (متر)، د) کاربری/پوشش اراضی، ذ) شاخص نرمال شده تفاوت پوشش گیاهی (NDVI) و ر) فاصله از جاده (متر).

Figure 3- Flood risk conditioning factors, a) precipitation (mm), b) slope angle, c) aspect, d) topographic humidity index, e) plan curvature, f) profile curvature, g) lithology, h) drainage density (km/km²), i) distance from river (m), j) land use/cover, k) normalized difference of vegetation index (NDVI) and l) distance from road (m).

دامنه مورد انتظار می‌شود، ازین‌رو بررسی همخطی عوامل مؤثر پیش از مدل‌سازی باید انجام شود (Zare Chahouki et.al., 2020). بررسی همخطی عوامل مؤثر به‌منظور شناسایی عوامل ضروری مرتبط با سیلاب براساس معیارهای آستانه تحمل (TOL^۳) و واریانس متورم (VIF^۴) انجام شد. این معیارها را

۲-۲-۳- بررسی همخطی^۲ عوامل مؤثر بر سیلاب عوامل پیش‌بینی‌کننده متعددی بر سیل مؤثر هستند که استفاده از همه عوامل می‌تواند بر عملکرد پیش‌بینی مدل تأثیر بگذارد و باعث کاهش دقت مدل شود. یکی از دلایل افزایش خطا و کاهش کارایی مدل همخطی است که منجر به پیش‌بینی‌های خارج از

³ Tolerance

⁴ Variance inflation factor

¹ Operational land imager

² Multi-collinearity test

می‌شود. از طریق تبدیل لجستیک، تابع پیوند $g(y_i)$ در رابطه (۴) نشان داده می‌شود.

$$P = \frac{\exp(c_0 + c_1X_1 + c_2X_2 + \dots + c_iX_i)}{1 + \exp(c_0 + c_1X_1 + c_2X_2 + \dots + c_iX_i)} \quad (3)$$

$$g(y_i) = c_0 + \sum c_iX_i + \varepsilon_i \quad (4)$$

P احتمال وقوع خطر y ، c_0 مقدار ثابت، c_i ضرایب‌های رگرسیون لجستیک، X_i متغیرهای مستقل و ε_i بیانگر خطای باقیمانده است.

۳-۲-۲-۴-۳ مدل تحلیل تفکیکی انعطاف‌پذیر (FDA)

مدل FDA یک روش طبقه‌بندی آماری غیرخطی است که توسط (Ramsay and Dalzell, 1991) ارائه شد. این مدل با ترکیب مدل‌ها و روش‌های مختلف (الگوریتم رگرسیون، تحلیل تفکیکی و تکنیک‌های طبقه‌بندی) چارچوبی را ارائه می‌نماید. برای دستیابی به انعطاف‌پذیری بالاتر در مرزهای تصمیم‌گیری، از مدل FDA برای مدل‌های غیرپارامتری رگرسیونی استفاده می‌شود. تحلیل تفکیکی خطی، چندین متغیر بهینه را از بین تعداد زیادی متغیر انتخاب و گروه‌های موجود در متغیر پاسخ را تعیین و تفکیک می‌نماید. در صورتی که در یک رابطه معادلات غیرخطی غالب باشند، از سری معادلات غیرخطی به جای معادلات خطی، به‌منظور برازش مقادیر استفاده می‌شود.

۴-۲-۲-۴-۴ مدل جنگل تصادفی (RF)

جنگل تصادفی یکی از روش‌های مدل‌سازی درخت تصمیم است، که شامل انبوهی از درخت‌های طبقه‌بندی و رگرسیونی هستند (Breiman, 2001). نمونه‌گیری تصادفی از عوامل پیش‌بینی کننده به صورت خودکار و تکرار شونده به‌منظور تولید انبوهی از درختان رگرسیونی انجام می‌شود. نتیجه یا برآورد نهایی در مورد ارتباط مخاطرات طبیعی و عوامل محیطی از تلفیق نتایج به دست آمده از درخت‌های مختلف ارائه می‌شود. میانگین‌گیری خطاهای ایجاد شده توسط هر درخت را کاهش می‌دهد، بنابراین نکته اصلی همبستگی پایین بین درختان است (Schoppa et al., 2020). مهم‌ترین ویژگی RF کارایی بالای آن در اندازه‌گیری اهمیت متغیرها و تعیین میزان نقش هر متغیر در پیش‌بینی پاسخ است. در این تحقیق اجرای مدل‌های مذکور توسط بسته sdm در محیط R انجام شد.

۵-۲-۲-۴-۵ ارزیابی عملکرد مدل‌ها

عملکرد مدل‌های مورد استفاده در پیش‌بینی نقاط حساس به سیل با استفاده از تحلیل منحنی مشخصه عملکرد گیرنده (ROC)

می‌توان براساس همبستگی (R^2) یک عامل پیش‌بینی کننده خاص و سایر عوامل مؤثر بر وقوع سیل تعیین کرد، که به ترتیب براساس رابطه‌های (۱) و (۲) محاسبه می‌شوند.

$$TOL = 1 - R^2 \quad (1)$$

$$VIF = \frac{1}{1 - R^2} \quad (2)$$

براساس مطالعات انجام شده $TOL < 0.10$ و $VIF > 5$ نشان‌دهنده مشکل چندخطی بودن است (Rahman et al., 2019).

۴-۲-۲-۴-۶ تهیه نقشه حساسیت خطر سیلاب با استفاده از مدل‌های یادگیری ماشین

تهیه نقشه پتانسیل خطر سیل بخشی از استان خوزستان با استفاده از مدل‌های ماشین بردار پشتیبان (SVM)، مدل خطی تعمیم‌یافته (GLM)، مدل تحلیل تفکیکی انعطاف‌پذیر (FDA) و جنگل تصادفی (RF) انجام شد. نقشه ۱۳ عامل مؤثر بر سیل به عنوان عوامل مستقل و با فرمت ASCII به همراه نقاط وقوع و عدم وقوع سیل به عنوان عامل وابسته، به مدل‌های مذکور وارد و در نرم‌افزار R اجرا شد.

۱-۲-۲-۴-۱ مدل ماشین بردار پشتیبان (SVM)

SVM یک مدل طبقه‌بندی کننده نظارت شده است که به‌طور گسترده در کمینه‌سازی ریسک ساختاری (SRM) برای حل مشکلات پیچیده رگرسیون و طبقه‌بندی استفاده شده است (Chen et al., 2016). زمانی که مجموعه داده‌ها دارای خطاهای تفکیک‌پذیری خطی پیچیده‌تر یا شرایط غیرخطی هستند، به دلیل کارایی آن برای تولید مرزهای منحنی پیچیده، مفیدتر است (Kalantar et al., 2018). SVM به‌طور مؤثر داده‌های دارای روابط غیرخطی را با استفاده از تابع کرنلی پردازش می‌کند (Hong et al., 2017). این مدل سعی می‌کند یک ابر صفحه^۱ جداکننده بهینه برای جداسازی دو کلاس پیدا کند (Kalantar et al., 2018).

۲-۲-۲-۴-۲ مدل خطی تعمیم یافته (GLM)

مدل خطی تعمیم‌یافته یک مدل رگرسیون خطی است که می‌تواند عوامل خاص و زمانی را ترکیب نماید (Goetz et al., 2015). این مدل برای عوامل عددی مناسب بوده و می‌توان برای داده‌های با توزیع غیرنرمال نیز به کار گرفته شود (Kalantar et al., 2020). احتمال وقوع خطر y براساس رابطه (۳) محاسبه

¹ Hyper-Plane

ارزیابی شد. محورهای منحنی ROC براساس معیارهای حساسیت و تشخیص به‌دست می‌آید که در نظر گرفتن این دو معیار به‌صورت هم‌زمان منجر به رفع نواقص هر یک از معیارهای حساسیت و تشخیص می‌شود (Rossi and Reichenbach, 2016). مساحت زیر منحنی (AUC) به‌عنوان معیار کمی برای اعتبارسنجی مورد استفاده قرار می‌گیرد. این مقدار بین ۰/۵ تا ۱ متغیر است که مقادیر بالاتر آن دقت بالاتر مدل‌سازی را نشان می‌دهد (Yalcin, 2008).

۳- نتایج و بحث

۳-۱- نتایج بررسی هم‌خطی بین عوامل مؤثر بر سیل

با توجه به آنچه گفته شد وجود هم‌خطی بین لایه‌های پیش‌بینی‌کننده منجر به کاهش کارایی و دقت مدل می‌شود. بنابراین لازم است لایه‌های دارای همبستگی شناسایی و حذف گردند. نتایج بررسی هم‌خطی بین عوامل مؤثر بر سیل در جدول ۱ ارائه شده است. بررسی نتایج جدول نشان داد معیارهای آستانه تحمل و تورم واریانس برای همه عوامل به‌ترتیب بیش‌تر از ۰/۱ و کم‌تر از ۵ به‌دست آمد، بنابراین شرط هم‌خطی برقرار نبوده و نیاز به حذف هیچ‌یک از ۱۳ عامل مؤثر بر سیل نبود.

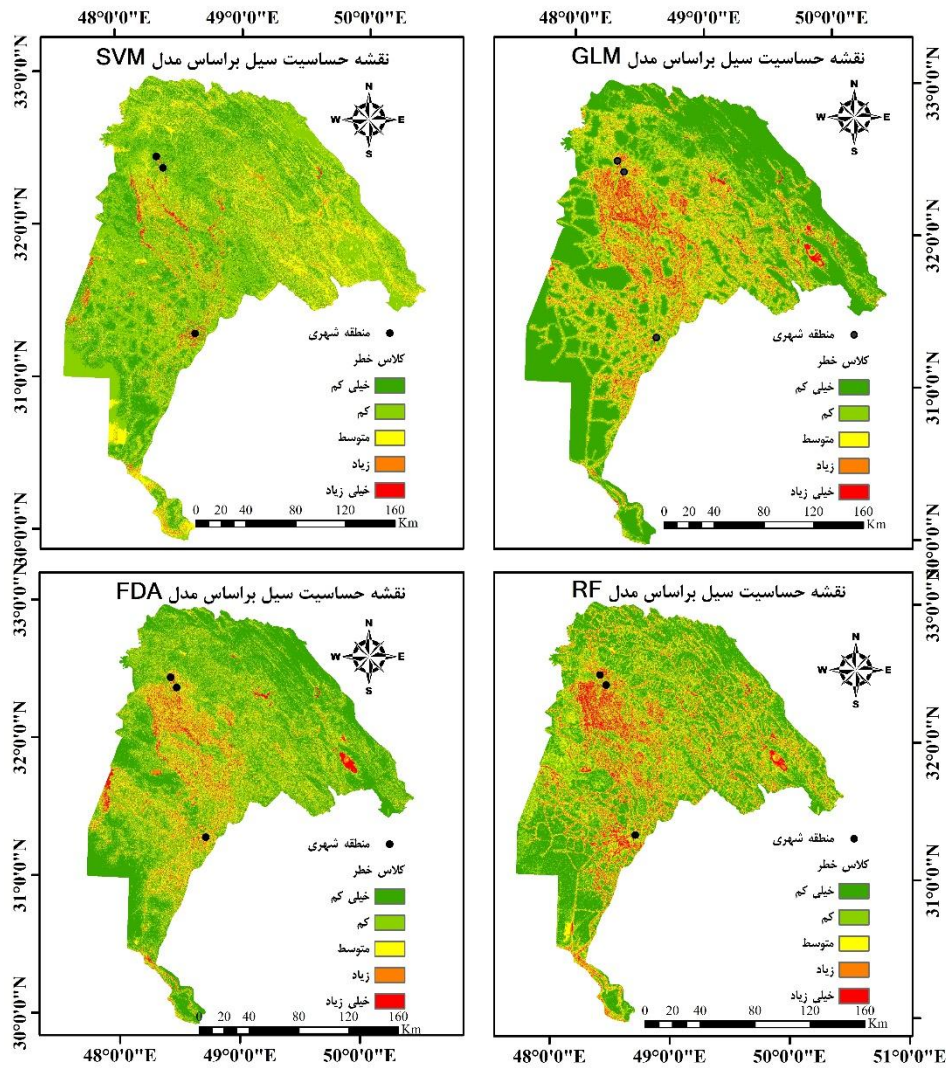
جدول ۱- نتایج بررسی هم‌خطی عوامل مؤثر بر سیل

Table 1- The results of the multi-collinearity test of conditioning factors for Flood risk

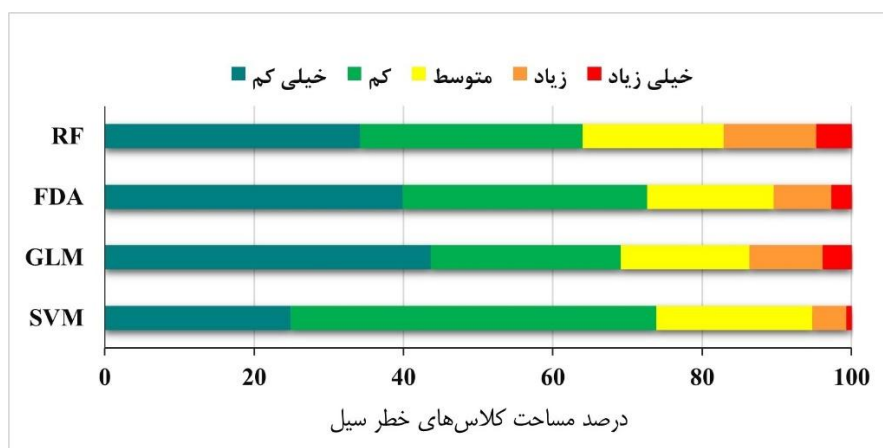
عوامل پیش‌بینی‌کننده	آستانه تحمل	عامل تورم واریانس
انحنای طولی	0.71	1.40
NDVI	0.77	1.28
پوشش/کاربری	0.83	1.20
سنگ‌شناسی	0.92	1.08
تراکم زهکش	0.52	1.91
فاصله از آبراهه	0.72	1.39
شاخص رطوبت توپوگرافی	0.47	2.09
ارتفاع	0.32	3.05
فاصله از جاده	0.85	1.17
بارش	0.25	4.02
شیب (درجه)	0.27	3.76
جهت شیب	0.87	1.15
انحنای عرضی	0.63	1.59

۳-۲- نقشه حساسیت وقوع سیل

موقعیت ۷۰ درصد از نقاط آموزش مربوط به محل وقوع و عدم وقوع سیل به‌همراه لایه‌های مربوط به عواملی که می‌توانند در الگوی خطر سیل در منطقه مؤثر باشند تهیه و مدل‌سازی احتمال وقوع سیل توسط چهار مدل GLM، SVM، FDA، RF انجام شد. نقشه حساسیت سیل حاصل از مدل‌های مذکور طبق روش Jenk's Natural Break در GIS به پنج طبقه خطر خیلی کم، کم، متوسط، زیاد و خیلی زیاد کلاس‌بندی شد (Wen et al., 2022). شکل ۴ نقشه حساسیت سیل را براساس هر یک از مدل‌های یادگیری ماشین مورد بررسی نشان می‌دهد. براین اساس مناطق کم‌ارتفاع و اطراف رودخانه‌ها و به سمت پایین‌دست بیش‌تر در معرض وقوع خطر می‌باشند. که یافته‌های این بخش با نتایج (Pourghasemi et al., 2020; Khosravi et al., 2016) هم‌خوانی دارد. از دلایل آن می‌توان به نزدیکی به خط‌القعر رودخانه، در دسترس بودن این مناطق جهت تغییر کاربری و تراکم بالای جمعیت اشاره کرد. همچنین، بررسی شکل ۴ نشان می‌دهد که موقعیت سه شهر پرجمعیت منطقه شامل اهواز، دزفول و اندیمشک و اراضی کشاورزی اطراف آن‌ها براساس روش‌های GLM، FDA، RF در کلاس‌های احتمال خطر خیلی زیاد قرار گرفته‌اند. شکل ۵ درصد فراوانی نسبی کلاس‌های مختلف خطر سیل را برای مدل‌های مورد بررسی نشان می‌دهد. با توجه به نمودار در هر چهار مدل، احتمال وقوع خطر سیل در بیش‌تر از نصف سطح حوزه خیلی کم و کم بوده است. به‌طوری‌که طبق نتایج مدل‌های GLM، SVM، FDA، RF به‌ترتیب ۷۳/۹ درصد، ۶۹ درصد، ۷۲/۶ درصد و ۶۳/۹ درصد از سطح حوزه به این کلاس‌ها تعلق داشت و بقیه آن در کلاس‌های متوسط تا خیلی زیاد قرار گرفت. همچنین براساس نتایج مدل‌های RF و GLM در سطح بیش‌تری از منطقه پتانسیل وقوع سیل زیاد و خیلی زیاد بود؛ به‌طوری‌که در مدل RF و GLM به‌ترتیب ۴/۷ و ۳/۹ درصد از سطح منطقه به کلاس خطر خیلی زیاد تعلق داشت.



شکل ۴- نقشه حساسیت سیل بخشی از استان خوزستان براساس مدل‌های SVM، GLM، FDA، RF و
 Figure 4 - Flood susceptibility map of a part of Khuzestan province based on the SVM, GLM, FDA, and RF models

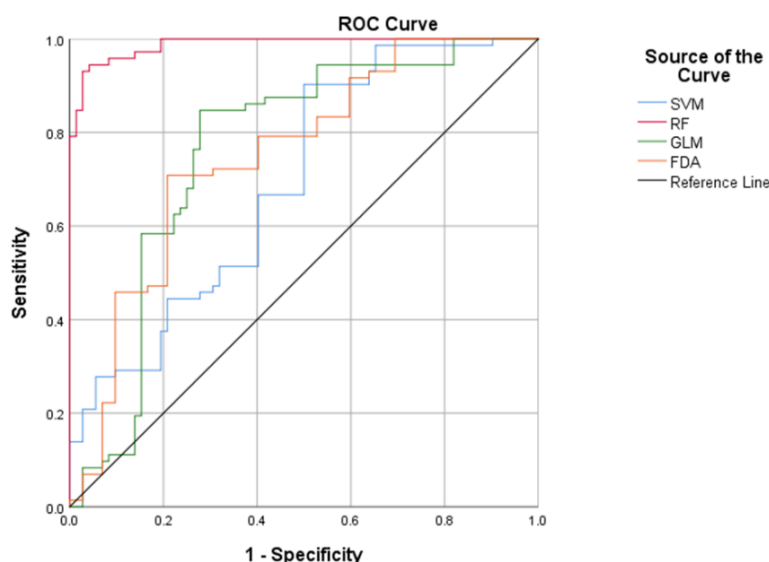


شکل ۵- درصد مساحت کلاس‌های خطر در مدل‌های مختلف یادگیری ماشین
 Figure 5- Percentage of area for risk classes in different machine learning models

۳-۳- نتایج ارزیابی عملکرد مدل‌ها

ارزیابی صحت مدل‌های مورد استفاده از طریق مقایسه نقشه‌های حاصل از مرحله آزمایش و ۳۰ درصد نقاط خطر مربوط به اعتبارسنجی صورت گرفت. همچنین، معیار ارزیابی در این مرحله سطح زیر منحنی نمودار ROC بود، که در شکل ۶ نشان داده شده است. همان‌گونه که مشاهده می‌شود، مدل RF در شبیه‌سازی خطر سیل از بالاترین کارایی برخوردار بود و مقدار سطح زیر منحنی برای این مدل ۹۸/۸ درصد به‌دست آمد و به‌ترتیب مدل‌های GLM، FDA و SVM از صحت پایین‌تری برخوردار بودند. از آنجایی که سطح زیر منحنی برای مدل‌های GLM و FDA به‌ترتیب ۷۶/۳ و ۷۵/۲ درصد و بالاتر از ۷۵ درصد به‌دست آمد (Guzzetti et al., 2005)، بنابراین نتایج این مدل‌ها نیز از صحت مناسبی برخوردار بود. اما مدل SVM با مقدار AUC ۶۹/۵ درصد برای شبیه‌سازی خطر سیل در این منطقه پیشنهاد نمی‌شود.

مدل جنگل تصادفی در بسیاری از مطالعات نه تنها در پیش‌بینی خطر سیل بلکه در مدل‌سازی سایر مخاطرات مانند زمین‌لغزش و آتش‌سوزی نیز به‌عنوان مدلی کارآمد شناخته شده است (Pourghasemi et al., 2023; Youssef et al., 2023). از دلایل کارایی بالای مدل RF می‌توان به ساختار این مدل اشاره کرد، به‌طوری که یک مدل ترکیبی از چندین درخت تصمیم است که نسبت به یک درخت تصمیم نتایج بهتری را ارائه می‌کند (Cutler et al., 2007). همچنین، هر درخت از چندین الگوریتم طبقه‌بندی استفاده می‌کند که نسبت به یک الگوریتم نتیجه بهتری به همراه دارد. بنابراین از آنجایی که در این مدل از چندین درخت استفاده می‌شود واریانس و خطای کم‌تری در طبقه‌بندی ایجاد می‌شود.



شکل ۶- منحنی مشخصه عامل گیرنده (ROC) مدل‌های مختلف

Figure 6- Receiver Operating Characteristic (ROC) curve for different models

۴- نتیجه‌گیری

این مطالعه به بررسی عملکرد چهار مدل یادگیری ماشین SVM، GLM، FDA و RF در زمینه تهیه نقشه حساسیت خطر سیل بخشی از استان خوزستان براساس شاخص سطح زیر منحنی ROC پرداخت. نتایج ارزیابی نشان داد که مدل جنگل تصادفی (RF) از صحت بسیار بالایی با سطح زیر منحنی ۹۸/۸ درصد برخوردار بوده است و به‌عنوان مناسب‌ترین مدل در پیش‌بینی نقاط خطر وقوع سیلاب شناسایی شد. طبق نتایج این مدل کلاس‌های با احتمال خیلی کم، کم، متوسط، زیاد و خیلی زیاد وقوع سیل به‌ترتیب ۳۴/۲، ۲۹/۷، ۱۸/۹، ۱۲/۴ و ۴/۷ درصد از سطح حوزه را به‌خود اختصاص داده بود؛ که مناطق پرخطر اغلب در حاشیه رواناب

سطحی و اراضی هموار واقع شده بودند. لازم به‌ذکر است که علاوه بر مدل RF مدل‌های GLM و FDA نیز براساس شاخص سطح زیر منحنی از دقت قابل قبولی برخوردار بودند (به‌ترتیب AUC برابر ۷۶/۳ درصد و ۷۵/۲ درصد). این نتایج بر توانایی مدل‌های یادگیری ماشین در پیش‌بینی احتمال وقوع خطر تأکید دارد. به‌دنبال افزایش جمعیت، توسعه مناطق شهری و زیرساخت‌ها در مناطق کوهستانی و دشت‌های سیلابی ارائه نقشه‌های حساسیت مخاطرات مختلف و نقشه‌های چندمخاطره‌ای به‌منظور توسعه پایدار ضروری به‌نظر می‌رسد. بنابراین، پیشنهاد می‌شود که در مطالعات پیش‌رو ارزیابی مدل‌های مختلف یادگیری ماشین و تهیه نقشه حساسیت برای سایر مخاطرات محتمل در منطقه مورد بررسی و در نهایت ارائه

مفهوم‌سازی، کنترل نتایج؛ حمیدرضا کشت‌کار: ویرایش و بازبینی مقاله، کنترل نتایج.

منابع

- چراغی قلعه سری، علی، حبیب‌نژاد روشن، محمود، و روشن، سیدحسین. (۱۳۹۹). تهیه نقشه حساسیت سیلاب با استفاده از مدل ماشین بردار پشتیبان (SVM) و سیستم اطلاعات جغرافیایی (GIS). *مخاطرات محیط طبیعی*، ۹(۲۵)، ۶۱-۸۰. doi:10.22111/jneh.2020.31018.1547
- رجبی‌زاده، یوسف، ایوب‌زاده، سیدعلی، و قمشی، مهدی (۱۳۹۸). بررسی سیل استان خوزستان طی سال آبی ۱۳۹۷-۱۳۹۸ و ارائه راهکارهای کنترل و مدیریت آن در آینده. *اکوهیدرولوژی*، ۶(۴)، ۱۰۶۹-۱۰۸۴. doi:10.22059/ije.2020.285854.1166
- زارع چاهوکی، محمدعلی، خلاصی اهواری، لیلا، و آذرنبوند، حسین (۱۳۹۰). مدل‌سازی پراکنش گونه‌های گیاهی بر اساس عوامل خاک و توپوگرافی با استفاده از روش رگرسیون لجستیک در مراتع شرق سمنان. *مرتع و آبخیزداری*، ۶۷(۱)، ۴۵-۵۹. doi: 10.22059/jrwm.2014.50827

References

- AlQahtany, A. M., & Abubakar, I. R. (2020). Public perception and attitudes to disaster risks in a coastal metropolis of Saudi Arabia. *Disaster Risk Reduction*, 44, 101422. doi:10.1016/j.ijdrr.2019.101422
- Bozorgmehr, S. (2019). Southwest Iran hit hard by flooding, evacuation underway in Ahvaz. Reuters. Retrieved 11 April 2019. <https://www.reuters.com/article/us-iran-foods/southwest-iran-hit-hard%02by-fooding-evacuation-underway-in-ahvaz-idUSKCN1RM1G6>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32. doi:10.1023/A:1010933404324
- Chen, W., Chai, H., Zhao, Z., Wang, Q., & Hong, H. (2016). Landslide susceptibility mapping based on GIS and support vector machine models for the Qianyang County, China. *Environmental Earth Sciences*, 75(6), 474. doi:10.1007/s12665-015-5093-0
- Cheraghi Ghalehsari, A., Habibnejad Roshan, M., & Roshun, S.H., (2020) Flood susceptibility mapping using a support vector machine

نقشه چندمخاطره‌ای مد نظر قرار گرفته شود. نتایج تحقیق حاضر به تصمیم‌گیران و سیاست‌گزاران کمک می‌کند تا تصمیمات مدیریتی منسجمی برای پیشرفت‌های دوره حاضر و آینده بگیرند.

سیاسگزاری

از همکاری و مساعدت دانشگاه تهران و دانشگاه شیراز در انجام مراحل مختلف این پژوهش تشکر و قدردانی می‌شود.

تضاد منافع نویسندگان

نویسندگان این مقاله اعلام می‌دارند که هیچ‌گونه تضاد منافی در خصوص نگارش و انتشار مطالب و نتایج این تحقیق وجود ندارند.

دسترسی به داده‌ها

تمامی اطلاعات و نتایج در متن مقاله ارائه شده است.

مشارکت نویسندگان

رویت قنواتی: نگارش نسخه اولیه مقاله و ویرایش، بصری‌سازی؛ علی سلاجقه: تهیه اطلاعات و داده‌های مورد نیاز، کنترل نتایج، راهنمایی؛ حمیدرضا پورقاسمی: انجام تحلیل‌های آماری و نرم‌افزاری، روش‌شناسی، بصری‌سازی؛ شهرام خلیقی سیگارودی:

models (SVM) and geographic information system (GIS), *Natural Environmental Hazards*, 9(25), 61-80. [In Persian] doi:10.22111/jneh.2020.31018.1547

- Conoscenti, C., Agnesi, V., Angileri, S., Cappadonia, C., Rotigliano, E., & Märker, M. (2013). A GIS-based approach for gully erosion susceptibility modelling: a test in Sicily, Italy. *Environmental Earth Sciences*, 70(3), 1179-1195. doi:10.1007/s12665-012-2205-y
- Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, 88(11), 2783-2792. <https://doi.org/10.1890/07-0539.1>
- Dehnavi, A., Aghdam, I. N., Pradhan, B., & Morshed Varzandeh, M. H. (2015). A new hybrid model using step-wise weight assessment ratio analysis (SWARA) technique and adaptive neuro-fuzzy inference system (ANFIS) for regional landslide hazard assessment in Iran. *Catena*, 135, 122-148. doi:10.1016/j.catena.2015.07.020
- Goetz, J. N., Brenning, A., Petschko, H., & Leopold, P. (2015). Evaluating machine learning and

- statistical prediction techniques for landslide susceptibility modeling. *Computers & Geosciences*, 81, 1–11. doi:10.1016/j.cageo.2015.04.007
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. *MIT Press*, 1, 98–164. <http://www.deeplearningbook.org>
- Gupta, L., & Dixit, J. (2022). A GIS-based flood risk mapping of Assam, India, using the MCDA-AHP approach at the regional and administrative level. *Geocarto International*, 37(26), 11867–11899. doi:10.1080/10106049.2022.2060329
- Guzzetti, F., Reichenbach, P., Cardinali, M., Galli, M., & Ardizzone, F. (2005). Probabilistic landslide hazard assessment at the basin scale. *Geomorphology*, 72(1–4), 272–299. doi:10.1016/j.geomorph.2005.06.002
- 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 Sensing*, 16(5), 858. doi:10.3390/rs16050858
- Hong, H., Pradhan, B., Bui, D. T., Xu, C., Youssef, A. M., & Chen, W. (2017). Comparison of four kernel functions used in support vector machines for landslide susceptibility mapping: a case study at Suichuan area (China). *Geomatics, Natural Hazards and Risk*, 8(2), 544–569. doi:10.1080/19475705.2016.1250112
- Kalantar, B., Pradhan, B., Naghibi, S. A., Motevalli, A., & Mansor, S. (2018). Assessment of the effects of training data selection on the landslide susceptibility mapping: a comparison between support vector machine (SVM), logistic regression (LR) and artificial neural networks (ANN). *Geomatics, Natural Hazards and Risk*, 9(1), 49–69. doi:10.1080/19475705.2017.1407368
- Kalantar, B., Ueda, N., Saeidi, V., Ahmadi, K., Halin, A. A., & Shabani, F. (2020). Landslide susceptibility mapping: machine and ensemble learning based on remote sensing big data. *Remote Sensing*, 12(11), 1737. doi:10.3390/rs12111737
- Khosravi, K., Pourghasemi, H. R., Chapi, K., & Bahri, M. (2016). Flash flood susceptibility analysis and its mapping using different bivariate models in Iran: a comparison between Shannon's entropy, statistical index, and weighting factor models. *Environmental Monitoring and Assessment*, 188(12), 656. doi:10.1007/s10661-016-5665-9
- Masoudi, M., & Elhaesahar, M. (2016). Trend assessment of climate changes in Khuzestan Province, Iran. *Natural Environment Change*, 2(2), 143–152. https://journals.ut.ac.ir/article_60997.html
- Nachappa, T., Ghorbanzadeh, O., Gholamnia, K., & Blaschke, T. (2020). Multi-hazard exposure mapping using machine learning for the state of Salzburg, Austria. *Remote Sensing*, 12(17), 2757. doi:10.3390/rs12172757
- Pourghasemi, H. R., Gayen, A., Edalat, M., Zarafshar, M., & Tiefenbacher, J. P. (2020). Is multi-hazard mapping effective in assessing natural hazards and integrated watershed management? *Geoscience Frontiers*, 11(4), 1203–1217. doi:10.1016/j.gsf.2019.10.008
- Pourghasemi, H. R., Pouyan, S., Bordbar, M., Golkar, F., & Clague, J. J. (2023). Flood, landslides, forest fire, and earthquake susceptibility maps using machine learning techniques and their combination. *Natural Hazards*, 116(3), 3797–3816. doi:10.1007/s11069-023-05836-y
- Pourghasemi, H. R., & Rahmati, O. (2018). Prediction of the landslide susceptibility: Which algorithm, which precision? *CATENA*, 162, 177–192. doi:10.1016/j.catena.2017.11.022
- Rahman, M., Chen, N., Islam, M. M., Mahmud, G. I., Pourghasemi, H. R., Alam, M., Rahim, M. A., Baig, M. A., Bhattacharjee, A., & Dewan, A. (2021). Development of flood hazard map and emergency relief operation system using hydrodynamic modeling and machine learning algorithm. *Cleaner Production*, 311, 127594. doi:10.1016/j.jclepro.2021.127594
- Rajabizadeh, Y., Ayyoubzadeh, S. A., & Gholami, M. (2020). Flood survey of Khuzestan province in 97-98 and providing solutions for its control and management in the future. *Ecohydrology*, 6(4), 1069–1084. [In Persian] doi:10.22059/ije.2020.285854.1166.
- Rahman, M., Ningsheng, C., Islam, M. M., Dewan, A., Iqbal, J., Washakh, R. M. A., & Shufeng, T. (2019). Flood susceptibility assessment in Bangladesh using machine learning and multi-criteria decision analysis. *Earth Systems and Environment*, 3(3), 585–601.

- doi:10.1007/s41748-019-00123-y
- Ramsay, J. O., & Dalzell, C. J. (1991). Some Tools for functional data analysis. *Royal Statistical Society. Series B (Methodological)*, 53(3), 539–572.
doi:10.1111/j.2517-6161.1991.tb01844.x
- Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., & Guzzetti, F. (2018). A review of statistically-based landslide susceptibility models. *Earth-Science Reviews*, 180, 60–91.
doi:10.1016/j.earscirev.2018.03.001
- Rossi, M., & Reichenbach, P. (2016). LAND-SE: a software for statistically based landslide susceptibility zonation, version 1.0. *Geoscientific Model Development*, 9(10), 3533–3543. doi:10.5194/gmd-9-3533-2016
- Rutgersson, A., Kjellström, E., Haapala, J., Stendel, M., Danilovich, I., Drews, M., Jylhä, K., Kujala, P., Larsén, X. G., Halsnæs, K., Lehtonen, I., Luomaranta, A., Nilsson, E., Olsson, T., Särkkä, J., Tuomi, L., & Wasmund, N. (2022). Natural hazards and extreme events in the Baltic Sea region. *Earth System Dynamics*, 13(1), 251–301. doi:10.5194/esd-13-251-2022
- Saharia, M., Jain, A., Baishya, R. R., Haobam, S., Sreejith, O. P., Pai, D. S., & Rafieenasab, A. (2021). India flood inventory: creation of a multi-source national geospatial database to facilitate comprehensive flood research. *Natural Hazards*, 108(1), 619–633.
doi:10.1007/s11069-021-04698-6
- Sahoo, S. N., & Sreeja, P. (2017). Development of flood inundation maps and quantification of flood risk in an urban catchment of Brahmaputra River. *ASCE-ASME Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 3(1).
doi:10.1061/AJRUA6.0000822
- Satarzadeh, E., Sarraf, A., Hajikandi, H., & Sadeghian, M. S. (2022). Flood hazard mapping in western Iran: assessment of deep learning vis-à-vis machine learning models. *Natural Hazards*, 111(2), 1355–1373.
doi:10.1007/s11069-021-05098-6
- Schoppa, L., Disse, M., & Bachmair, S. (2020). Evaluating the performance of random forest for large-scale flood discharge simulation. *Hydrology*, 590, 125531.
doi:10.1016/j.jhydrol.2020.125531
- Segond, M.-L., Wheeler, H. S., & Onof, C. (2007). The significance of spatial rainfall representation for flood runoff estimation: A numerical evaluation based on the Lee catchment, UK. *Hydrology*, 347(1–2), 116–131. doi:10.1016/j.jhydrol.2007.09.040
- Segue, W. S., Njilah, I. K., Fossi, D. H., & Nsangou, D. (2024). Advancements in mapping landslide susceptibility in Bafoussam and its surroundings area using multi-criteria decision analysis, statistical methods, and machine learning models. *African Earth Sciences*, 213, 105237. doi:10.1016/j.jafrearsci.2024.105237
- Towfiqul Islam, A. R. M., Talukdar, S., Mahato, S., Kundu, S., Eibek, K. U., Pham, Q. B., Kuriqi, A., & Linh, N. T. T. (2021). Flood susceptibility modelling using advanced ensemble machine learning models. *Geoscience Frontiers*, 12(3), 101075.
doi:10.1016/j.gsf.2020.09.006
- Vafakhah, M., Mohammad Hasani Loor, S., Pourghasemi, H., & Katebikord, A. (2020). Comparing performance of random forest and adaptive neuro-fuzzy inference system data mining models for flood susceptibility mapping. *Geosciences*, 13(11), 417.
doi:10.1007/s12517-020-05363-1
- Vojtek, M., & Vojteková, J. (2019). Flood susceptibility mapping on a national scale in Slovakia using the analytical hierarchy process. *Water*, 11(2), 364.
doi:10.3390/w11020364
- Wen, T., Tiewang, W., Arabameri, A., Asadi Nalivan, O., Pal, S. C., Saha, A., & Costache, R. (2022). Land-subsidence susceptibility mapping: assessment of an adaptive neuro-fuzzy inference system–genetic algorithm hybrid model. *Geocarto International*, 37(26), 12194–12218.
doi:10.1080/10106049.2022.2066198
- Yalcin, A. (2008). GIS-based landslide susceptibility mapping using analytical hierarchy process and bivariate statistics in Ardesen (Turkey): Comparisons of results and confirmations. *catena*, 72(1), 1–12.
doi:10.1016/j.catena.2007.01.003
- Youssef, A. M., Mahdi, A. M., Al-Katheri, M. M., Pouyan, S., & Pourghasemi, H. R. (2023). Multi-hazards (landslides, floods, and gully erosion) modelling and mapping using machine learning algorithms. *African Earth Sciences*, 197, 104788.
doi:10.1016/j.jafrearsci.2022.104788
- Yu, L., Wang, Y., & Pradhan, B. (2024). Enhancing

- landslide susceptibility mapping incorporating landslide typology via stacking ensemble machine learning in Three Gorges Reservoir, China. *Geoscience Frontiers*, 15(4). doi: 10.1016/j.gsf.2024.101802
- Zare Chahouki, M.A., Khalsi Ahvazi, L., & Azarnivand, H. (2020) Plant species distribution modeling using logistic regression models in the North East of Semnan, *Rangeland and watershed*, 67(1), 45-59. [In Persian] doi:10.22059/jrwm.2014.50827
- Zhou, C., Yin, K., Cao, Y., Ahmed, B., Li, Y., Catani, F., & Pourghasemi, H. R. (2018). Landslide susceptibility modelling applying machine learning methods: A case study from Longju in the Three Gorges Reservoir area, China. *Computers & Geosciences*, 112, 23–37. doi:10.1016/j.cageo.2017.11.019