



Preparing a flood sensitivity map in the Eghlid urban basin using machine learning algorithms

Saeed Negahban¹ , Mehri Marhamat² 

1. (Corresponding Author) Department of Geography, Faculty of Economics, Management and Social Sciences, Shiraz University, Shiraz, Iran

Email: snegahban@shirazu.ac.ir

2. Department of Physical Geography, Faculty of Social Sciences, Mohaghegh Ardabili University, Ardabil, Iran

Email: meri.marhamat90@gmail.com

Article Info

Article type:

Research Article

Article History:

Received:

3 June 2025

Revised:

2 September 2025

Accepted:

8 October 2025

Available online:

10 November 2025

Keywords:

Flood Hazard Sensitivity Map,
Machine Learning,
Random Forest,
Support Vector Machine,
Arc GIS Pro.

ABSTRACT

Flooding is among the most destructive natural hazards, causing substantial human casualties and economic losses worldwide each year. The preparation of flood risk maps and the identification of flood-prone areas play a critical role in reducing the potential damages associated with flooding. In this study, a flood sensitivity map was developed for the urban watershed of Eqlid in Fars Province using a range of flood conditioning factors, including slope, TRASP, topographic position index, stream power index, distance from the river, drainage density, surface curvature, rainfall, normalized difference vegetation index (NDVI), geology, land use, and soil type. Spatial data processing and factor preparation were conducted using ArcGIS Pro software. Subsequently, flood risk maps were generated using Random Forest (RF) and Support Vector Machine (SVM) machine learning algorithms implemented in the R programming environment, and the results of the two models were systematically compared. Model performance was evaluated using the receiver operating characteristic (ROC) curve, yielding accuracy values of 0.83 for the RF model and 0.84 for the SVM model, indicating good and nearly equivalent predictive performance. Feature importance analysis revealed that rainfall, soil type, land use, and drainage density were the most influential factors in the RF model, while rainfall, slope, soil type, and drainage density were identified as the dominant variables in the SVM model. The maps and datasets produced in this study can serve as valuable tools for urban planners and decision-makers in preventive planning, enhancing public awareness, and designing flood-resilient infrastructure.

Cite this article: Negahban, S., & Marhamat, M. (2025). Preparing a flood sensitivity map in the Eghlid urban basin using machine learning algorithms. *Physical Geography Research Quarterly*, 57 (3), 43-62.

<http://doi.org/10.22059/jphgr.2025.406510.1007909>



© The Author(s).

Publisher: University of Tehran Press

Extended Abstract

Introduction

Flooding is one of the major natural hazards resulting from heavy rainfall, river overflows, and tidal influences. In recent years, increasing urbanization and climate change have intensified flood occurrences, particularly urban flooding, across many regions of the world. This phenomenon poses substantial social and economic risks, making accurate flood prediction essential for minimizing potential damages. Recent methodological advances, particularly in machine learning, have significantly enhanced the accuracy of flood hazard prediction. In this study, a flood hazard zoning map was developed for the urban watershed of Eqlid using advanced machine learning algorithms, namely Random Forest and Support Vector Machine. This approach integrates multi-source spatial data to support a comprehensive flood risk management strategy, with the potential to reduce damages associated with urban flooding.

Methodology

The study area covers approximately 40/623 hectares and has an elevation range from 1234 to 3213 meters above sea level, located in the southern part of Iran. Geographically, the watershed is located between 52°24'0" and 52°52'0" East longitude and between 30°43'40" and 31°4'20" North latitude. Initially, a flood point inventory map was prepared using 150 points, including both flood occurrence and non-occurrence locations, identified through field surveys, topographic maps, and data obtained from the General Directorate of Natural Resources of Fars Province. Additionally, 15 variables were selected for modeling based on methodologies commonly adopted in previous studies. These variables include elevation, slope, TRASP, drainage density, distance from the river, land use, geology, soil texture, Stream Power Index (SPI), Topographic Wetness Index (TWI), Topographic Roughness Index (TRI), Topographic Position Index (TPI), average annual rainfall, and the Normalized Difference Vegetation Index (NDVI). Multicollinearity among the selected variables was assessed using the Variance Inflation Factor (VIF) to identify potential

linear correlations. Subsequently, machine learning algorithms, namely Random Forest (RF) and Support Vector Machine (SVM), were employed for flood hazard modeling. Finally, model performance was evaluated using statistical indices, including Accuracy, Sensitivity, Specificity, and the Kappa coefficient, in addition to the receiver operating characteristic (ROC) curve.

Results and discussion

Examining multicollinearity among the selected variables is essential for the reliable production of flood sensitivity maps. Among the 15 selected variables, elevation, TWI, and TRI were excluded because of high VIF values. The flood sensitivity maps indicate that areas with low to moderate susceptibility are predominantly located in the northern, northwestern, and southwestern parts of the watershed, whereas zones with high flood susceptibility are mainly concentrated in the central area. Model performance evaluation showed that Sensitivity values for the training dataset were 0.93 for the RF model and 0.84 for the SVM model, while corresponding values for the validation dataset were 0.86 and 0.80, respectively. Specificity values reached 0.94 for RF and 0.83 for SVM in the training phase, and 0.86 and 0.73 in the validation phase. Accuracy values were 0.94 for RF and 0.89 for SVM for the training data, and 0.73 and 0.77 for the validation data, respectively. In addition, Kappa coefficients were 0.86 for RF and 0.73 for SVM in the training dataset, decreasing to 0.75 and 0.63 in the validation dataset. The AUC values for the RF model were 99% for the training data and 83% for the testing data, while the SVM model achieved AUC values of 93% and 84%, respectively, indicating high predictive accuracy for both models.

The variable importance analysis revealed that rainfall, soil type, land use, slope, and drainage density are the primary factors influencing flood occurrence, each contributing significantly to flood hazard formation. Although rainfall intensity tends to increase at higher elevations, the accumulation of runoff in lower elevation areas significantly elevates flood risk in these zones. Aridisols soils, characterized by low permeability and calcareous structure,

promote increased surface runoff and accelerate flood occurrence. Land use types such as urban areas, agricultural lands, and moderately vegetated zones contribute to intensified flooding as a result of increased impervious surfaces and reduced infiltration capacity. Gentle slopes reduce flow velocity and prolong water retention time, thereby increasing the likelihood of flooding. Although high drainage density can facilitate rapid water conveyance, it also concentrates runoff in downstream areas, potentially leading to sudden and destructive flood events.

These findings are consistent with previous studies and highlight the importance of considering the combined effects of natural and anthropogenic factors in urban planning and water resource management to mitigate flood hazards. Overall, the machine learning models employed in this study produced accurate and reliable flood hazard maps that can effectively support urban management decisions and contribute to reducing flood-related damages in the urban watershed of Eqlid.

Conclusion

In this study, advanced machine learning models, including Random Forest and Support Vector Machine, were employed to generate flood hazard potential maps for the urban watershed of Eqlid with a high level of accuracy. The findings indicate that the interaction of climatic, geological, and hydrological variables, together with land use changes and urbanization expansion, plays a fundamental role in flood occurrence. The proposed models demonstrated a strong capability to accurately predict flood sensitivity and to provide valuable information for decision-makers involved in flood damage reduction planning. However, certain limitations, including the focus on a single watershed and constraints related to data availability, should be taken into account in future research. Overall, this research emphasizes that the application of data-driven methods and modern modeling approaches, alongside scientific analyses, can serve as an effective strategy for reducing the destructive impacts of floods in urban areas and supporting policymakers in informed decision-making.

Funding

There is no funding support.

Authors' Contribution

Conceptualization and formal analysis and initial methodology and investigation by S.Negahban.; project administration and supervision and visualization and software writing review and editing by M.Marhamat.;

Conflict of Interest

Authors declared no conflict of interest.

Acknowledgments

We are grateful to all the scientific consultants of this paper.

تهیه نقشه حساسیت سیل در حوزه شهری اقلید با استفاده از الگوریتم‌های یادگیری ماشینی

سعید نگهبان^۱ ✉، مهری مرحمت^۲ ^۱ - نویسنده مسئول، گروه جغرافیا، دانشکده اقتصاد، مدیریت و علوم اجتماعی، دانشگاه شیراز، شیراز، ایران. رایانامه: snegahban@shirazu.ac.ir^۲ - گروه جغرافیای طبیعی، دانشکده علوم اجتماعی، دانشگاه محقق اردبیلی، اردبیل، ایران. رایانامه: meri.marhamat90@gmail.com

چکیده	اطلاعات مقاله
سیل از جمله بلایای طبیعی است که هر ساله خسارات مالی و جانی زیادی را در سراسر جهان به بار می‌آورد. با تهیه نقشه خطر سیل و تعیین مناطق مستعد سیل، می‌توان خسارات احتمالی این پدیده را کاهش داد. برای تهیه نقشه حساسیت سیل در حوزه شهری اقلید (استان فارس) عوامل مؤثر بر وقوع سیل شامل شیب، شاخص تابش خورشیدی، شاخص موقعیت توپوگرافی، شاخص قدرت جریان، فاصله از رودخانه، تراکم زهکشی، انحنای سطح، بارندگی، شاخص پوشش گیاهی، سنگ‌شناسی، کاربری اراضی و نوع خاک استفاده شد. سپس، برای تهیه نقشه خطر سیل، از الگوریتم جنگل تصادفی (RF) و ماشین بردار پشتیبان (SVM) به عنوان دو مدل یادگیری ماشینی در نرم‌افزار R پیاده‌سازی و نتایج آنها با هم مقایسه شده است. مدل‌های RF و SVM بر اساس منحنی ROC به ترتیب با دقت (۰/۸۳ و ۰/۸۴) اعتبارسنجی شده‌اند که هر دو عملکرد خوب و تقریباً یکسانی داشته‌اند. با تحلیل اهمیت ویژگی‌های لایه‌ها در مدل RF مشخص شد که مؤلفه‌های بارندگی، خاک، کاربری اراضی و تراکم زهکشی بیش‌ترین تأثیر را بر خطر سیل دارند؛ همچنین در مدل SVM، بارندگی، شیب، خاک و تراکم زهکشی به ترتیب مهم‌ترین متغیرها برای پیش‌بینی خطر سیل شناخته شدند. نقشه‌ها و داده‌های حاصل از این مطالعه می‌توانند به مسئولین و مدیران شهری به عنوان ابزارهای کلیدی برای برنامه‌ریزی‌های پیشگیرانه، آگاه‌سازی عمومی و طراحی زیرساخت‌های مقاوم در برابر سیل کمک کنند.	نوع مقاله: مقاله پژوهشی تاریخ دریافت: ۱۴۰۴/۰۳/۱۳ تاریخ بازنگری: ۱۴۰۴/۰۶/۱۱ تاریخ پذیرش: ۱۴۰۴/۰۷/۱۶ تاریخ چاپ: ۱۴۰۴/۰۸/۱۹ واژگان کلیدی: جنگل تصادفی، ماشین بردار پشتیبان، نقشه حساسیت خطر سیل، یادگیری ماشینی، ARCGISPro

استناد: نگهبان، سعید و مرحمت، مهری. (۱۴۰۴). تهیه نقشه حساسیت سیل در حوزه شهری اقلید با استفاده از الگوریتم‌های یادگیری ماشینی. مجله پژوهش‌های جغرافیای طبیعی، ۵۷ (۳)، ۴۳-۶۲.

<http://doi.org/10.22059/jphgr.2025.406510.1007909>

ناشر: مؤسسه انتشارات دانشگاه تهران

© نویسندگان



مقدمه

یکی از مخاطرات طبیعی، سیل است و زمانی رخ می‌دهد که آب ناشی از بارندگی شدید، طغیان رودخانه‌ها یا امواج جزر و مد، مناطقی را که معمولاً خشک هستند، زیر آب ببرد (Ma et al, 2021). سیل‌ها اغلب در اثر بارندگی شدید در یک دوره کوتاه، که در آن سیستم‌های زهکشی قادر به مدیریت حجم آب نیستند، یا به دلیل خرابی زیرساخت‌ها (مانند خاکریزها یا سدها) ایجاد می‌شوند (Ward et al, 2020). در سال‌های اخیر، شتاب تدریجی شهرنشینی و افزایش تغییرات آب‌وهوایی مشاهده شده است. این امر منجر به وقوع فجایع قابل توجه سیل شهری در شهرهای سراسر جهان شده و خسارات اجتماعی - اقتصادی زیادی به بار آورده است، به طوری که جان و مال مردم را تهدید می‌کند (Alfieri et al, 2017). پیش‌بینی‌ها نشان می‌دهد که در ۳۰ سال آینده، خطر سیل در سراسر جهان تا ده برابر افزایش خواهد یافت. این افزایش عمدتاً ناشی از وقوع سیل‌های مکرر و شدیدتر به دلیل تغییرات آب‌وهوایی، افزایش جمعیت و گسترش اقتصادی در مناطق آسیب‌پذیر است. به عبارت دیگر سیل شهری به عنوان یک نگرانی مهم در زمینه ایمنی شهری مطرح شده است که مانع پیشرفت‌های اقتصادی و اجتماعی می‌شود (Xiao et al, 2023).

ارزیابی ریسک سیلاب شهری برای اقدامات پیشگیرانه جهت حفاظت از شهر حیاتی است. روش‌های رایج مورد استفاده برای ارزیابی سیل شهری را می‌توان به مدل‌سازی تجربی، شبیه‌سازی عددی و یادگیری ماشینی طبقه‌بندی کرد. سیستم‌های سنتی پیش‌بینی سیل اغلب دقت و قطعیت کافی ندارند که این موضوع به آمادگی ناکافی منجر می‌شود. بنابراین، سیستم‌های مدرن که از انواع داده‌ها و الگوریتم‌های مختلف بهره می‌برند، به شدت مورد نیاز هستند تا پیش‌بینی سیل را بهبود دهند (Duan et al, 2022). در حال حاضر محققان از روش‌های یادگیری ماشین برای ارزیابی سیل شهری استفاده کرده‌اند به عنوان مثال (امیری و همکاران، ۱۴۰۱) پژوهشی با عنوان آینده‌پژوهی و پتانسیل سنجی خطر وقوع سیل با استفاده از الگوریتم یادگیری ماشین (SVM) برای تحلیل داده‌های مکانی و محیطی مرتبط با وقوع سیل و تهیه نقشه پتانسیل خطر سیلاب شهری در شهر تهران پرداختند و به این نتیجه رسیدند که مدل SVM دقت مناسبی در تعیین مناطق خطر سیل دارد و می‌تواند به عنوان ابزاری کارآمد در مدیریت مخاطرات سیلاب شهری مورد استفاده قرار گیرد. (رضایی و همکاران، ۱۴۰۳) به مدل‌سازی مکانی و تهیه نقشه پتانسیل سیل‌گیری با استفاده از الگوریتم‌های یادگیری ماشین (SVM، GAM، RF) در استان بوشهر پرداختند و به این نتیجه رسیدند که عوامل مؤثر مهم بر پتانسیل سیل‌گیری در منطقه مورد مطالعه شامل ارتفاع، بارندگی و سنگ‌شناسی بوده‌اند که بیشترین تأثیر را دارند و مدل جنگل تصادفی (Random Forest) بهترین دقت را داشته و توانسته است نقشه باکیفیت و قابل اتکایی برای مدیریت ریسک سیل در استان بوشهر فراهم کند. (حیفی نیا و عبقری، ۱۴۰۴) به پیش‌بینی مناطق مستعد سیل با استفاده از مدل‌های یادگیری ماشین (خطی تعمیم‌یافته و بیشینه آنتروپی) در حوضه آبخیز زیوه (ارومیه) پرداختند و به این نتیجه رسیدند که مدل بیشینه آنتروپی با شاخص AUC برابر با ۰/۹۱۶ و مدل خطی تعمیم‌یافته با AUC برابر با ۰/۹۰۲ عملکرد بسیار خوبی در پهنه‌بندی حساسیت به سیل از خود نشان داده‌اند. همچنین شاخص کاپا تحلیل‌ها نشان داده که عوامل محیطی مهمی مانند زمین‌شناسی، فاصله از آبراهه، ارتفاع و شیب بیشترین تأثیر را در وقوع سیل داشته‌اند و برخی عوامل مانند شاخص انحنا، پروفیل، کاربری اراضی و شاخص تعادل جرم تأثیر کمتری داشته‌اند. (Noymanee et al, 2017) به پیش‌بینی سیل سطحی با استفاده از داده‌های آزاد (تصاویر ماهواره، داده‌های ایستگاه‌های بارش و هواشناسی، ارتفاع، کاربری زمین، هیدرومتری) و تکنیک‌های یادگیری ماشینی (ANN، SVM، Random Forest، LSTM، GBM) در حوضه پاتانی (تایلند) پرداختند و به این نتیجه رسیدند که مدل‌های یادگیری ماشین به طور کلی توانسته‌اند پیش‌بینی‌های دقیق‌تری نسبت به روش‌های سنتی ارائه کنند و استفاده از داده‌های آزاد در کنار این الگوریتم‌ها سبب افزایش دقت مدل‌ها شده است. معمولاً جنگل تصادفی (Random Forest) یا

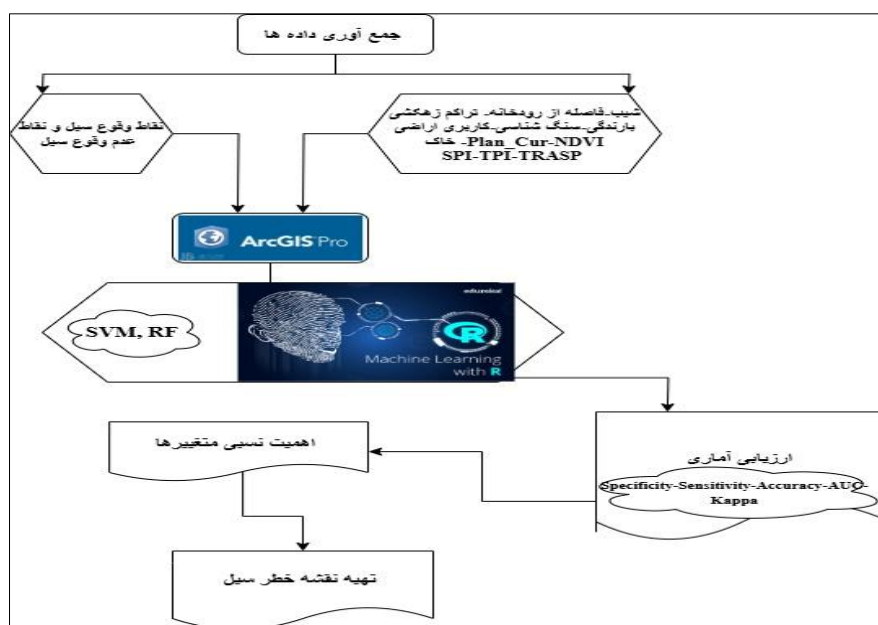
مدل‌های تقویتی گرادیان (Gradient Boosting Machines) در پیش‌بینی سیل عملکرد بالاتری دارند. (گنجی‌راد و دلاور، ۲۰۲۳) به نقشه‌برداری ریسک سیل با استفاده از مدل‌های یادگیری ماشین (Random Forest (RF) و (Support Vector Machine (SVM) برای ارزیابی خطر سیل در بخش Calcasieu از ایالت لوئیزیانا، آمریکا پرداختند و به این نتیجه رسیدند که نقشه‌های ریسک سیل تولید شده توسط مدل RF قادر به تمایز بهتر مناطق با ویژگی‌های مختلف جغرافیایی و خطر سیل هستند و مناطق با ارتفاع پایین و سیل‌خیز مانند جنوب غرب Calcasieu در کلاس ریسک بسیار بالا قرار گرفته‌اند. همچنین اهمیت فاکتورهای ورودی در مدل RF نشان داد که ارتفاع، شیب و انحناهای پلان بیشترین تأثیر را در پیش‌بینی ریسک سیل داشتند. (Dai et al, 2024) به پیش‌بینی سیلاب شهری با استفاده از شبکه عصبی مصنوعی ترکیبی (Ensemble ANN) به منظور بررسی بهبود عدم قطعیت مدل در شهر ساحلی ماکائو (Macao) در چین پرداختند و به این نتیجه رسیدند که استفاده از مدل‌های ترکیبی ANN و بهینه‌سازی ساختار و ورودی‌ها می‌تواند به کاهش ناپایداری و عدم قطعیت پیش‌بینی سیلاب‌های شهری کمک کند و باعث افزایش دقت و قابلیت اطمینان در مدیریت بحران و کاهش خسارات گردد. (Rizal et al, 2025) به بهبود پیش‌بینی سیلاب در مناطق شهری با استفاده از رویکرد ماشینی (Gaussian Naive Bayes, k-Nearest Neighbor, Linear Discriminant Analysis, Logistic Regression) در شهر ماکاسار (اندونزی) پرداختند و به این نتیجه رسیدند مدل Random Forest بهترین عملکرد را در پیش‌بینی سیلاب‌های شهری نشان داده و از این رو قابل اطمینان‌ترین مدل در پردازش داده‌های پیچیده سیلاب شهری بوده است. بررسی منابع نشان می‌دهد که برای انتخاب بهترین الگوریتم دسته‌بندی، باید به دقت به شرایط و ویژگی‌های مسئله توجه کنیم (مانند نوع داده‌ها، میزان قدرت پردازش کامپیوتر و نیازهای عملکردی). هر الگوریتمی دارای نقاط ضعف و قوت است که ممکن است برای بعضی موارد مناسب باشد؛ ولی در برخی موارد دیگر اصلاً مناسب نباشد. بنابراین، برای گرفتن بهترین نتیجه، باید الگوریتمی را انتخاب کنیم که با شرایط پروژه و هدف موردنظر بیشترین هماهنگی را داشته باشد.

حوزه شهری اقلید به دلیل ترکیبی از عوامل طبیعی (بارندگی زیاد، شیب متوسط، ارتفاع زیاد) و همچنین فعالیت‌های انسانی (شهرنشینی سریع، کاهش مناطق آبخیز و زهکشی ناکافی) با چالش‌های پیچیده سیل بخصوص در بخش مرکزی حوضه روبرو است. از نظر دینامیکی، سیل یکی از فعال‌ترین و پیچیده‌ترین بلایای طبیعی محسوب می‌شود که جلوگیری کامل از آن ممکن نیست، اما با تهیه نقشه‌های حساسیت‌پذیری، خطر و ریسک، می‌توان شدت خسارات، اثرات منفی و میزان تلفات ناشی از آن را کاهش داد. در این راستا، شناسایی عوامل کلیدی مؤثر بر وقوع و افزایش شدت سیل و همچنین تهیه نقشه‌های حساسیت‌پذیری، از مهم‌ترین راه‌کارها برای کاهش مخاطرات ناشی از سیل به شمار می‌آید. هدف اصلی این مطالعه، استفاده از تکنیک‌های یادگیری ماشین برای تهیه نقشه پهنه‌بندی مناطق پرخطر سیل است و اهداف فرعی آن شناسایی مهم‌ترین عوامل مؤثر بر این پدیده می‌باشد. نوآوری این تحقیق در طراحی و پیاده‌سازی یک چارچوب یکپارچه و نوین است که منابع داده‌های متنوع را با الگوریتم‌های پیشرفته یادگیری ماشین ترکیب می‌کند. این چارچوب به‌طور خاص برای پیش‌بینی سیل‌های شهری طراحی شده و قادر است به‌صورت دقیق‌تری عوامل مؤثر در وقوع سیل را شبیه‌سازی کند.

روش پژوهش

در رابطه با پهنه‌بندی نقاط حساس به سیل، داده‌های مربوط به شیب، شاخص تابش خورشیدی، شاخص قدرت جریان، انحنا سطح، شاخص موقعیت توپوگرافی، فاصله از رودخانه، تراکم زهکشی، بارندگی سالانه، کاربری اراضی، پوشش گیاهی،

سنگ‌شناسی، نوع خاک و عوامل مرتبط دیگر (نقاط سیل و عدم وقوع سیل) جمع‌آوری و تحلیل می‌شوند تا احتمال وقوع سیل مشخص شود. روند اجرای این مطالعه در شکل ۱ نمایش داده شده است.



شکل ۱. روش شناسی تحقیق

آماده‌سازی نقشه نقاط سیلابی

تحلیل داده‌های تاریخی مربوط به رویدادهای سیلابی گذشته، ابزار کلیدی در پیش‌بینی احتمال وقوع سیلاب در مناطق مختلف محسوب می‌شود. به‌طوری‌که مناطق مجاور رویدادهای سیلابی گذشته حساسیت بالایی در برابر وقوع مجدد سیل از خود نشان می‌دهند (رحمتی و همکاران، ۲۰۱۶). در منطقه مورد مطالعه، با بهره‌گیری از فناوری GPS و انجام پیمایش‌های میدانی همراه با مشاهده مستقیم مناطق تحت سیلاب و استفاده از داده‌های آماری اداره منابع طبیعی استان فارس، تعداد ۵۰ موقعیت سیلابی شناسایی و ثبت گردید. علاوه بر این، به‌منظور تعیین ۱۰۰ نقطه غیرسیلابی، با بهره‌گیری از نقشه توپوگرافی، بازدیدهای میدانی و سامانه Google Earth، نقاطی در مناطق مرتفع مانند کوه‌ها و تپه‌ها که سیلاب توان نفوذ به آنها را ندارد، شناسایی گردید. در این پژوهش، به‌طور تصادفی ۷۰ درصد از نقاط جهت آموزش مدل‌ها و ۳۰ درصد باقیمانده برای ارزیابی دقت آنها اختصاص یافت؛ این تقسیم‌بندی، طبق استانداردهای رایج در مطالعات حساسیت‌سنجی مخاطرات طبیعی به‌عنوان نسبت بهینه توصیه شده است (Nachappa et al, 2020).

تهیه و تدوین لایه‌های اطلاعاتی منطقه مورد بررسی

مطابق با تحقیقات و دسترسی به داده‌ها، ۱۵ شاخص مستقل متشکل از پارامترهای هیدرولوژیکی، توپوگرافی، زمین‌شناسی و محیطی شامل ارتفاع، شیب، شاخص تابش خورشیدی، انحنای سطح، کاربری زمین، بارندگی، فاصله تا آبراهه و تراکم زهکشی، سنگ‌شناسی، نوع خاک، شاخص پوشش گیاهی (NDVI)، شاخص رطوبت توپوگرافی (Topographic Wetness Index)، شاخص موقعیت توپوگرافی (Topographic Position Index)، شاخص قدرت جریان (Stream Power Index) و شاخص ناهمواری زمین (Terrain Ruggedness Index) به‌عنوان متغیرهای پیش‌بینی حساسیت سیل در منطقه مطالعه شده تعیین شدند. که در ادامه ارتفاع، شاخص ناهمواری زمین و شاخص رطوبت توپوگرافی حذف

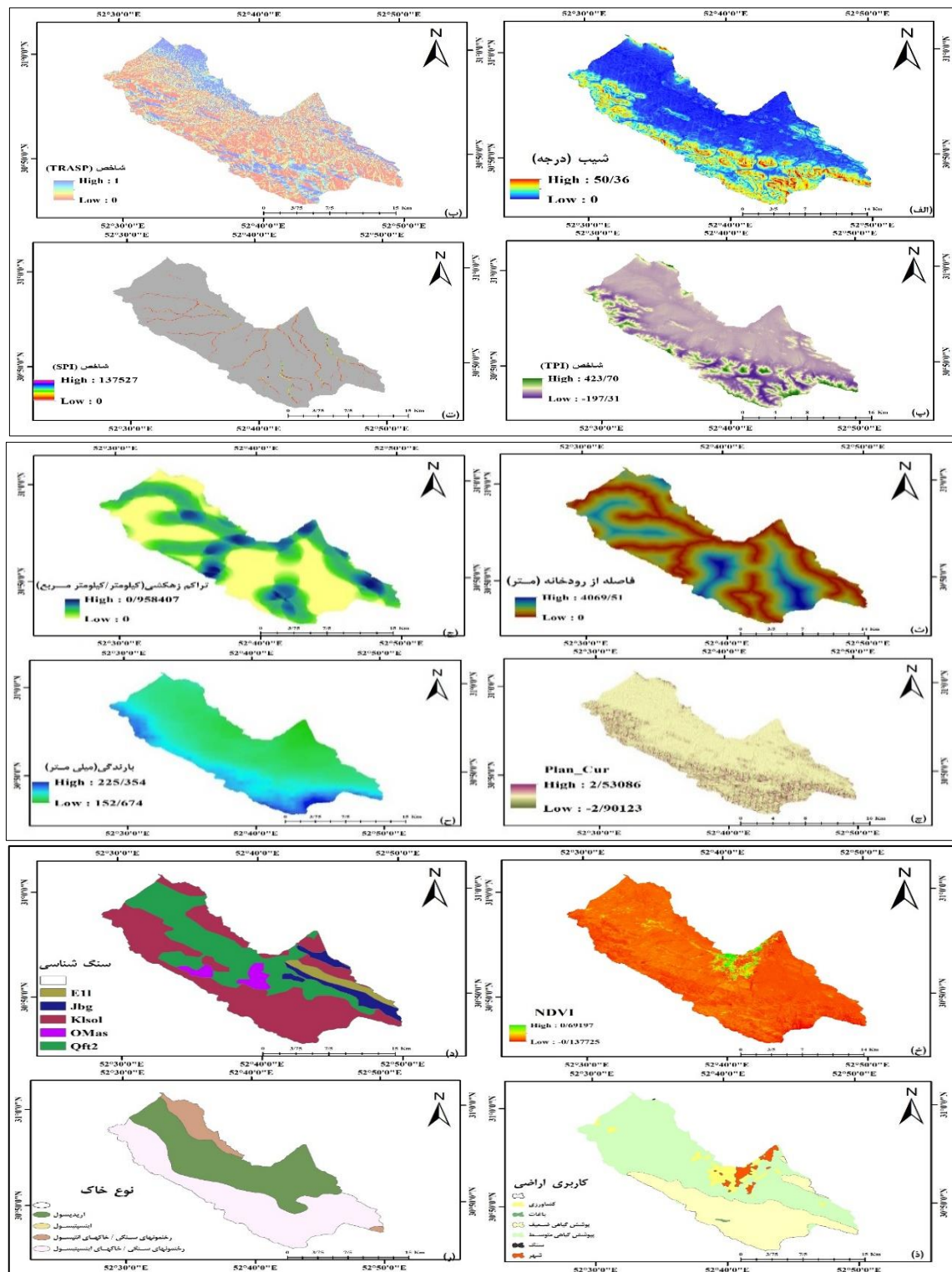
شدند. شیب: مناطقی که ارتفاع کمتری دارند، معمولاً شیب کمتری داشته و در معرض ریسک بالاتری از سیل قرار می‌گیرند؛ همچنین این نواحی قادرند رواناب‌ها را با سرعت بیشتری تخلیه کنند (تهرانی و کومار، ۲۰۱۸). لایه شیب منطقه مورد مطالعه با واحد درجه از مدل رقومی ارتفاعی با استفاده از تجزیه و تحلیل رستری زمین در نرم‌افزار ArcGISPro تهیه شد. شیب منطقه مورد مطالعه بین صفر تا ۵۰/۳۶ درجه است. شیب‌های کم مربوط به قسمت‌های شمالی نقشه و شیب کم مربوط به بخش‌های جنوبی و غربی منطقه مورد مطالعه است. شاخص تابش خورشیدی: جهت جغرافیایی با استفاده از رابطه زیر به شاخص تابش خورشیدی (TRASP) تبدیل گردید. مقدار شاخص تابش خورشیدی بین صفر و یک است و جهت شمال و شمال شرقی دارای مقدار صفر (خنک‌ترین دامنه) و جهت جنوب و جنوب غربی دارای مقدار ۱ (گرم‌ترین دامنه) است (علوی و همکاران، ۱۳۹۸).

$$\text{TRASP} = (1 - \cos((\pi/180) * (\text{Aspect} - 30))) / 2 \quad \text{رابطه ۱}$$

در منطقه مورد مطالعه مقادیر پایین شاخص TRASP در نقاطی که سیل رخ داده است، نشان‌دهنده جهت‌گیری سطوح زمین به گونه‌ای است که تابش خورشیدی کمتری دریافت می‌کنند. این شرایط ممکن است با تأثیرگذاری بر عوامل محیطی مانند پوشش گیاهی، تبخیر و نفوذ آب در خاک، نقش غیرمستقیمی در افزایش احتمال وقوع سیل ایفا کند. **انحنای سطح:** باتوجه به اینکه انحنای نمایانگر شکل سطح زمین و توانایی آن در نگهداری آب است، احتمال بروز سیل با افزایش انحنای کاهش می‌یابد (Costache, 2019). این لایه اطلاعاتی با استفاده از داده‌های ارتفاعی و ابزار Curvature موجود در نرم‌افزار Arc GIS PRO با وضوح ۱۲/۵ متر به دست آمده است. اصولاً در منطقه مورد مطالعه سیل در بخش‌های که انحنای به صورت مسطح است رخ داده است و این پارامتر در منطقه بین ۲/۹۰۱۲۳- تا ۲/۵۳۰۸۶ متغیر است. **کاربری اراضی:** کاربری زمین جزء متغیرهای مهم تأثیرگذار بر سیل است؛ پژوهش (Apollonio et al, 2016) تأکید می‌کند که تغییرات در کاربری زمین با افزایش خطر سیل، خصوصاً در نواحی که عایق رطوبتی ناشی از گسترش شهرنشینی افزایش یافته، ارتباط قابل توجهی دارد. برای استخراج لایه رستری کاربری اراضی منطقه مورد مطالعه، تصاویر ماهواره‌ای لندست ۸ به کار گرفته شد. در منطقه مورد مطالعه ۶ کاربری اراضی (کشاورزی، باغات، پوشش گیاهی ضعیف، پوشش گیاهی متوسط، شهری و سنگ) با استفاده از طبقه بندی نظارت شده (الگوریتم Random Forest) شناسایی شد. بیشتر رخداد سیل در منطقه مورد مطالعه در کاربری‌های کشاورزی، شهری و پوشش گیاهی متوسط رخ داده است. **بارندگی:** میزان بارندگی ارتباط مستقیمی با احتمال وقوع سیل دارد (گنجی راد و دلواری، ۲۰۲۲). با بهره‌گیری از داده‌های یک ایستگاه سینوپتیک (اقلید) و دو ایستگاه باران‌سنجی (اقلید و تیماران) (سال‌های ۱۴۰۰-۱۳۷۳)، نقشه بارندگی به وسیله روش درون‌یابی IDW در بستر ArcGIS Pro تهیه شد. میزان بارندگی در منطقه مورد مطالعه بین ۱۵۲/۶۷۴ تا ۲۲۵/۳۵۴ میلی‌متر متغیر است. سیل اصولاً در قسمت‌های که بارندگی کمتری رخ داده است به وقوع پیوسته است. **فاصله از آبراهه و تراکم زهکشی:** تراکم خطوط آبی و فاصله کم از آبراهه‌ها تأثیر مستقیم و چشمگیری بر زمان تمرکز و بزرگی جریان سیلاب دارد، به گونه‌ای که با افزایش تراکم زهکشی، احتمال وقوع سیل‌های شدیدتر افزایش می‌یابد (Pallard et al, 2009). همچنین، مناطق پیرامون رودخانه‌ها با خطر بیشتری از سیل مواجه هستند. این دو لایه از لایه رستری DEM تهیه شد. فاصله از آبراهه در منطقه مورد مطالعه بین صفر تا ۴۰۶۹/۵۱ متر متغیر است و سیل اصولاً در فاصله کمی از آبراهه‌ها رخ داده است. در رابطه با تراکم زهکشی این پارامتر بین صفر تا ۰/۹۵۸۴۰۷ کیلومتر در کیلومتر مربع رخ داده است. اصولاً سیل در قسمت‌های از منطقه که دارای تراکم زهکشی بالایی است رخ داده است. **سنگ‌شناسی:** تفاوت‌های چشمگیر در ناپایداری دامنه‌ها ناشی از تنوع واحدهای سنگ‌شناسی است. با استفاده از نقشه زمین‌شناسی کشور به مقیاس ۱:۱۰۰۰۰۰، نقشه سنگ‌شناسی

منطقه مطالعه شده در قالب ۵ کلاس مجزا (Ell, Jbg, Klsol, OMas و Qft2) تهیه شد. در منطقه مورد بررسی، بیشترین تعداد سیلاب‌ها مربوط به سازندهای Qft2 (دوره کوآترنری، دوران سنوزوئیک) و Klsol (دوره کرتاسه، دوران مزوزوئیک) است. در رسوبات کم‌ارتفاع فن کوهپایه‌ای و تراس‌های دره‌ای (Qft2) با وجود نفوذپذیری مناسب این سازندها، واقع شدن آن‌ها در نواحی کم‌شیب و کم‌ارتفاع باعث شده تا آب‌های سطحی در این مناطق بیشتر تجمع پیدا کند و در نتیجه سیل‌گرفتگی شدت یابد. همچنین سازند سنگ آهک ضخیم بستر (Klsol) معمولاً نفوذناپذیر است و جریان سطحی را افزایش می‌دهد، اما شکستگی‌ها و ناهمگنی‌های آن توانسته است مسیرهای سیلابی ایجاد کند که نقش مهمی در سیل حوضه دارد. در رابطه با شاخص پوشش گیاهی نرمال شده (NDVI): مقادیر NDVI منفی به آب اختصاص دارد و مقادیر مثبت نمایانگر حضور پوشش گیاهی هستند. همچنین NDVI رابطه منفی با وقوع سیل دارد؛ به این معنا که هر چه مقدار NDVI بیشتر باشد، احتمال بروز سیل کمتر است. این شاخص از باند ۴ و ۵ تصاویر لندست ۸ استخراج شد. مقدار NDVI در منطقه مورد مطالعه بین ۱۳۷۷۲۵- تا ۰/۶۹۱۹۷ متغیر است. در مناطقی که پوشش گیاهی وجود ندارد یا پوشش گیاهی ضعیف است به دلیل کاهش جذب و نفوذ بیشترین سیل‌گیری اتفاق افتاده است. این در حالی است که قسمت‌های از منطقه مورد مطالعه که دارای پوشش گیاهی خوب و نرمالی است بازهم سیل رخ داده است که باید سایر عوامل مورد بررسی قرار گیرد. زیرا پوشش گیاهی خوب باعث افزایش زمان نفوذ آب در خاک و کاهش رواناب سطحی می‌شود که منجر به کاهش سیلاب می‌گردد. نوع خاک: خاک یکی از اجزای اساسی در تنظیم نفوذپذیری، میزان جریان رواناب و در نتیجه بروز رخدادهای طبیعی خطرناک به شمار می‌رود (Kanani-Sadat et al, 2019). این لایه توسط اداره مرکز تحقیقات، آموزش و ترویج کشاورزی و منابع طبیعی استان فارس تأمین گردید. این لایه شامل ۴ کلاس (اریدیسول، اینسپتی‌سول، رخنمون-های سنگی/خاک‌های انتیسول، رخنمون‌های سنگی/خاک‌های اینسپتی‌سول) است. از بین این ۴ کلاس بیشتر رخداد سیل در خاک‌های اریدیسول که از نظر داشتن ماده عالی بسیار فقیر و جز خاک‌های مناطق بیابانی و خشک و حاصل رسوبات ثانویه است، اتفاق افتاده است. شاخص قدرت جریان (Stream Power Index): شاخص توان آبراهه، نمایانگر ترکیبی از نیم‌رخ عرضی و طولی آبراهه است که با مقایسه نسبت مساحت هر نقطه به شیب زمین بالادستش، می‌تواند مناطق تجمع آب و رسوب‌گذاری احتمالی در حوضه آبخیز را شناسایی کند؛ این شاخص ابزاری کلیدی در تحلیل ساختار و عملکرد جریان‌های آبی است (شهبابی و همکاران، ۲۰۱۴). مقدار این شاخص در منطقه مورد مطالعه بین صفر تا ۱۳۷۵۲۷ متغیر است. فرمول SPI به صورت ساده نشان می‌دهد که هر چه شیب زمین بیشتر و مساحت حوضه آبریز بزرگ‌تر باشد، توان جریان آب بالاتر است و به تبع آن، قدرت فرسایش و تخریب توسط جریان آب بیشتر خواهد بود. در منطقه مورد مطالعه با افزایش مقدار SPI، حساسیت سیل کاهش یافته است. یکی از دلایل این امر این است که افزایش شاخص قدرت جریان اغلب به افزایش عرض آبراهه‌ها منجر می‌شود که می‌تواند عمق جریان را کاهش دهد و از شدت سیل بکاهد. شاخص موقعیت توپوگرافی (Topographic Position Index): شاخص موقعیت توپوگرافی به طور تدریجی و دقیق، برای سنجش موقعیت شیب توپوگرافی به کار می‌رود و با مقایسه ارتفاع هر سلول نسبت به میانگین ارتفاع سلول‌های همجوار در یک ناحیه مشخص، نقش کلیدی در تحلیل توپوگرافی ایفا می‌کند (Derue et al, 2011). این شاخص مقایسه موقعیت شیب توپوگرافی با تعیین ارتفاع هر سلول نسبت به میانگین ارتفاع سلول‌های همجوار، باعث شناسایی مناطق با شیب تند و ناهموار می‌شود که این مناطق با افزایش سرعت جریان آب در بارندگی شدید، خطر سیل و فرسایش را افزایش می‌دهند. به عبارت دیگر، این الگوریتم با شناسایی نقاط حساس به تجمع و حرکت سریع آب، نقش مهمی در پیش‌بینی و مدیریت سیل دارد. در منطقه مورد مطالعه این شاخص بین ۱۹۷/۳۱- تا ۴۲۳/۷۰ متغیر است. مقادیر پایین نشان‌دهنده‌ی دره یا گودال است معمولاً این نقاط محل تجمع جریان‌های آب و رواناب هستند و احتمال وقوع سیل در این مناطق بیشتر

است. این در حالی است که مقادیر بالای این شاخص معمولاً نشان‌دهنده نقاط مرتفع مانند تپه‌ها یا قله‌هاست که آب از آنها به سمت پایین جریان می‌یابد و احتمال تجمع آب و سیل در آن‌ها کمتر است (شکل ۲).



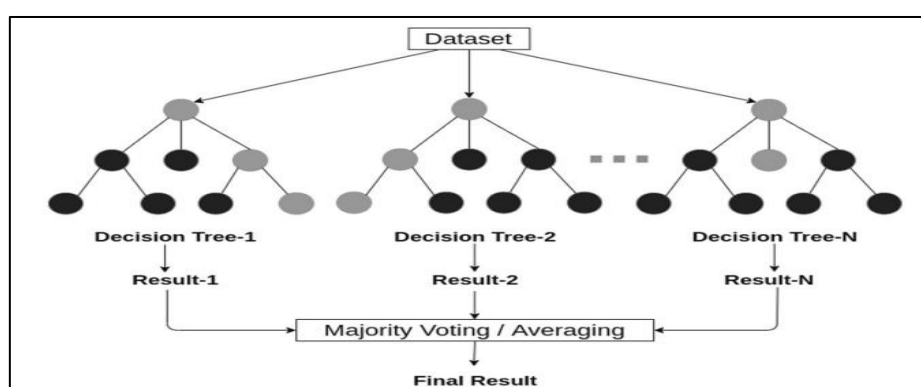
شکل ۲. مجموعه‌ای از لایه‌های تأثیرگذار در سیلاب (الف) شیب، (ب) شاخص Trap، (پ) شاخص TPI، (پ) شاخص SPI (ث) فاصله از رودخانه، (ج) تراکم زهکشی، (چ) انحنای سطح، (ح) بارندگی، (خ) پوشش گیاهی، (د) سنگ‌شناسی، (ذ) کاربری اراضی و (ر) نوع خاک

مدل‌های یادگیری ماشین

تحقیقات گذشته رویکردهای مختلفی برای پیش‌بینی سیل ارائه داده‌اند که در میان آنها، یادگیری ماشینی (ML) به‌عنوان یک راه‌حل امیدوارکننده مطرح شده است. مطالعات اخیر نشان داده‌اند که الگوریتم‌های ML می‌توانند به طور مؤثر به پیش‌بینی دقیق سیل بپردازد (Schmidt et al, 2020).

الف- جنگل تصادفی

الگوریتم جنگل تصادفی یک الگوریتم یادگیری ماشین است که برای وظایف طبقه‌بندی و رگرسیون استفاده می‌شود و اولین بار توسط Breiman (۲۰۰۱) مطرح شد. هر درخت بر روی یک نمونه از داده‌های آموزشی آموزش داده می‌شود. سپس، اگر هدف طبقه‌بندی باشد، پیش‌بینی با رأی اکثریت درختان انجام می‌شود. اگر هدف رگرسیون باشد، میانگین خروجی درختان به دست می‌آید (Svetnik et al, 2004) (شکل ۳).

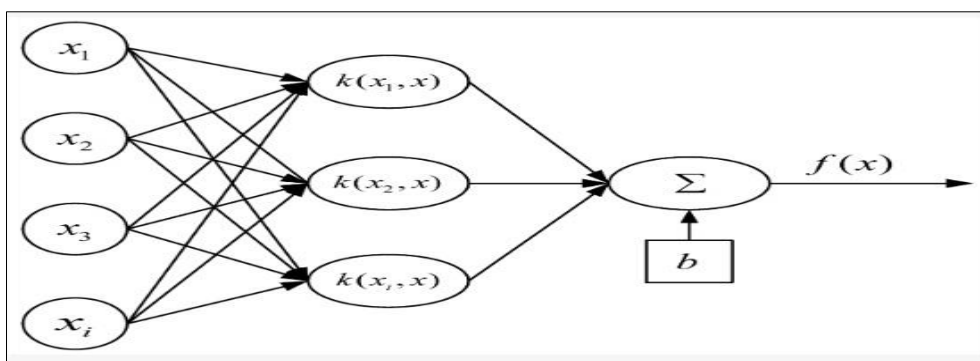


شکل ۳. ساختار کلی مدل RF (Sarker, 2021)

الگوریتم جنگل تصادفی می‌تواند با تجزیه و تحلیل منابع داده‌های مختلف (شیب، شاخص تابش خورشیدی، فاصله از رودخانه، تراکم زهکشی، شاخص قدرت جریان، شاخص موقعیت توپوگرافی، شاخص پوشش گیاهی، سنگ‌شناسی، خاک، کاربری اراضی، بارندگی، انحنای سطح) در پیش‌بینی سیل مورد استفاده قرار گیرد. این الگوریتم با ایجاد چندین درخت تصمیم‌گیری که هر کدام بر اساس زیرمجموعه‌ای از داده‌های موجود هستند، کار می‌کند. درخت‌های تصمیم‌گیری با انتخاب تصادفی ویژگی‌ها و نمونه‌ها از مجموعه داده‌ها ایجاد می‌شوند. پیش‌بینی نهایی با ترکیب پیش‌بینی‌های همه درخت‌های تصمیم‌گیری انجام می‌شود (Hasti et al, 2009). همچنین قابلیت مدیریت داده‌های از دست‌رفته و داده‌های پرت را که می‌توانند در مجموعه داده‌های دنیای واقعی رایج باشند، دارد.

ب- ماشین بردار پشتیبان

ماشین بردار پشتیبان (SVM) یک الگوریتم یادگیری ماشین تحت نظارت است که در مدل‌سازی (طبقه‌بندی و رگرسیون) استفاده می‌شود (Choubin et al, 2019). این مدل بر اساس اصل بهینه‌سازی است و سعی می‌کند یک ابر صفحه را روی مجموعه داده‌های آموزشی برای جداسازی کلاس‌های مختلف برآزش دهد. ابر صفحه به گونه‌ای جهت‌گیری شده است که تا حد امکان از نزدیک‌ترین نقاط داده از هر یک از کلاس‌ها دور باشد. این نزدیک‌ترین نقاط، بردارهای پشتیبان نامیده می‌شوند (Koggalage & Halgamuge, 2004). یک ابر صفحه همچنین به عنوان مرز تصمیم‌گیری شناخته می‌شود. طبق شکل (۴)، هسته مورد نظر روی ویژگی‌ها پیاده‌سازی می‌شود و در نهایت طبقه‌بندی با اعمال سیگما و بایاس روی خروجی انجام می‌شود.



شکل ۴. ساختار کلی مدل SVM (Xu et al, 2021)

همان‌طور که در شکل (۴) نشان داده شده است، x_i ویژگی ورودی، k هسته و نتیجه نهایی نقشه خطر سیل است. به‌طور کلی، ماشین بردار پشتیبان با دریافت ورودی‌های متنوع (هیدرولوژیکی، توپوگرافی و اقلیمی و ...) قادر است شرایط منطقه مورد مطالعه را تحلیل کرده و خطر سیل را پیش‌بینی نماید. این توانمندی باعث شده از SVM در مدل‌سازی‌های مختلف پیش‌بینی سیل و مدیریت خطر سیل استفاده شود.

در نهایت می‌توان گفت مدل‌های جنگل تصادفی (RF) و ماشین بردار پشتیبان (SVM) قابلیت‌های برتر خود را در پیش‌بینی خطر سیل با دقت بالا را دارند (Yu et al, 2023). مدل جنگل تصادفی از طریق رویکرد گروهی خود عملکرد برتر خود را در مدیریت مجموعه داده‌های پیچیده و بدون ساختار نشان داده است (Zhang & Zhu, 2022). در حالی که ماشین بردار پشتیبان در مدیریت داده‌های نامتعادل و کاهش نویز قدرت خود را نشان داده است (Qian et al, 2021).

پس از طراحی مدل، گام اساسی بعدی، ارزیابی دقت و کارایی آن است. فرض کنیم مدلی برای پیش‌بینی احتمال وقوع ریسک توسعه یافته است؛ یکی از پرسش‌های کلیدی در این زمینه، نحوه سنجش عملکرد مدل ساخته شده است. در این مرحله، لازم است صحت پیش‌بینی‌های مدل به دقت بررسی شود؛ به عبارت دیگر، باید مشخص شود که مدل تا چه اندازه قابلیت پیش‌بینی صحیح وقوع ریسک را در برابر داده‌ها و شرایطی که در زمان آموزش با آن‌ها مواجه نشده است، دارد. معمولاً چندین مدل متنوع توسعه می‌یابد و چالش اساسی انتخاب بهترین مدل از میان آن‌ها برای پیش‌بینی دقیق‌تر است. به طور معمول، مدل‌ها بر روی داده‌های آموزشی ساخته شده و سپس با استفاده از داده‌های آزمایشی، مورد ارزیابی عملکرد قرار می‌گیرند. معیارهای مختلفی مانند دقت (Accuracy)، حساسیت (Sensitivity)، ویژگی (Specificity)، شاخص کاپا (Kappa) و سطح زیر منحنی ROC (AUC) به عنوان معیارهای استاندارد برای سنجش قدرت پیش‌بینی مدل‌ها به کار گرفته شد (رضایی مقدم و رحیم‌پور، ۱۴۰۲).

اطلاعات مربوط به معیارهای سنجش در قالب ماتریس درهم‌ریختگی (Confusion Matrix) در جدول شماره ۱ ارائه شده است. این ماتریس روابط بین نتایج واقعی و پیش‌بینی‌های مدل را نشان می‌دهد و امکان ارزیابی دقیق کارایی مدل را فراهم می‌آورد. در این جدول (True Positive)، تعداد نمونه‌هایی است که مدل به درستی مثبت شناسایی کرده است (True Negative)، نمونه‌های منفی صحیح (False Positive)، نمونه‌های منفی که اشتباهاً مثبت شناسایی شده‌اند، و (False Negative) نمونه‌های مثبت که به اشتباه منفی طبقه‌بندی شده‌اند. این چهار معیار پایه‌ای برای ارزیابی عملکرد مدل‌های طبقه‌بندی هستند (نادری و همکاران، ۱۴۰۴).

جدول ۱. ماتریس در هم ریختگی

برچسب پیش‌بینی		
برچسب واقعی	منفی (۰)	مثبت (۱)
منفی (۰)	TN	FN
مثبت (۱)	FP	TP

معیار دقت (Accuracy) درصد نمونه‌هایی است که مدل به‌درستی شناسایی کرده است. روش محاسبه این معیار در رابطه شماره ۲ ارائه شده است.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad \text{رابطه ۲}$$

معیار حساسیت (Sensitivity) که به نرخ مثبت درست نیز شناخته می‌شود، نشان‌دهنده درصد نمونه‌های مثبتی است که مدل توانسته به‌درستی آن‌ها را پیش‌بینی کند. روش محاسبه این معیار در رابطه شماره ۳ ارائه شده است.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad \text{رابطه ۳}$$

معیار ویژگی (Specificity) نشان‌دهنده درصد نمونه‌های منفی است که مدل به‌درستی تشخیص داده است. روش محاسبه این معیار در رابطه شماره ۴ ارائه شده است.

$$\text{Specificity} = \frac{TN}{FP + TN} \quad \text{رابطه ۴}$$

شاخص کاپا (Kappa)

ضریب کاپا میزان توافق بین مقادیر پیش‌بینی شده و واقعی را نشان می‌دهد و. مقادیر این ضریب بین +۱ تا -۱ متغیر است. تفسیر این ضریب به این صورت است که مقادیر کمتر از ۰/۴ ضعیف، بین ۰/۴ تا ۰/۵۵ قابل قبول، ۰/۵۵ تا ۰/۷ خوب، ۰/۷ تا ۰/۸۵ بسیار خوب و بالاتر از ۰/۸۵ عالی است

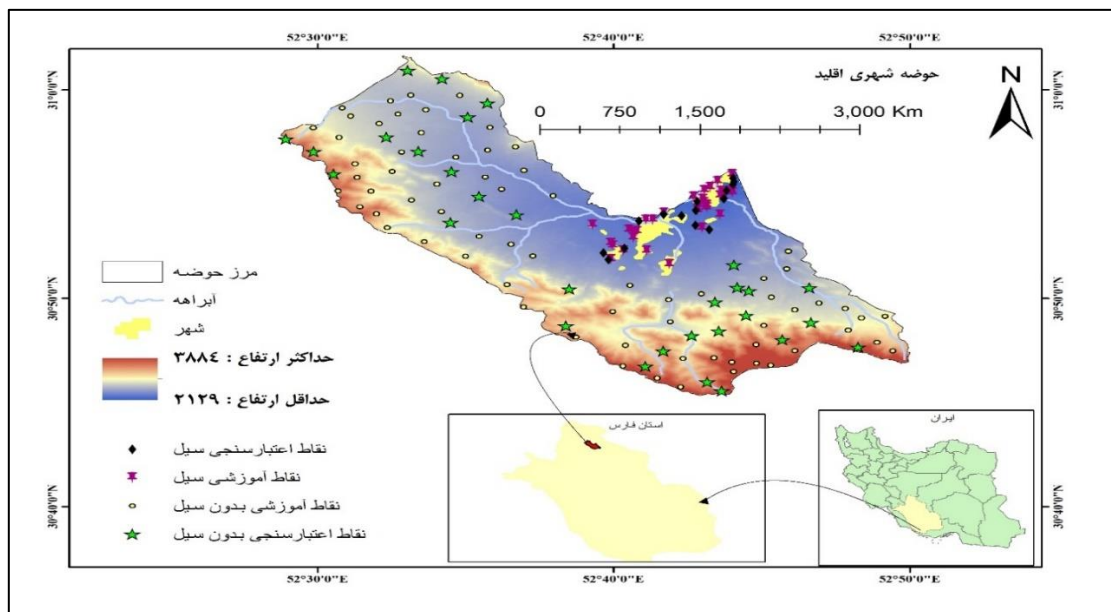
سطح زیر منحنی (AUC)

برای سنجش کارایی مدل از شاخص مساحت زیر منحنی (AUC) در چارچوب تحلیل منحنی مشخصه عملکرد گیرنده (ROC) بهره گرفته شد. مقدار AUC نشان‌دهنده توان مدل در تمایز بین کلاس‌های مختلف است و دقت عملکرد مدل بر اساس مقادیر زیر طبقه‌بندی می‌شود: ۰/۵ تا ۰/۶ به‌عنوان ضعیف، ۰/۶ تا ۰/۷ متوسط، ۰/۷ تا ۰/۸ خوب، ۰/۸ تا ۰/۹ بسیار خوب و ۰/۹ تا ۱ عالی تلقی می‌گردد. این معیار به شکلی مؤثر عملکرد مدل را در تفکیک نمونه‌های مثبت و منفی اندازه‌گیری کرده و بهترین ابزار برای قضاوت در مورد قابلیت پیش‌بینی مدل محسوب می‌شود (عرب عامری، ۲۰۱۹).

محدوده مورد مطالعه

حوزه شهری اقلید در شرق استان فارس که از جنوب به شهرستان‌های مرو دشت و سپیدان، از شرق به شهرستان خرمبید و از غرب به استان‌های اصفهان و کهگیلویه و بویراحمد محدود شده است. این حوزه در گستره جغرافیایی ۵۲ درجه، ۲۴ دقیقه و صفر ثانیه تا ۵۲ درجه، ۵۲ دقیقه و صفر ثانیه طول شرقی و ۳۰ درجه، ۳۴ دقیقه و ۴۰ ثانیه تا ۳۱ درجه، ۴ دقیقه و ۲۰ ثانیه عرض شمالی قرار گرفته است. حوزه مورد مطالعه به‌طور کلی در حدود ۴۰/۶۲۳ هکتار مساحت، ارتفاع متوسط ۲۶۱۹/۹ متر و شیب متوسط ۱۶/۴ درصد است که جزء حوضه‌های مرتفع و با شیب متوسط محسوب می‌گردد. تغییرات ارتفاعی منطقه از ۱۲۳۴ تا ۳۲۱۳ متر بالاتر از سطح دریا گسترش یافته است. ارتفاعات بالاتر در سمت غرب منطقه واقع

شده است و قسمت‌های مرکزی منطقه مورد مطالعه ارتفاع کمتری دارند و احتمال سیل در این مناطق بیشتر است. حوزه شهری اقلید دارای چندین خروجی و آبراهه اصلی است که به شهر اقلید ختم می‌شود. مکان جغرافیایی منطقه مورد مطالعه در شکل (۵) نمایش داده شده است.



شکل ۵. موقعیت جغرافیایی منطقه مورد مطالعه در استان فارس و ایران

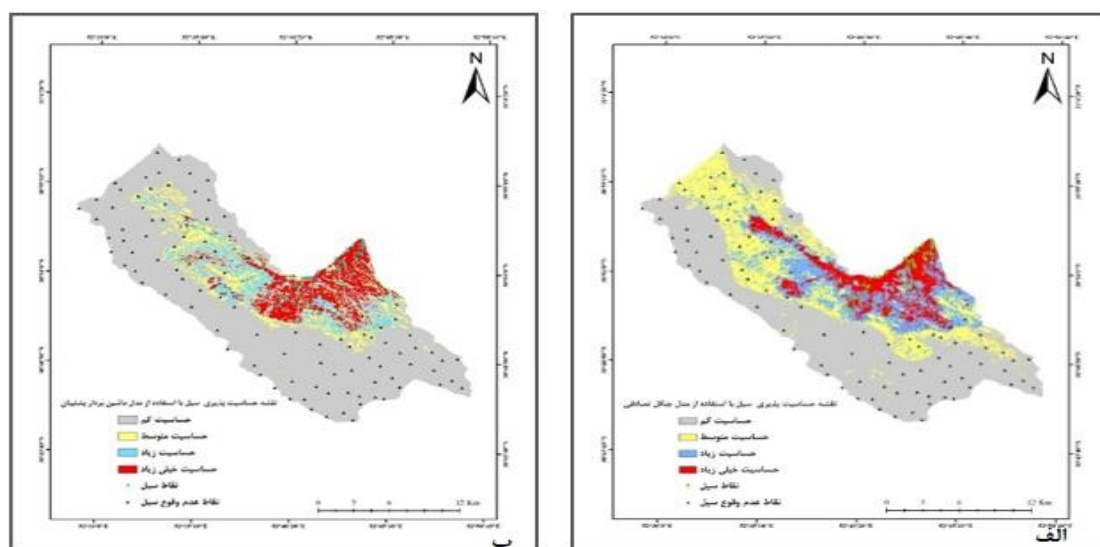
یافته‌ها

قبل از به‌کارگیری متغیرها بر گسترش سیل، لازم است آزمون هم‌خطی چندگانه بین پارامترهای مورد استفاده انجام گیرد. هم‌خطی بیان‌کننده این موضوع است که آیا بین متغیرهای مستقل همبستگی وجود دارد یا خیر (Bui et al, 2015). بدین منظور از شاخص فاکتور تورم واریانس (Variance Inflation Factor) یا VIF استفاده شد. با بررسی آزمون هم‌خطی چندگانه (VIF) بین متغیرهای مستقل سه متغیر ارتفاع، شاخص رطوبت توپوگرافی و شاخص ناهمواری زمین به دلیل $VIF > 10$ از ادامه روند کار حذف شدند. سپس، با بهره‌گیری از مجموعه داده‌های آموزشی تکمیل‌شده، مدل‌های ماشین بردار پشتیبان (SVM) و جنگل تصادفی (RF) با استفاده از زبان برنامه‌نویسی R پیاده‌سازی شدند. مسئله مورد بررسی به شکل یک مسئله رگرسیونی تعریف شده است و هدف نهایی، برآورد احتمال وقوع سیل در هر یک از پیکسل‌های منطقه مورد مطالعه است. با استفاده از روش شکست طبیعی هر کدام از نقشه‌های نهایی در ۴ دسته طبقه‌بندی شدند. داده‌ها به نسبت ۷۰ درصد برای آموزش مدل و ۳۰ درصد برای ارزیابی عملکرد آن تقسیم‌بندی شده‌اند.

نتایج حاصل از الگوریتم جنگل تصادفی و ماشین بردار پشتیبان

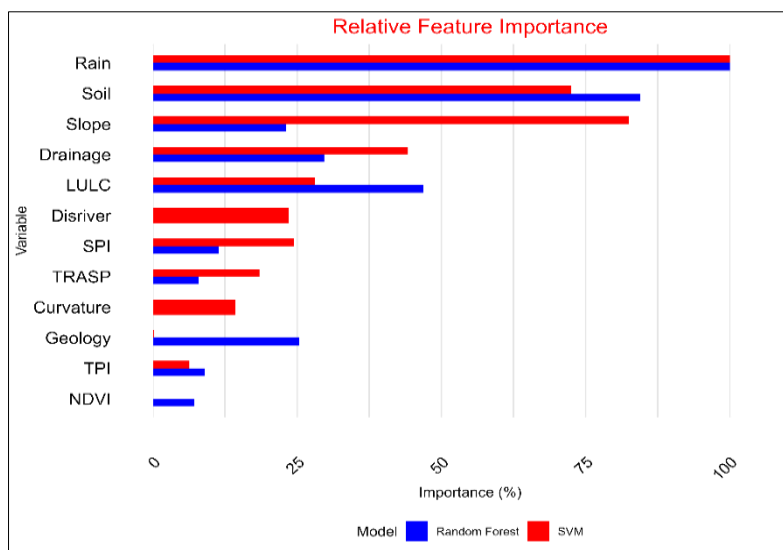
نقشه‌های نهایی به ۴ کلاس شامل مناطق با خطر سیل کم، متوسط، زیاد و بسیار زیاد طبقه‌بندی شدند (شکل ۶). بر اساس نتایج بدست آمده از مدل جنگل تصادفی، ۲۰۹/۰۱۲۴ کیلومتر مربع (۵۲/۷۳ درصد) از مساحت حوزه دارای حساسیت کم، ۸۸/۷۵۹۸ کیلومتر مربع (۲۲/۳۹ درصد) از مساحت حوزه دارای حساسیت متوسط، ۶۲/۱۰۲۷ کیلومتر مربع (۱۵/۶۶ درصد) از مساحت حوزه دارای حساسیت زیاد و ۳۶/۴۹۸۶ کیلومتر مربع (۹/۲۰ درصد) از مساحت حوزه دارای حساسیت خیلی زیاد نسبت به سیل است. بیشترین مساحت حساسیت سیل مربوط به طبقه حساسیت کم است (شکل ۶).

الف). این در حالی است که در مدل ماشین بردار پشتیبان $286/1892$ کیلومتر مربع ($72/20$ درصد) از مساحت حوضه دارای حساسیت کم، $38/3211$ کیلومتر مربع ($9/66$ درصد) دارای حساسیت متوسط، $33/7041$ کیلومتر مربع ($8/50$ درصد) دارای حساسیت زیاد و $38/1591$ کیلومتر مربع ($9/62$ درصد) دارای حساسیت خیلی زیاد است. بیشترین مساحت حساسیت در این مدل مربوط به طبقه حساسیت کم است (شکل ۶ ب). نتایج نشان می‌دهد که هر دو مدل به یکسان توانسته است مناطق با خطر بسیار زیاد را بر اساس ویژگی‌های آن‌ها در خطر سیل تشخیص دهد. به عبارت دیگر در حوضه مورد مطالعه سیل در قسمت مرکزی که دارای شیب کم، شاخص تابش خورشیدی پایین، دارای انحنای مسطح، کاربری اراضی شهری-کشاورزی و پوشش گیاهی متوسط، فاصله کم از آبراهه، تراکم بالای زهکشی، در رسوبات کواترنری و سنگ آهک کرتاسه، خاک اریدیسول، بالا بودن مقدار شاخص قدرت جریان و از نظر شاخص موقعیت توپوگرافی دره یا گودال بوده است، رخ داده است.



شکل ۶. الف) نقشه حساسیت به سیلاب با استفاده از مدل جنگل تصادفی ب) نقشه حساسیت به سیلاب با استفاده از مدل ماشین بردار پشتیبان

ارزیابی دیگری که می‌توان در مدل RF و SVM انجام داد، تعیین اهمیت نسبی عوامل (ویژگی‌های) مؤثر به کاررفته است. اهمیت ویژگی‌ها با استفاده از تحلیل حساسیت که امتیازی را برای همه ویژگی‌های ورودی محاسبه می‌کند، تعیین می‌شود. هر ویژگی با امتیاز بالاتر به این معنی است که آن ویژگی تأثیر بیشتری در پیش‌بینی خطر سیل دارد. در مدل جنگل تصادفی به ترتیب متغیرهای بارندگی (100 درصد)، خاک ($84/4$ درصد)، کاربری اراضی ($46/9$) و تراکم زهکشی ($29/7$ درصد) بیشترین تأثیر دارد این در حالی است که انحنای سطح و فاصله از رودخانه هیچ اهمیتی ندارد. در مدل ماشین بردار پشتیبان به ترتیب متغیرهای بارندگی (100 درصد)، شیب ($82/48$)، خاک ($72/48$) و تراکم زهکشی ($44/21$ درصد) بیشترین تأثیر را دارد در حالی که شاخص پوشش گیاهی هیچ تأثیری ندارد (شکل ۷).



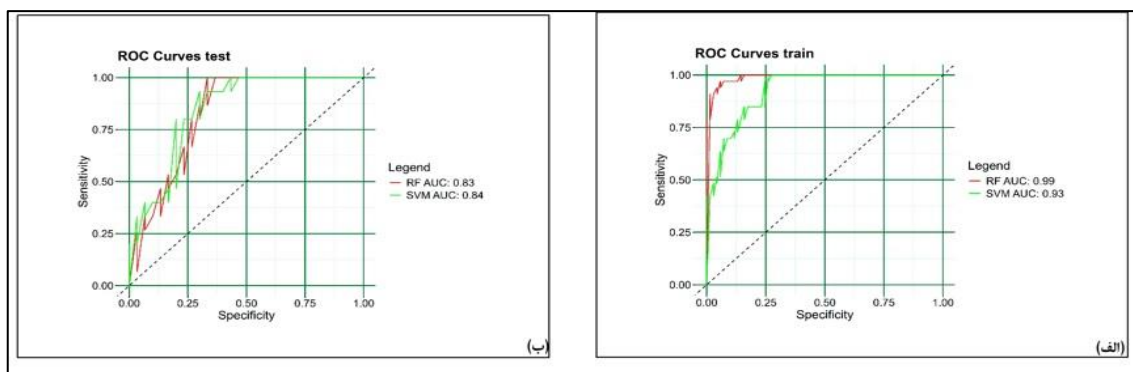
شکل ۷. اهمیت نسبی متغیرها

در ادامه برای ارزیابی مدل‌ها در فرایندهای آموزشی و آزمایشی از شاخص‌های آماری از قبیل Sensitivity, Specificity, Accuracy, Kappa و AUC استفاده شد. نتایج حاصل Sensitivity برای داده‌های آموزشی ($RF=0/93$ و $SVM=0/84$) و برای داده‌های اعتبارسنجی ($RF=0/86$ و $SVM=0/80$) به دست آمد. مقدار Specificity برای داده‌های آموزشی ($RF=0/94$ و $SVM=0/83$) و برای داده‌های اعتبارسنجی ($RF=0/86$ و $SVM=0/73$) به دست آمد. نتایج حاصل از مقدار Accuracy برای مقدار آموزشی ($RF=0/94$ و $SVM=0/89$) و برای داده‌های اعتبارسنجی ($RF=0/73$ و $SVM=0/77$) به دست آمد. در نهایت مقدار Kappa برای داده‌های آموزشی ($RF=0/86$ و $SVM=0/73$) و برای داده‌های اعتبارسنجی ($RF=0/75$ و $SVM=0/63$) به دست آمد. نتایج این پژوهش نشان داد که هر دو مدل، عملکرد قابل قبولی در تهیه نقشه‌های پتانسیل خطر وقوع سیل در حوزه مورد مطالعه دارند و دقت نقشه‌های تولید شده توسط این دو مدل از نظر کمی و کیفی عملکرد بالایی دارند (جدول ۲).

جدول ۲. ارزیابی آماری مدل‌های مورد استفاده

مدل	مدل FR		مدل SVM	
	شاخص آماری	داده‌های آموزشی	داده‌های آموزشی	داده‌های اعتبارسنجی
True positive		۳۱	۷	۲۸
True negative		۶۹	۲۶	۵۷
False positive		۴	۴	۱۲
False negative		۲	۸	۵
Sensitivity(%)		0/93	0/86	0/84
Specificity(%)		0/94	0/86	0/83
Accuracy(%)		0/94	0/73	0/89
Kappa(%)		0/86	0/75	0/73

در شکل (۸) منحنی ROC را برای مجموعه داده‌های آموزشی و آزمایشی مدل‌ها نشان داده شده است. مقدار سطح زیر منحنی برای داده‌های آموزش ($RF=0/99$ و $SVM=0/93$) و برای داده‌های آزمایشی ($RF=0/83$ و $SVM=0/84$) به دست آمد. این نتایج نشان می‌دهد که هر دو مدل عملکرد بالایی در پیش بینی سیل در منطقه مورد مطالعه داشته‌اند.



شکل ۸. منحنی ROC (الف) داده‌های آموزشی (ب) داده‌های اعتبارسنجی

بحث

شناسایی عوامل مؤثر بر سیل و تهیه نقشه خطر، یکی از راهکارهای اساسی برای کنترل این خطر طبیعی است. برای تخمین شدت خطر سیل، روش‌های متعددی توسط محققان مختلف ارائه شده است. در این مطالعه، با پیاده‌سازی مدل‌های یادگیری ماشین شامل RF و SVM، نقشه‌های خطر سیل در حوزه شهری اقلید تهیه شده است که نشان می‌دهد بیشتر خطر سیل در این حوزه مربوط به بخش مرکزی است.

بررسی اهمیت نسبی متغیرها نشان داد در هر دو مدل متغیرهای بارندگی، خاک، کاربری اراضی، شیب و تراکم زهکشی بیشترین تأثیر را در مدل‌سازی داشته است.

در حوزه شهری اقلید، میزان بارندگی در مناطق کم‌ارتفاع کمتر بوده، درحالی‌که مناطق مرتفع‌تر با بارندگی‌های بیشتری مواجه شده‌اند، این تفاوت توزیع بارش، همراه با شیب منطقه، می‌تواند موجب افزایش رواناب در ارتفاعات و تسریع وقوع سیل در بخش‌های پایین‌دست شود، بنابراین رابطه مستقیم بین افزایش ارتفاع و میزان بارندگی، همراه با اثرات توپوگرافی، نقش مهمی در شدت و فراوانی سیلاب‌ها در این حوزه ایفا می‌کند. این نتایج با نتایج (کرمی و همکاران، ۱۴۰۲) همخوانی دارد به این صورت که بر اساس تحلیل شاخص بارش، روند کلی نشان می‌دهد که با افزایش ارتفاع از سطح دریا، میزان و احتمال وقوع بارندگی نیز افزایش می‌یابد. مناطق مرتفع به دلیل شرایط جوی خاص، مانند برخورد جریان‌های هوایی مرطوب و سرد با سطوح بالاتر جو، معمولاً بارش‌های بیشتری را تجربه می‌کنند. این پدیده در بسیاری از مطالعات (مانند Smith, 1979) تایید قرار گرفته است، که نشان می‌دهد تغییرات دما و فشار در این مناطق به طور قابل توجهی الگوهای بارش را تحت تأثیر قرار می‌دهد. با این حال، باوجود بارش بیشتر در ارتفاعات بالاتر، خطر وقوع سیل عمدتاً در ارتفاعات پایین‌تر بیشتر است. این موضوع به دلیل جمع شدن رواناب‌ها و جریان‌های آبی حاصل از بارش‌ها در نقاط پست‌تر رخ می‌دهد که ظرفیت جذب و هدایت آب در این مناطق محدودتر است. به عبارت دیگر، آب باران جمع‌آوری شده در ارتفاعات بالا پس از جاری شدن به سمت پایین، در مناطق کم‌ارتفاع تجمع می‌یابد و منجر به افزایش سطح آب‌ها و ایجاد سیلاب می‌شود. بنابراین، هرچند بارش در ارتفاعات بالاتر بیشتر است، اما مخاطره سیل بیشتر متوجه مناطق پایین‌دست و با ارتفاع کمتر خواهد بود. این نتایج با نتایج (پورنبی درزی و همکاران، ۱۴۰۰) همخوانی دارد به این صورت که بارش بیشتر در ارتفاعات رخ می‌دهد، لیکن مخاطرات سیلابی غالباً در مناطق با ارتفاع پایین‌تر و پایین‌دست متمرکزند که تحت تأثیر هیدروگراف تجمعی قرار دارند. از نظر متغیر خاک، در منطقه مورد مطالعه وقوع سیل عموماً در مناطقی که دارای خاک (اریدیسول) است، رخ داده است. این نوع خاک پوشش گیاهی کمی دارند و به دلیل وجود افق‌های زیرسطحی غنی از کربنات کلسیم و گچ، نفوذ آب محدود و رواناب سطحی سریع تجمع می‌یابد. میزان کم مواد آلی و ساختار ضعیف خاک

باعث افزایش رواناب و کاهش ظرفیت جذب آب می‌شود. این ویژگی‌ها سبب می‌شوند که در هنگام بارش‌های ناگهانی و شدید در حوضه مورد مطالعه، آب به سرعت در سطح جمع شده و احتمال وقوع سیل را افزایش دهد. این نتایج با نتایج (میربایی و همکاران، ۱۴۰۰) همخوانی دارد به این صورت که کمبود مواد آلی در خاک‌های خشک منجر به کاهش ظرفیت جذب آب و افزایش رواناب است.

در رابطه با کاربری اراضی در منطقه مورد مطالعه در کاربری‌های مربوط به شهری، کشاورزی و پوشش گیاهی متوسط هر یک به نحوی در وقوع سیل نقش دارند. در مناطق شهری به دلیل افزایش سطح نفوذناپذیر (مانند آسفالت و ساختمان‌ها)، بیشترین رواناب سطحی ایجاد می‌شود که جریان سریع آب و احتمال سیل را افزایش می‌دهد. اراضی کشاورزی با وجود داشتن پوشش گیاهی، در صورتی که شیوه‌های مدیریت مناسب رعایت نشود (مانند تخریب خاک)، نفوذ آب کاهش یافته و رواناب سطحی افزایش می‌یابد و احتمال بروز سیل بیشتر است. پوشش گیاهی متوسط به اندازه کافی نمی‌تواند آب باران را جذب و رواناب را کنترل کند، پس سرعت جریان آب را کاهش نداده و محافظت خاک ضعیف‌تر است که نهایتاً موجب افزایش احتمال وقوع سیل می‌شود؛ بنابراین ترکیب این سه نوع کاربری اراضی می‌تواند شرایط را برای بروز سیل‌های مخرب در حوضه مورد مطالعه فراهم کند. این نتایج با نتایج رحمتی و همکاران (۲۰۱۶) مطابقت دارد به این صورت که در مناطق شهری به دلیل ساخت‌وساز و تغییر کاربری، سیل در منطقه مورد مطالعه افزایش یافته است. به عبارت دیگر کاربری اراضی نقش تعیین‌کننده‌ای در شدت و سرعت نفوذ رواناب دارد. با تغییرات کاربری اراضی که ناشی از دخالت‌های انسانی است، توان جذب و هدایت آب‌های سطحی کاهش یافته و این امر به عنوان شاخصی کلیدی در پیش‌بینی و تحلیل وقوع سیلاب به شمار می‌آید.

در منطقه مورد مطالعه شیب کم باعث کاهش سرعت جریان آب شده است که به همان نسبت زمان ماندگاری و تجمع آب بر روی سطح زمین افزایش یافته است. این تجمع طولانی مدت در مناطق کم‌شیب اجازه نمی‌دهد آب به سرعت به مسیرهای طبیعی خود هدایت شود و امکان نفوذ به زمین نیز کاهش می‌یابد، زیرا معمولاً خاک مناطق کم‌شیب چگالی بیشتری دارد و نفوذپذیری کمتر است. همچنین شیب کم با کاهش انرژی جریان آب، فرایند فرسایش را کاهش داده؛ اما باعث افزایش جریان سطحی گسترده و تجمع حجم زیاد آب در منطقه شده است که احتمال وقوع سیل را افزایش داده است. این نتایج با نتایج رضایی مقدم و رحیم‌پور (۱۴۰۲) همخوانی دارد به این صورت که سیلاب‌های منطقه مورد مطالعه عمدتاً در نواحی‌ای رخ داده‌اند که دارای شیب بسیار کم و ملایمی بین ۰ تا ۱۰ درصد هستند. این شیب پایین باعث می‌شود حرکت رواناب به صورت کندتر صورت گیرد و تجمع آب در سطح زمین افزایش یابد که در نهایت احتمال وقوع سیلاب را بالا می‌برد. چنین شرایطی اغلب منجر به کاهش نفوذ آب به داخل خاک و افزایش رواناب سطحی شده و بار فشار هیدرولیکی بر شبکه‌های زهکشی و آبراهه‌ها افزوده می‌شود که خود عاملی مؤثر در تشدید سیلاب‌ها به شمار می‌آید.

در رابطه با تراکم زهکشی در منطقه مورد مطالعه باید گفت که در جاهای که تراکم زهکشی زیاد بوده است، سیل رخ داده است. به عبارت دیگر افزایش تراکم زهکشی در منطقه، به معنای توسعه گسترده شبکه آبراهه‌ها و مسیرهای رواناب است که سرعت تخلیه آب باران را بالا می‌برد و می‌تواند باعث تجمع ناگهانی حجم زیاد رواناب در نقاط پایین دست شود. این پدیده علی‌رغم اینکه موجب انتقال سریع آب می‌شود، به دلیل تمرکز رواناب و تشکیل جریان‌های قوی سبب افزایش احتمال سیل‌های ناگهانی و مخرب در منطقه مورد مطالعه شده است. این نتایج با نتایج (Rafiei-Sardooi et al, 2021) همخوانی دارد به این صورت که با افزایش تراکم زهکشی، خاک قادر به جذب و نفوذ آب نیست و در نتیجه سیل با شدت بیشتری رخ می‌دهد. این در حالی است که این نتایج با نتایج رضایی و همکاران (۱۴۰۳) همخوانی ندارد به این صورت که سیلاب‌ها بیشتر در مناطقی رخ داده‌اند که تراکم شبکه زهکشی در آن‌ها پایین‌تر است. در مناطقی که این شبکه کمتر

توسعه یافته یا کم تراکم است، رواناب‌های ناشی از بارش باران نمی‌توانند به سرعت و به طور مؤثر از زمین خارج شوند. در نتیجه، آب روی سطح زمین باقی می‌ماند و تجمع می‌کند که این موضوع باعث افزایش حجم آب در نقاط خاص شده و احتمال آب‌گرفتگی و سیلاب را افزایش می‌دهد؛ بنابراین، تراکم پایین زهکشی به معنای کاهش ظرفیت منطقه در کنترل و هدایت آب‌های سطحی است که عامل مهمی در بروز سیلاب‌های شدید به شمار می‌رود. نتایج منحنی ROC و مقدار سطح زیر منحنی (AUC) نشان داد که هر دوی این مدل‌ها (داده‌های آموزشی ($RF=0/99$ و $SVM=0/93$) و داده‌های آزمایشی ($RF=0/83$ و $SVM=0/84$)) مناطق حساس سیل را به خوبی پیش‌بینی کرده‌اند.

نتیجه‌گیری

سیلاب‌ها به‌عنوان پدیده‌های ژئومورفولوژیکی ناشی از تعامل پیچیده عوامل اقلیمی، زمین‌شناسی و هیدرولوژیکی، هر ساله سبب بروز خسارات جانی و اقتصادی گسترده‌ای در مناطق مختلف کشور می‌شوند. تغییرات اقلیمی و رشد سریع شهرنشینی باعث افزایش تعداد و شدت سیلاب‌های شهری شده‌اند. به همین دلیل، پیش‌بینی دقیق سیل اهمیت بیشتری پیدا کرده است. این بررسی باهدف تهیه نقشه‌های پتانسیل خطر سیلاب در حوزه شهری اقلید انجام شده است تا زمینه مناسبی برای برنامه‌ریزی در مدیریت این مخاطره فراهم شود. در این مطالعه، از ۱۵ متغیر مؤثر بر وقوع سیلاب به همراه داده‌های مناطق دارای سیل (۵۰) و فاقد سیل (۱۰۰) در محیط ARCGISPro و R و از دو مدل ماشینی جنگل تصادفی (Random Forest) و ماشین بردار پشتیبان (Support Vector Machine) استفاده گردید تا نقشه‌های پهنه‌بندی خطر تولید شوند. برای آموزش مدل‌های موردنظر و ارزیابی دقت آن‌ها، داده‌های مربوط به نقاط سیل و عدم وقوع سیل به دو مجموعه تقسیم شدند؛ ۷۰ درصد برای آموزش و ۳۰ درصد برای اعتبارسنجی مدل‌ها اختصاص یافت. متغیرهای بارندگی، خاک، شیب، کاربری اراضی و تراکم زهکشی بیشترین اهمیت را در فرایند مدل‌سازی دارا بودند. نتایج حاصل از سطح زیر منحنی ($RF=0/83$ و $SVM=0/84$) نشان داد هر دو مدل تقریباً یکسان اعتبارسنجی کردند. همچنین از این مقاله می‌توان نتیجه گرفت که مدل‌های یادگیری ماشینی مانند جنگل تصادفی و ماشین بردار پشتیبان می‌توانند در پیش‌بینی حساسیت سیل در مناطق شهری به کار گرفته شوند. این مطالعه نشان داد که این مدل‌ها دقت بالایی در اعتبارسنجی دارند و می‌توانند اطلاعات مفیدی را در اختیار تصمیم‌گیرندگان قرار دهند تا نواحی اولویت‌دار برای مدیریت کاهش اثرات سیل را شناسایی کنند. نتایج این مطالعه می‌تواند به بهبود مدیریت سیل در مناطق شهری کمک کند، چون مبنای علمی قوی برای تصمیم‌گیری‌های مرتبط با کاهش خسارات سیل فراهم می‌آورد. با این حال، همان‌طور که اشاره شده، محدودیت‌هایی نیز وجود دارد که هنگام تفسیر یافته‌ها و توصیه‌های این مطالعه باید به آن‌ها توجه شود. از جمله این محدودیت‌ها می‌توان به تمرکز روی یک حوزه شهری خاص (اقلید)، محدودیت در تعداد داده‌های رخدادهای سیل، عدم در نظر گرفتن اثرات تغییرات اقلیمی و چالش‌هایی در تفسیر مدل‌های پیچیده اشاره کرد؛ بنابراین، نتایج این مطالعه ارزشمند هستند؛ ولی بهتر است با احتیاط و همراه با تحقیقات تکمیلی در مناطق و شرایط مختلف مورد استفاده قرار بگیرند. در مجموع، این مطالعه باعث می‌شود که مدیریت سیل در مناطق شهری به‌جای تکیه بر حدس و روش‌های سنتی، بر پایه تحلیل‌های علمی، داده‌محور و مدل‌های پیشرفته انجام شود که منجر به کاهش خسارات جانی و مالی و افزایش امنیت و پایداری شهری خواهد شد.

حامی مالی

این اثر حامی مالی نداشته است.

سهم نویسندگان در پژوهش

نویسندگان در تمام مراحل و بخش‌های انجام پژوهش سهم برابر داشتند.

تضاد و منافع

نویسندگان اعلام می‌دارند که هیچ تضاد منافی در رابطه با نویسندگی و یا انتشار این مقاله ندارند.

تقدیر و تشکر

نویسندگان از همه کسانی که در انجام این پژوهش به ما یاری رساندند، به‌ویژه کسانی که کار ارزیابی کیفیت مقاله را انجام دادند، تشکر و قدردانی می‌نمایند.

منابع

- امیری، پویا؛ عقیقی، محمدابراهیم و موغلی، مرضیه. (۱۴۰۱). آینده‌پژوهی و پتانسیل سنجی خطر وقوع سیل در شهر تهران با استفاده از الگوریتم ماشین پشتیبان بردار SVM. *جغرافیای طبیعی*، ۵۵(۱۴)، ۷۷-۹۳. DOI: 20.1001.1.20085656.1401.15.55.5.7
- پورینی درزی، سمیه؛ وفاخواه، مهدی و رجبی، محمد رسول. (۱۴۰۰). پهنه‌بندی خطر سیل با استفاده از مدل هیدرولیکی HEC-RAS و Arc GIS (مطالعه موردی: حوزه آبخیز چشمه کیله شهرستان تنکابن). *مخاطرات محیط طبیعی*، ۱۰(۲۸)، ۱۵-۲۸. <https://doi.org/10.22111/jneh.2021.28694.1603>
- حنیفی نیا، عبدالعزیز و عبقری، هیراد. (۱۴۰۴). پیش‌بینی مناطق مستعد سیل با استفاده از مدل‌های یادگیری ماشین خطی تعمیم‌یافته و بیشینه آنتروپی. *مخاطرات محیط طبیعی*، ۱۴(۴۳)، ۱۲-۳۴. <https://doi.org/10.22111/jneh.2024.47730.2021>
- رضایی، فاطمه؛ پور قاسمی، حمیدرضا؛ فلاح شمسی، سید رشید؛ خسروی شرف‌آبادی، رضا و کریمی‌نژاد، نرگس. (۱۴۰۳). مدل‌سازی مکانی و تهیه نقشه پتانسیل سیل‌گیری با استفاده از الگوریتم‌های یادگیری ماشین (مطالعه موردی: استان بوشهر). *مدیریت جامع حوزه‌های آبخیز*، ۴(۲)، ۸۱-۹۶. <https://doi.org/10.22034/iwm.2024.2014966.1113>
- رضایی مقدم، محمدحسین و رحیم‌پور، توحید. (۱۴۰۲). تهیه نقشه پتانسیل خطر وقوع سیل با استفاده از دو روش نسبت فراوانی و شاخص آماری (مطالعه موردی: حوزه آبریز آجی چای). *مدیریت مخاطرات محیطی*، ۱۰(۴): 803. 10.22059/jhsci.2024.369163.803
- علوی، سیدجلیل؛ احمدی، کوروش؛ حسینی، سید محسن؛ طبری، مسعود و نوری، زهرا. (۱۳۹۸). اهمیت متغیرهای اقلیمی، توپوگرافی و خاک در پراکنش سرخدار و اولویت‌بندی رویشگاه‌های مطلوب برای حفاظت و احیا در جنگل‌های خزری، مجله جنگل ایران. *انجمن جنگل‌بانی ایران*، ۱۱(۴)، ۴۷۷-۴۹۲.
- کریمی، پیمان؛ اسلامی نژاد، سید احمد؛ افتخاری، مبین؛ اکبری، محمد و راستگو، ملیکا. (۱۴۰۲). پهنه‌بندی حساسیت سیلاب با استفاده از روش‌های یادگیری ماشین بهبود یافته توسط الگوریتم ژنتیک. *محیط زیست طبیعی*، ۷۶(۱)، ۴۳-۶۰. <http://doi.org/10.22059/jne.2022.350170.2485>
- میربابایی، سیده مهرانوش؛ شعبان پور شهرستانی، محمود؛ خالدیان، محمدرضا و ذوالفقاری، علی‌اصغر. (۱۴۰۰). بررسی رابطه بین آب‌گریزی طبیعی و ویژگی‌های فیزیک و شیمیایی خاک در کاربری‌های مختلف در مناطق ساحلی غرب گیلان. *مجله تحقیقات آب‌و‌خاک ایران (مجله علوم کشاورزی ایران)*، ۵۲(۷)، ۱۸۲۳-۱۸۰۷. <https://doi.org/10.22059/ijswr.2021.323653.668972>
- نادری، حامد. رستگار سرخه، محمدعلی. استادی، بختیار و کارگری، مهرداد. (۱۴۰۴). پیش‌بینی احتمال وقوع ریسک عملیاتی در صنعت بانک داری با استفاده از الگوریتم‌های یادگیری ماشینی. *مدیریت دارایی و تامین مالی*، ۱۳(۴)، ۷۷-۹۶. <https://doi.org/10.22108/amf.2025.143030.1929>

References

- Alfieri, L., Bisselink, B., Dottori, F., Naumann, G., de Roo, A., Salamon, P., Wyser, K., & Feyen, L. (2017). Global projections of river flood risk in a warmer world. *Earth's Future*, 5(2), 171–182. <https://doi.org/10.1002/2016EF000485>

- Alavi, S. J., Ahmadi, K., Hosseini, S. M., Tabari, M., & Nouri, Z. (2019). Importance of climatic, topographic, and soil variables in the distribution of Persian ironwood and prioritization of suitable habitats for conservation and restoration in the Caspian forests. *Iranian Journal of Forest*, 11(4), 477–492. In Persian
- Amiri, P., Afifi, M. A., & Moghli, M. (2022). Foresight and potential assessment of flood risk in Tehran city using Support Vector Machine (SVM) algorithm. *Journal of Natural Geography*, 55(14), 77–93. DOI: 20.1001.1.20085656.1401.15.55.5.7 [In Persian]
- Apollonio, C., Balacco, G., Novelli, A., Tarantino, E., & Piccinni, A. F. (2016). Land use change impact on flooding areas: The case study of Cervaro Basin (Italy). *Sustainability*, 8(10), 996. <https://doi.org/10.3390/su8100996>
- Arabameri, A., Pradhan, B., Rezaei, Kh., & Conoscent, Ch. (2019). Gully erosion mapping using GIS-based multi-criteria decision analysis techniques. *Catena*, 180, 282–297. <https://doi.org/10.1016/j.catena.2019.04.032>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Bui, D. T., Pradhan, B., Revhaug, I., Nguyen, D. B., Pham, H. V., & Bui, Q. N. (2015). A novel hybrid evidential belief function-based fuzzy logic model in spatial prediction of rainfall-induced shallow landslides in the Lang Son city area (Vietnam). *Geomatics, Natural Hazards and Risk*, 6(3), 243–271. <https://doi.org/10.1080/19475705.2013.843206>
- Choubin, B., Moradi, E., Golshan, M., Adamowski, J., Sajedi-Hosseini, F., & Mosavi, A. (2019). An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. *Science of the Total Environment*, 651, 2087–2096. <https://doi.org/10.1016/j.scitotenv.2018.10.064>
- Costache, R. (2019). Flood susceptibility assessment by using bivariate statistics and machine learning models—a useful tool for flood risk management. *Water Resources Management*, 33(9), 3239–3256. <https://doi.org/10.1007/s11269-019-02301-z>
- Dai, W., Tang, Y., Liao, N., Zou, S., & Cai, Z. (2024). Urban flood prediction using ensemble artificial neural network: An investigation on improving model uncertainty. *Applied Water Science*, 14(6), 144. <https://doi.org/10.1007/s13201-024-02201-7>
- Derue, D. S., Nahrgang, J. D., Wellman, N. E., & Humphrey, S. E. (2011). Trait and behavioral theories of leadership: An integration and meta-analytic test of their relative validity. *Personnel Psychology*, 64(1), 7–52. <https://doi.org/10.1111/j.1744-6570.2010.01201.x>
- Duan, C., Zhang, J., Chen, Y., Lang, Q., Zhang, Y., Wu, C., & Zhang, Z. (2022). Comprehensive risk assessment of urban waterlogging disaster based on MCDA-GIS integration: The case study of Changchun, China. *Remote Sensing*, 14(13), 3101. <https://doi.org/10.3390/rs14133101>
- Ganjirad, M., & Delavar, M. R. (2023). Flood risk mapping using random forest and support vector machine. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 10, 201–208. <https://doi.org/10.5194/isprs-annals-X-4-W1-2022-201-2023>
- Hanifi Nia, A., & Abghari, H. (2025). Prediction of flood-prone areas using generalized linear models and maximum entropy. *Natural Environmental Hazards*, 14(43), 12–34. <https://doi.org/10.22111/jneh.2024.47730.2021> In Persian
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer. <https://doi.org/10.1007/978-0-387-84858-7>
- Kanani-Sadat, Y., Arabsheibani, R., Karimipour, F., & Nasser, M. (2019). A new approach to flood susceptibility assessment in data-scarce and ungauged regions based on GIS-based hybrid multi criteria decision-making method. *Journal of Hydrology*, 572, 17–31. <https://doi.org/10.1016/j.jhydrol.2019.02.034>
- Karami, P., Eslami Nejad, S. A., Eftkhari, M., Akbari, M., & Rastgoo, M. (2023). [Flood susceptibility zoning using machine learning methods improved by genetic algorithm]. *Natural Environment*, 76(1), 43–60. <http://doi.org/10.22059/jne.2022.350170.2485>
- Koggalage, R., & Halgamuge, S. (2004). Reducing the number of training samples for fast support vector machine classification. **Neural Information Processing-Letters and Reviews*, 2*(3), 57–65.

- Ma, M., Zhao, G., He, B., Li, Q., Dong, H., Wang, S., & Wang, Z. (2021). XGBoost-based method for flash flood risk assessment. *Journal of Hydrology*, 598, 126382. <https://doi.org/10.1016/j.jhydrol.2021.126382>
- Mirbabayi, S. M., Shabanpour Shahrastani, M., Khalidian, M. R., & Zolfaghari, A. A. (2021). [Investigation of the relationship between natural soil water repellency and physicochemical properties in different land uses in the coastal areas of western Gilan]. *Iranian Journal of Soil and Water Research*, 52(7), 1807–1823. <https://doi.org/10.22059/ijswr.2021.323653.668972> In Persian
- Nachappa, T. G., Piralilou, S. T., Gholamnia, K., Ghorbanzadeh, O., Rahmati, O., & Blaschke, T. (2020). Flood susceptibility mapping with machine learning, multi-criteria decision analysis and ensemble using Dempster Shafer Theory. *Journal of Hydrology*, 590, 125275. <https://doi.org/10.1016/j.jhydrol.2020.125275>
- Naderi, H., Rastegar Sorkhe, M. A., Estadi, B., & Karagari, M. (2025). [Prediction of operational risk probability in the banking industry using machine learning algorithms]. *Asset Management and Financing*, 13(4), 77–96. <https://doi.org/10.22108/amf.2025.143030.1929> In Persian
- Noymanee, J., Nikitin, N. O., & Kalyuzhnaya, A. V. (2017). Urban pluvial flood forecasting using open data with machine learning techniques in pattani basin. *Procedia Computer Science*, 119, 288–297. <https://doi.org/10.1016/j.procs.2017.11.187>
- Pallard, B., Castellarin, A., & Montanari, A. (2009). A look at the links between drainage density and flood statistics. *Hydrology and Earth System Sciences*, 13(7), 1019–1029. <https://doi.org/10.5194/hess-13-1019-2009>
- Pournabi Derazi, S., Vafakhvah, M., & Rajabi, M. R. (2021). [Flood hazard zoning using HEC-RAS hydraulic model and ArcGIS (Case study: Cheshmeh Kileh watershed, Tonekabon County)]. *Natural Environmental Hazards*, 10(28), 15–28. <https://doi.org/10.22111/jneh.2021.28694.1603>
- Qian, Y., Aghaabbasi, M., Ali, M., Alqurashi, M., Salah, B., Zainol, R., Moeinaddini, M., & Hussein, E. E. (2021). Classification of imbalanced travel mode choice to work data using adjustable SVM model. *Applied Sciences*, 11(24), 11916. <https://doi.org/10.3390/app112411916>
- Rafiei-Sardooi, E., Azareh, A., Choubin, B., Mosavi, A. H., & Clague, J. J. (2021). Evaluating urban flood risk using hybrid method of TOPSIS and machine learning. *International Journal of Disaster Risk Reduction*, 66, 102614. <https://doi.org/10.1016/j.ijdrr.2021.102614> In Persian
- Rahmati, O., Pourghasemi, H. R., & Zeinivand, H. (2016). Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the Golastan Province, Iran. *Geocarto International*, 31(1), 42–70. <https://doi.org/10.1080/10106049.2015.1041559> In Persian
- Rezaei, F., Pourghasemi, H. R., Fallah Shamsi, S. R., Khosravi Sharafabadi, R., & Karimi Nejad, N. (2024). [Spatial modeling and flood potential mapping using machine learning algorithms (Case study: Bushehr Province)]. *Watershed Management Research*, 4(2), 81–96. <https://doi.org/10.22034/iwm.2024.2014966.1113>
- Rezaei Moghadam, M. H., & Rahimpour, T. (2023). [Flood hazard potential mapping using frequency ratio and statistical index methods (Case study: Ajichay watershed)]. *Environmental Hazard Management*, 10(4), 13–24. <https://doi.org/10.22059/jhsci.2024.369163.803>
- Rizal, H. M., Hariadi, M., Arif, Y. M., & Warni, E. (2025). Enhancing flood prediction in urban areas: A machine learning approach for Makassar City. *Engineering, Technology & Applied Science Research*, 15(2), 21678–21684. <https://doi.org/10.48084/etasr>
- Sarker, I. H. (2021). Data science and analytics: an overview from data-driven smart computing, decision-making and applications perspective. *SN Computer Science*, 2(5), 377. <https://doi.org/10.1007/s42979-021-00765-8>
- Schmidt, L., Heße, F., Attinger, S., & Kumar, R. (2020). Challenges in applying machine learning models for hydrological inference: A case study for flooding events across Germany. *Water Resources Research*, 56(5), e2019WR025924. <https://doi.org/10.1029/2019WR025924>
- Shahabi, H., Khezri, S., Ahmad, B. B., & Hashim, M. (2014). Landslide susceptibility mapping at central Zab basin, Iran: A comparison between analytical hierarchy process, frequency ratio and logistic regression models. *Catena*, 115, 1–12. <https://doi.org/10.1016/j.catena.2013.11.014>

- Smith, R. B. (1979). The influence of mountains on the atmosphere. *Advances in Geophysics*, 21, 87–230. [https://doi.org/10.1016/S0065-2687\(08\)60262-9](https://doi.org/10.1016/S0065-2687(08)60262-9)
- Svetnik, V., Liaw, A., Tong, C., & Wang, T. (2004). Application of Breiman's random forest to modeling structure-activity relationships of pharmaceutical molecules. In F. Roli, J. Kittler, & T. Windeatt (Eds.), *Multiple classifier systems* (pp. 334–343). Springer. https://doi.org/10.1007/978-3-540-25966-4_33
- Tehrany, M. S., & Kumar, L. (2018). The application of a Dempster–Shafer-based evidential belief function in flood susceptibility mapping and comparison with frequency ratio and logistic regression methods. *Environmental Earth Sciences*, 77(13), 490. <https://doi.org/10.1007/s12665-018-7667-0>
- Ward, P. J., de Ruiter, M. C., Mård, J., Schröter, K., Van Loon, A., Veldkamp, T., von Uexkull, N., Wanders, N., AghaKouchak, A., Arnbjerg-Nielsen, K., Capewell, L., Coughlan de Perez, E., Dai, X., Day, R., Dewals, B., Di Baldassarre, G., Huning, L. S., Jupner, R., Kelman, I., ... Wens, M. (2020). The need to integrate flood and drought disaster risk reduction strategies. *Water Security*, 11, 100070. <https://doi.org/10.1016/j.wasec.2020.100070>
- Xiao, S., Zou, L., Xia, J., Dong, Y., Yang, Z., & Yao, T. (2023). Assessment of the urban waterlogging resilience and identification of its driving factors: A case study of Wuhan City, China. *Science of the Total Environment*, 866, 161321. <https://doi.org/10.1016/j.scitotenv.2022.161321>
- Xu, D., Qi, X., Li, C., Sheng, Z., & Huang, H. (2021). Wise information technology of med: Human pose recognition in elderly care. *Sensors*, 21(21), 7130. <https://doi.org/10.3390/s21217130>
- Yu, H., Luo, Z., Wang, L., Ding, X., & Wang, S. (2023). Improving the accuracy of flood susceptibility prediction by combining machine learning models and the expanded flood inventory data. *Remote Sensing*, 15(14), 3601. <https://doi.org/10.3390/rs15143601>
- Zhu, Z., & Zhang, Y. (2022). Flood disaster risk assessment based on random forest algorithm. *Neural Computing and Applications*, 34(5), 3443–3455. <https://doi.org/10.1007/s00521-021-05757->