

Research Paper

Comparing the Performance of Machine Learning Models in Flood Risk Zoning in Nekarood Watershed

SeidMohamad Mousavi,¹ Karim Solaimani,^{2*} Fatemeh Shokrian,³ Omid Ghorbanzadeh,⁴

¹PhD Student, Sari Agricultural Sciences and Natural Resources University, Sari, Iran.

²Professor, Environmental RS & GIS Research Institute, Department of Watershed Management, Sari Agricultural Sciences and Natural Resources University, Sari, Iran

³Associate Professor, Department of Watershed Management, Sari Agricultural Sciences and Natural Resources University, Sari, Iran.

⁴Assistant Professor, Institute of Geomatics, University of Natural Resources and Life Sciences, Vienna, Austria.



10.22125/iwe.2026.562734.1911

Received:
November 7, 2025
Accepted:
March 11, 2026
Available online:
June 25, 2026

Keywords:
Flood, Machine Learning, Gradient Boosting, Support Vector Machine, Nekarood

Abstract

Flooding stands as one of the most destructive hydro-climatic hazards, inflicting substantial annual losses on human communities and economic infrastructure. Identifying flood-prone zones and creating sensitivity maps are crucial for effective risk management and natural resource planning. This study focused on assessing the performance of three machine learning algorithms—logistic regression (LR), support vector machine (SVM), and gradient boosting (GB)—for flood risk mapping in the Nekarood watershed of Mazandaran province. Thirteen environmental and hydrological factors, including rainfall, slope, elevation, proximity to waterways, land use, and morphometric indices, were analyzed alongside 152 recorded flood occurrence points. The models were evaluated using metrics such as AUC, overall accuracy, and the kappa coefficient. Results indicated that the GB model achieved the highest performance, with an AUC of 0.896, an overall accuracy of 87%, and a kappa coefficient of 0.84. The SVM model followed with an AUC of 0.872 and an accuracy of 83%, while the LR model, scoring an AUC of 0.853 and an accuracy of 80%, showed the weakest performance among the three. Analysis of variable importance revealed that rainfall, slope, distance from waterways, and elevation are the most significant factors influencing flood occurrence. Consequently, reinforcement learning-based algorithms could serve as effective tools for enhancing predictive accuracy and minimizing uncertainty in flood risk mapping efforts. The findings from this study offer valuable insights for strengthening early warning systems, restricting development in high-risk zones, and reducing both human and economic damages caused by flooding in similar regions.

1. Introduction

Floods are among the most destructive hydroclimatic hazards, causing loss of human lives, damage to infrastructure, and severe economic impacts every year (Costache, 2019; Islam et al., 2021). Climate change, land-use/land-cover (LULC) changes, and rapid urbanization have intensified the frequency and magnitude of floods in many parts of the world, with projections estimating that by 2050 global

* **Corresponding Author:** Name

Address: Department, University, Country,

Email: Academic Email

Tel: Number

flood-related damages may exceed one trillion USD (Ali et al., 2019; Serraj & Pingali, 2019). Iran, due to its varied climatic and topographic conditions, is highly vulnerable to floods, which occur across almost all watersheds of the country. The Nekarood watershed, located in Mazandaran Province, northern Iran, is a critical region due to its dense settlements, agricultural activities, and vital infrastructures. Identifying flood-prone areas and preparing reliable susceptibility maps in such regions is vital for disaster risk reduction, sustainable land-use planning, and climate adaptation strategies (Parsian et al., 2021).

Traditional hydrological models, while widely used, suffer from significant challenges, including high data requirements, difficulty in simulating multidimensional watershed conditions, and limited adaptability to new environments. In recent years, the development of machine learning (ML) techniques has opened new opportunities for flood susceptibility modeling (Meliho et al., 2021). ML models can capture complex and nonlinear interactions between environmental and hydrological variables, thereby enhancing predictive performance compared to conventional methods. Among the wide spectrum of ML algorithms, ensemble methods such as Gradient Boosting (GB) have demonstrated superior robustness and accuracy in hazard prediction tasks (Friedman, 2001). This study aims to employ and compare three ML models—Logistic Regression (LR), Support Vector Machine (SVM), and Gradient Boosting (GB)—to map flood susceptibility in the Nekarood watershed and provide insights into the most effective predictive approaches for flood risk management.

2. Materials and Methods

The Nekarood watershed, covering approximately 1,048 km² in northern Iran, was selected as the case study area due to its high flood vulnerability. To build the susceptibility models, a comprehensive dataset was developed integrating 15 conditioning factors representing topographic, hydrological, climatic, geological, and land-use characteristics. These include rainfall, elevation, slope, aspect, curvature, distance from rivers, drainage density, soil type, land use, NDVI, TWI, SPI, and lithology. Data sources included DEMs from ASTER (30 m resolution), Landsat 8 OLI and Sentinel-2 images for land cover indices, ERA5 precipitation datasets, and geological and hydrological maps from national surveys. All layers were standardized to a 30 m spatial resolution and normalized to reduce scale effects.

Flood inventory mapping was conducted by collecting 152 flood occurrence points based on historical flood reports, remote sensing analysis, and field surveys. To balance the dataset, an equal number of non-flood points were randomly generated outside flood-affected zones. The dataset was divided into training (70%) and testing (30%) subsets.

Three ML algorithms were applied:

Logistic Regression (LR) – A benchmark linear classifier suitable for binary classification.

Support Vector Machine (SVM) – Applied with a radial basis function (RBF) kernel to model nonlinear relations.

Gradient Boosting (GB) – Implemented as an ensemble of decision trees with parameters: 1000 estimators, maximum depth = 20, learning rate = 0.01, and subsample = 0.8.

The models were evaluated using Overall Accuracy, Precision, Recall, F1-score, Kappa coefficient, and AUC–ROC metrics. Feature importance was derived from the GB model to quantify the relative contribution of each variable.

3. Results

The comparative analysis revealed significant differences in the predictive performance of the models. The GB model consistently outperformed the others with AUC = 0.896, overall accuracy = 87%, Kappa = 0.84, precision = 94.93%, recall = 93.20%, and F1-score = 94.05%. The SVM model achieved AUC = 0.872, accuracy = 83%, and Kappa = 0.79, while the LR model obtained AUC = 0.853, accuracy = 80%, and Kappa = 0.74. These results underline the strength of boosting-based approaches in capturing nonlinear patterns and complex feature interactions compared to linear or kernel-based models.

The variable importance analysis indicated that rainfall (28%), slope (21%), distance from rivers (19%), and elevation (15%) were the most influential predictors of flood susceptibility. This aligns with findings from previous studies (Pourghasemi et al., 2020; Kubra et al., 2024; Wida et al., 2024), reinforcing the role of climatic and geomorphological factors in shaping flood hazards. Land-use changes, particularly the expansion of impermeable surfaces, also had a measurable effect, highlighting the link between human activities and increased flood risks.

4. Discussion and Conclusion

The susceptibility maps revealed that approximately 19.41% of the watershed lies within high to very high flood-prone zones, which correspond to densely populated areas, agricultural lands, and infrastructure corridors. These results carry substantial implications for land-use planning and disaster management, as they provide actionable insights for prioritizing high-risk areas for flood mitigation measures.

In conclusion, the study demonstrates that Gradient Boosting offers a robust and reliable tool for flood susceptibility mapping in complex watersheds. Its superior accuracy and interpretability of variable contributions make it particularly suitable for decision-making contexts. Nevertheless, some limitations remain, including reliance on static flood inventories and the need for higher-resolution multi-temporal datasets. Future studies should integrate deep learning models, finer-scale remote sensing data, and real-time hydrological monitoring to further enhance flood prediction capabilities.

5. Six important references

- 1) Costache, R. (2019). Flash-flood Potential Index mapping using weights of evidence, decision trees models and their novel hybrid integration. *Stochastic Environmental Research and Risk Assessment*, 33(7), 1375–
- 2) Islam, A. R. M. T., et al. (2021). Flood susceptibility modelling using advanced ensemble machine learning models. *Geoscience Frontiers*, 12(3), 101075.
- 3) Kubra, A., Lokesh, K. J., Shakthi, G., Surya, M., Mohammad Jawad, A., & Melquier, A. (2024). Enhanced flood prediction and management for urban underpass using machine learning. In *International Conference on Computer & Communication Technologies* (pp. 385–395). Springer.
- 4) Meliho, M., Khattabi, A., & Asinyo, J. (2021). Spatial modeling of flood susceptibility using machine learning algorithms. *Arabian Journal of Geosciences*, 14(21), 2243.
- 5) Pourghasemi, H. R., Razavi-Termeh, S. V., Kariminejad, N., Hong, H., & Chen, W. (2020). An assessment of metaheuristic approaches for flood assessment. *Journal of Hydrology*, 582, 124536.
- 6) Wida, A., et al. (2024). Application of machine learning for flood susceptibility mapping in the U.S. watershed. *Journal of Hydrological Sciences*, 69(4), 585–601

مقایسه عملکرد مدل‌های یادگیری ماشین در پهنه‌بندی خطر سیل حوزه آبخیز نکارود

سیدمحمد موسوی^۱، کریم سلیمانی^{۱*}، فاطمه شکران^۲ و امید قربانزاده^۳

تاریخ ارسال: ۱۴۰۴/۰۸/۱۶

تاریخ پذیرش: ۱۴۰۴/۱۲/۲۰

مقاله پژوهشی

چکیده

شناسایی مناطق مستعد سیل و تهیه نقشه‌های پهنه‌بندی خطر، نقش کلیدی در مدیریت ریسک و برنامه‌ریزی فضایی ایفا می‌کند. هدف این پژوهش، ارزیابی و مقایسه عملکرد سه الگوریتم یادگیری ماشین شامل رگرسیون لجستیک (LR)، ماشین بردار پشتیبان (SVM) و گرادیان بوستینگ (GB) در پهنه‌بندی خطر سیل در حوزه آبخیز نکارود استان مازندران است. بدین منظور، ۱۳ عامل محیطی، هیدرولوژیکی و ژئومورفولوژیکی از جمله بارندگی، شیب، ارتفاع، فاصله از رودخانه و کاربری اراضی استخراج و با ۱۵۲ نقطه وقوع سیل واقعی تلفیق شد. مدل‌ها پس از پیش‌پردازش داده‌ها آموزش داده شده و عملکرد آن‌ها با استفاده از شاخص‌های AUC، دقت کلی و ضریب کاپا ارزیابی گردید. نتایج نشان داد که مدل گرادیان بوستینگ با مقدار AUC برابر با ۰.۸۹۶، دقت کلی ۰.۸۷٪ و ضریب کاپا ۰.۸۴ بهترین عملکرد را در شناسایی مناطق سیلابی و غیرسیلابی ارائه می‌دهد. مدل SVM با AUC برابر با ۰.۸۷۲ و دقت ۰.۸۳٪ در رتبه دوم قرار گرفت، در حالی که مدل رگرسیون لجستیک با AUC برابر با ۰.۸۵۳ و دقت ۰.۸۰٪ ضعیف‌ترین عملکرد را در میان مدل‌های مورد بررسی داشت. مقایسه نتایج بیانگر آن است که مدل‌های مبتنی بر درخت تصمیم و تقویت یادگیری، در مقایسه با مدل‌های خطی کلاسیک، توانایی بیشتری در بازنمایی الگوهای پیچیده مرتبط با وقوع سیل دارند. تحلیل اهمیت متغیرها نیز نشان داد که بارندگی، شیب، فاصله از رودخانه و ارتفاع مهم‌ترین عوامل مؤثر بر حساسیت به سیل در منطقه مطالعاتی هستند. نتایج نشان می‌دهد الگوریتم‌های یادگیری ماشین، به‌ویژه گرادیان بوستینگ، می‌تواند روش مناسب در پهنه‌بندی خطر سیل در مناطق مشابه باشد.

واژه‌های کلیدی: سیل، یادگیری ماشین، گرادیان بوستینگ، ماشین بردار پشتیبان، نکارود

^۱ استاد پژوهشکده سنجش از دور و GIS محیطی و گروه مهندسی آبخیزداری، دانشگاه علوم کشاورزی و منابع طبیعی ساری
*(نویسنده مسئول): ka.solaimani@sanru.ac.ir solaimani2001@yahoo.co.uk

^۲ دانشیار گروه مهندسی آبخیزداری، دانشکده منابع طبیعی ساری، دانشگاه علوم کشاورزی و منابع طبیعی ساری

^۳ استادیار، موسسه ژئوماتیک، دانشگاه منابع طبیعی و علوم زیستی، وین، اتریش



مقدمه

سیل یکی از جدی‌ترین و ویرانگرترین مخاطرات هیدرواقليمی در سطح جهان است که سالانه باعث بروز خسارات انسانی و اقتصادی گسترده می‌شود (Costache, 2019; Islam et al., 2021). براساس آمارهای جهانی، این پدیده نه تنها تهدیدی مستقیم برای سلامت و امنیت جوامع انسانی به شمار می‌رود، بلکه آثار مخربی بر زیرساخت‌ها، منابع طبیعی و روند توسعه اجتماعی-اقتصادی نیز بر جای می‌گذارد (شادمانی و همکاران، ۱۳۸۷). پیش‌بینی‌ها نشان می‌دهد که تحت تأثیر تغییرات اقلیمی، تغییر الگوهای بارش، تغییر کاربری زمین و افزایش تراکم جمعیت، شدت و فراوانی وقوع سیلاب‌ها تا سال ۲۰۵۰ به‌طور قابل توجهی افزایش یافته و خساراتی بالغ بر یک تریلیون دلار ایجاد خواهد کرد (Ali et al., 2019; Serraj and Pingali, 2019). در ایران نیز، وقوع سیل پدیده‌ای فراگیر است که تمامی نواحی کشور را تحت تأثیر قرار می‌دهد و محدود به منطقه یا موقعیت خاصی نیست. نوع سیل و میزان خسارات ناشی از آن بسته به ویژگی‌های محیطی، توپوگرافی و شرایط اقلیمی هر منطقه متفاوت است (کرمی و همکاران، ۱۴۰۲). این شرایط به‌ویژه در مناطق پرجمعیت و یا دارای زیرساخت‌های آسیب‌پذیر بحرانی‌تر است.

به‌طور کلی، سیل‌ها به چهار دسته اصلی شامل سیل‌های ناگهانی، رودخانه‌ای، ساحلی و شهری تقسیم می‌شوند (Costache et al., 2020). در میان آن‌ها، سیل‌های ناگهانی بیشترین قدرت تخریب را دارند؛ زیرا به‌دلیل ماهیت غیرمنتظره، شدت جریان بالا و سرعت گسترش سریع، می‌توانند در مدت زمان کوتاهی تلفات انسانی و خسارات سنگینی به بار آورند (Bui et al., 2020). بنابراین، شناسایی مناطق حساس به سیل و تهیه نقشه‌های پهنه‌بندی آسیب‌پذیری به‌عنوان یک ابزار حیاتی برای مدیریت بحران، کاهش ریسک و برنامه‌ریزی پایدار منابع

طبیعی و شهری ضروری است (Parsian et al., 2021).

در این راستا، پژوهش‌های اخیر تمرکز خود را بر توسعه روش‌های نوین برای ارزیابی خطرات سیل معطوف کرده‌اند. این مطالعات تلاش کرده‌اند تا با بهره‌گیری از مدل‌های کمی و کیفی، عوامل محیطی و انسانی مؤثر بر بروز سیل را شناسایی و ارتباط آن‌ها با شدت و گستره سیل تبیین کنند (Pollack et al., 2022; Seydi et al., 2022). مفهوم «حساسیت به سیل» در این زمینه اهمیت ویژه‌ای دارد؛ زیرا به توصیف و تحلیل قابلیت وقوع سیل در مناطق مختلف بر اساس شرایط جغرافیایی و محیطی می‌پردازد (Parizi et al., 2022). پهنه‌بندی سیل با بهره‌گیری از سامانه اطلاعات جغرافیایی (GIS)، به‌عنوان یکی از رویکردهای غیرسازهای، ابزاری مؤثر برای مدیریت و کاهش مخاطرات ناشی از سیل محسوب می‌شود (سلیمانی و محسنی، ۱۴۰۳). با وجود توسعه مدل‌های هیدرولوژیکی کلاسیک، محدودیت‌هایی همچون نیاز به داده‌های فراوان، پیچیدگی‌های محاسباتی و دشواری در بازنمایی شرایط چندبعدی حوضه‌ها، کاربرد آن‌ها را در بسیاری از مناطق با چالش مواجه کرده است. در سال‌های اخیر، پیشرفت‌های چشمگیری در حوزه تکنیک‌های یادگیری ماشینی (ML) صورت گرفته است که توانسته‌اند این محدودیت‌ها را کاهش داده و دقت تحلیل‌های مکانی را به‌طور قابل توجهی بهبود بخشند (Meliho et al., 2021). مدل‌های یادگیری ماشینی با توانایی درک روابط پیچیده و غیرخطی میان متغیرها، ابزاری قدرتمند برای پیش‌بینی و پهنه‌بندی سیل به شمار می‌روند و زمینه‌ساز تولید نقشه‌های دقیق‌تر و قابل اعتمادتر شده‌اند.

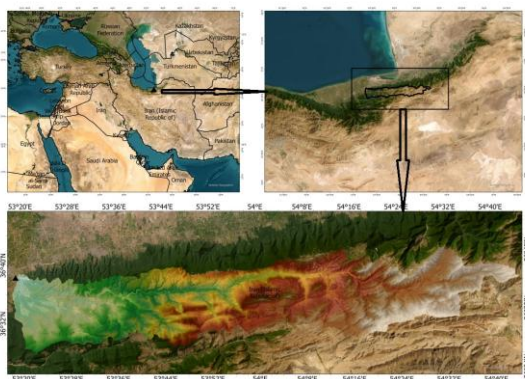
در مجموع، ادغام داده‌های مکانی و تکنیک‌های یادگیری ماشینی می‌تواند به شناسایی نواحی پرخطر، تحلیل دقیق‌تر عوامل مؤثر و در نهایت بهبود تصمیم‌گیری‌های مدیریتی کمک کند. این رویکرد نه

کارآمدی برای مدیریت ریسک و کاهش خسارات ناشی از سیل فراهم آورد.

مواد و روش ها

معرفی محدوده مورد مطالعه

حوزه آبخیز نکارود یکی از زیرحوضه‌های مهم شمال ایران در استان مازندران به‌شمار می‌رود که سرچشمه آن از ارتفاعات البرز مرکزی در جنوب شهرستان نکا آغاز شده و پس از طی مسیری با جهت کلی جنوب به شمال، به دریای خزر منتهی می‌شود (شکل ۱). این حوضه با وسعت تقریبی ۱۰۴۸ کیلومتر مربع، بین عرض‌های جغرافیایی ۳۶°۱۵' تا ۳۶°۳۰' شمالی و طول‌های جغرافیایی ۵۲°۳۰' تا ۵۲°۴۵' شرقی واقع شده است. از جنوب به ارتفاعات البرز و از شمال به جلگه‌های ساحلی دریای خزر محدود می‌شود. اقلیم منطقه مرطوب خزری با میانگین بارش سالانه بیش از ۱۰۰۰ میلی‌متر است. بخش‌های بالادست عمدتاً پوشیده از جنگل‌های هیرکانی بوده و در بخش‌های پایین‌دست اراضی کشاورزی، باغات و سکونتگاه‌های انسانی گسترده‌اند (شکل ۱). ترکیب ویژگی‌های طبیعی و انسانی موجب شده است که این حوضه یکی از مناطق حساس و آسیب‌پذیر در برابر سیلاب‌های فصلی و ناگهانی به‌شمار آید.



شکل (۱): منطقه مورد مطالعه

تنها موجب ارتقای کیفیت پیش‌بینی‌ها در مناطق حساس می‌شود، بلکه قابلیت تعمیم آن برای سایر حوضه‌ها و مناطق مشابه نیز وجود دارد. از این رو، پژوهش حاضر با تمرکز بر حوزه آبخیز نکارود استان مازندران، به ارزیابی و پهنه‌بندی خطر سیل با استفاده از مدل‌های یادگیری ماشینی پرداخته و تلاش می‌کند تا گامی در جهت بهبود مدیریت منابع آب و کاهش خسارات ناشی از سیل‌های فصلی و ناگهانی بردارد.

پژوهش‌های بین‌المللی و داخلی نیز اهمیت استفاده از مدل‌های یادگیری ماشین در پیش‌بینی سیل را تأیید کرده‌اند. به‌عنوان نمونه، Kubra و همکاران (۲۰۲۴) با ترکیب داده‌های بلادرنگ و مدل‌های یادگیری نظارت‌شده و بدون نظارت، چارچوبی برای مدیریت سیل ارائه کردند که نسبت به روش‌های سنتی دقت بالاتری داشت. همچنین، Wida و همکاران (۲۰۲۴) در ایالات متحده با بهره‌گیری از شش الگوریتم یادگیری ماشین از جمله RF و XGBoost، نشان دادند که متغیرهایی مانند فاصله از رودخانه، نوع خاک و بارش بیشترین نقش را در حساسیت به سیل دارند. در ایران نیز مطالعات مختلفی به این حوزه پرداخته‌اند؛ از جمله خدائی و زندی (۱۴۰۱) که با استفاده از تحلیل شبکه‌ای و شبکه عصبی مصنوعی در حوضه خداآفرین، ارتفاع رواناب، شیب و بارش را مهم‌ترین عوامل وقوع سیل شناسایی کردند. قره‌داغی و قاسم‌زاده (۱۴۰۳) نیز با ادغام داده‌های Sentinel و الگوریتم‌های یادگیری ماشین، پیش‌بینی دقیقی از نواحی مستعد سیل در حوضه کارون ارائه کردند و حنیفی‌نیا و عبقری (۱۴۰۴) نشان دادند که مدل‌های پیشینه آنتروپی و رگرسیون خطی تعمیم‌یافته در پهنه‌بندی خطر سیل حوزه زیوه عملکرد مناسبی دارند. این یافته‌ها در مجموع بیانگر آن است که استفاده از رویکردهای ترکیبی و مدل‌های نوین می‌تواند با کاهش عدم قطعیت و ارتقای دقت، مبنای



متغیرهای مؤثر بر وقوع سیل

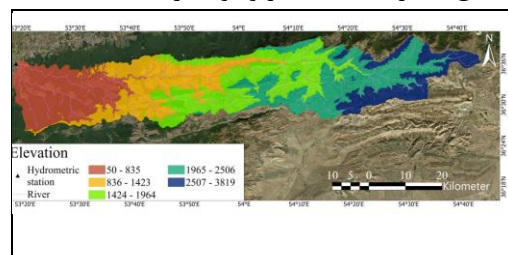
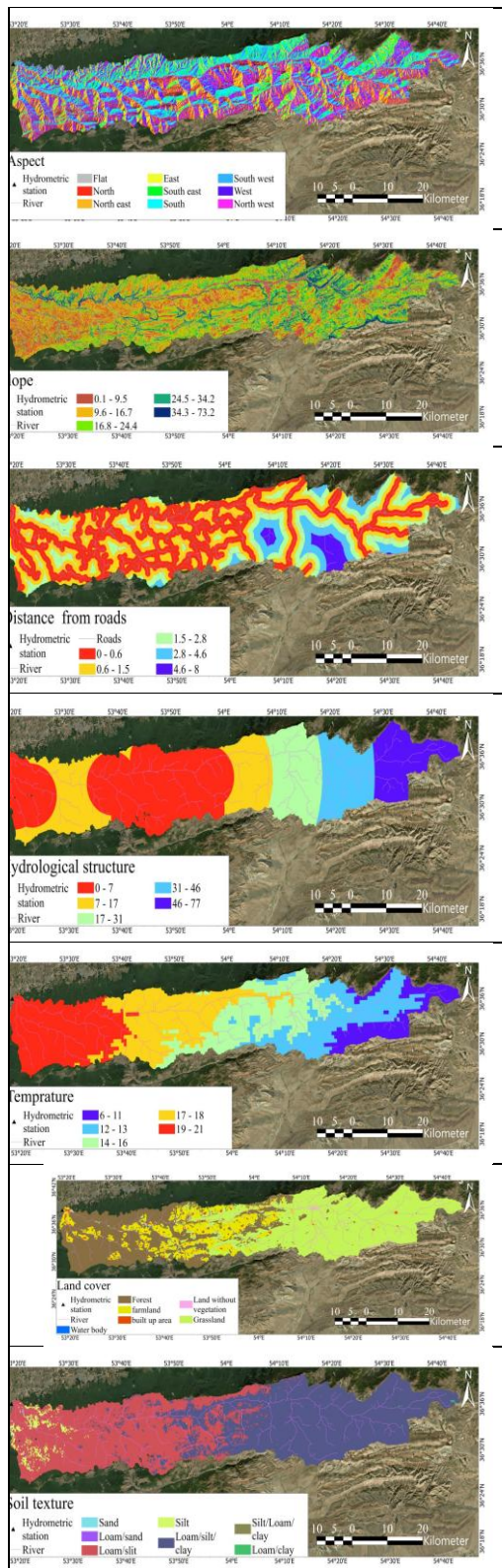
انتخاب متغیرهای مؤثر در مدل‌سازی خطر سیل نقشی کلیدی در افزایش دقت پیش‌بینی و ارتقای کیفیت نتایج ایفا می‌کند (Ahmad et al., 2025). بر این اساس، در پژوهش حاضر مجموعه‌ای از شاخص‌های محیطی و انسانی به‌عنوان عوامل اصلی وقوع و شدت سیل مورد استفاده قرار گرفتند. این متغیرها با استناد به مطالعات معتبر داخلی و بین‌المللی و با در نظر گرفتن شرایط ژئومورفولوژیکی، اقلیمی و کاربری اراضی در حوزه آبخیز نکارود انتخاب شدند. شاخص‌های مذکور شامل شیب زمین، ارتفاع از سطح دریا، جهت شیب، فاصله از رودخانه، بارندگی سالانه، نوع خاک، کاربری اراضی، شاخص پوشش گیاهی (NDVI)، فاصله از سازه‌های هیدرولوژیکی، شاخص رطوبت توپوگرافی (TWI)، شاخص مساحت حوضه و انحنای زمین هستند (شکل ۲).

ارتفاع از سطح دریا یکی از مهم‌ترین عوامل مؤثر بر بروز سیل محسوب می‌شود. به‌طور کلی، نواحی کم‌ارتفاع به دلیل تجمع رواناب، بیش از مناطق مرتفع در معرض سیل قرار دارند (Botzen et al., 2013). علاوه بر آن، شیب زمین عامل تعیین‌کننده دیگری است که بر سرعت رواناب، شدت جریان، فرآیند نفوذ و رسوب‌گذاری اثرگذار بوده و شرایط فیزیکی لازم برای وقوع سیل را فراهم می‌سازد (Samanta et al., 2018). مناطق با شیب ملایم معمولاً حساسیت بالاتری به سیل دارند، در حالی که شیب‌های تند گرچه منجر به افزایش سرعت جریان می‌شوند، اما الگوی خطر متفاوتی ایجاد می‌کنند (Kaur et al., 2017). جهت شیب نیز با تأثیر بر تابش خورشیدی، رطوبت خاک و تبخیر و تعرق، بر ذخیره و پایداری رطوبت دامنه‌ها اثرگذار است. در مجموع، متغیرهای توپوگرافی که عمدتاً از مدل‌های ارتفاع رقومی (DEM) استخراج می‌شوند، از مهم‌ترین ورودی‌ها در مدل‌سازی سیل محسوب می‌شوند (Kia et al., 2012).

کاربری اراضی و پوشش گیاهی نیز در تغییر الگوهای رواناب و انتقال رسوب نقش مهمی دارند. مناطق شهری به دلیل نفوذپذیری پایین سطح زمین و افزایش رواناب، به‌طور ویژه مستعد سیل هستند؛ در حالی که اراضی جنگلی با افزایش قابلیت نفوذ، شدت خطر را کاهش می‌دهند (Doudangeh et al., 2020). همچنین مطالعات متعددی نشان داده‌اند که تراکم پوشش گیاهی رابطه‌ای معکوس با احتمال وقوع سیل دارد. از سوی دیگر، فاصله از رودخانه‌ها نیز شاخصی تعیین‌کننده در حساسیت به سیل است؛ زیرا نواحی نزدیک به بستر رودخانه بیشترین آسیب‌پذیری را در زمان طغیان تجربه می‌کنند (Zeng et al., 2017). در میان شاخص‌های هیدرولوژیکی، شاخص رطوبت توپوگرافی (Topographic wetness index) جایگاه ویژه‌ای دارد. این شاخص توزیع مکانی رطوبت در مقیاس حوضه را بازتاب داده و بر مبنای مدل ارتفاعی زمین، الگوی تجمع آب و مسیر جریان‌های هیدرولوژیکی را کمی‌سازی می‌کند (Jahanbani et al., 2024). بارش سالانه نیز عامل اصلی در وقوع سیل بوده و تغییرات آن رابطه مستقیمی با دبی رودخانه‌ها و پتانسیل سیل‌خیزی حوضه دارد (Pourghasemi et al., 2020). در این پژوهش داده‌های بارش از ترکیب منابع سازمان هواشناسی ایران و پایگاه ERA5 (ERA5 — پایگاه داده اقلیمی Climate Reanalysis Database) استخراج شد تا دقت مکانی و زمانی تحلیل افزایش یابد. ویژگی‌های خاک نیز با کنترل فرآیند نفوذ و تبدیل بارش به رواناب، تأثیر مستقیم بر وقوع سیل دارند (Jahanbani et al., 2024). علاوه بر این، فاصله از سازه‌های هیدرولوژیکی همچون پل‌ها، سدها و بندها نیز به‌عنوان شاخصی مؤثر لحاظ شد؛ زیرا این سازه‌ها می‌توانند الگوی طبیعی جریان آب را تغییر داده یا موجب تمرکز رواناب در بخش‌های خاصی از حوضه شوند. سایر متغیرها مانند شاخص مساحت حوضه و انحنای زمین نیز در تحلیل شرایط

هیدرولوژیکی اهمیت دارند. انحنای مقعر با تمرکز جریان و افزایش حساسیت به سیل مرتبط است، در حالی که انحنای محدب پراکنش رواناب را افزایش داده و شدت خطر را کاهش می‌دهد (Saleh et al., 2022).

به منظور تولید و آماده‌سازی این متغیرها، از منابع متعددی استفاده شد: مدل ارتفاعی ASTER برای تولید متغیرهای توپوگرافی، تصاویر ماهواره‌ای Landsat 8 و Sentinel-2 برای استخراج شاخص NDVI و نقشه‌های کاربری اراضی، داده‌های خاک از پایگاه FAO و منابع داخلی، داده‌های بارش از ERA5 و سازمان هواشناسی ایران، و اطلاعات مربوط به رودخانه‌ها، جاده‌ها و سازه‌های هیدرولوژیکی از نقشه‌های توپوگرافی ۱:۲۵۰۰۰ سازمان نقشه‌برداری کشور. تمامی لایه‌ها به تفکیک مکانی ۳۰ متر بازنمونه‌برداری شده و در سیستم مختصات UTM Zone 39N (WGS84) یکپارچه گردیدند. به منظور کاهش ناهمگنی داده‌ها، کلیه متغیرها پیش از ورود به مدل، استانداردسازی (استانداردسازی یعنی تبدیل داده‌ها به توزیعی با میانگین = ۰ و انحراف معیار = ۱) و نرمال‌سازی (نرمال‌سازی یعنی تبدیل مقادیر یک ویژگی به یک بازه‌ی مشخص معمولاً ۰ تا ۱) شدند در نهایت، این پایگاه داده‌ی مکانی یکپارچه، مبنای آموزش و اعتبارسنجی مدل‌های یادگیری ماشین در ادامه تحقیق قرار گرفت و امکان ارزیابی دقیق اثر هر یک از متغیرها در بروز خطر سیل را فراهم ساخت. این رویکرد نه تنها موجب ارتقای دقت پیش‌بینی‌های مکانی گردید، بلکه قابلیت تعمیم‌پذیری آن برای سایر مناطق با شرایط مشابه نیز وجود دارد.

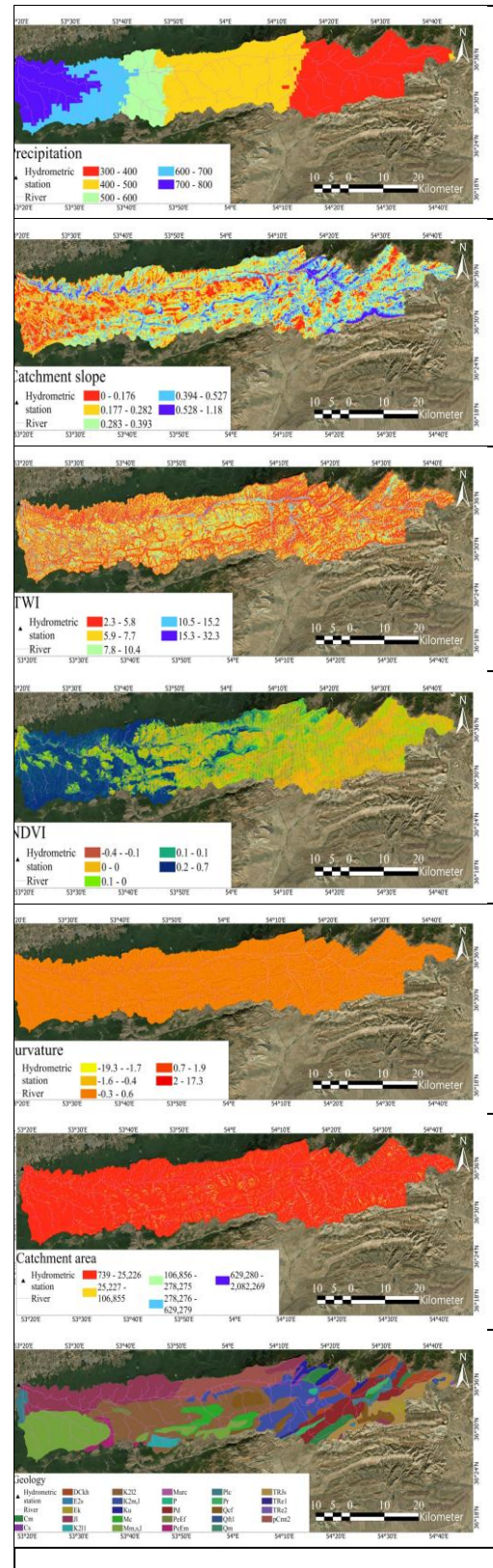


مدل‌های مورد استفاده در پژوهش

در این مطالعه، هدف ما ارزیابی و مقایسه عملکرد برخی از مدل‌های یادگیری ماشین است. ما از سه مدل، شامل ماشین بردار پشتیبان (SVM)، رگرسیون لجستیک (LR) و CatBoost (CB) استفاده کردیم. جزئیات هر مدل در ادامه شرح داده شده است:

الگوریتم ماشین بردار پشتیبان (Support Vector Machine – SVM)

یک مدل یادگیری ماشین بسیار محبوب، SVM است که به همراه مجموعه‌ای از توابع شاخص خطی، برای مسائل تعیین تابع به کار گرفته شده است و توسط Vapnik (۲۰۱۳) معرفی شده است. برای تبدیل داده‌ها در مدل SVM، از تابع سیستم ریاضی هسته استفاده شده است. با استفاده از مجموعه داده‌های آموزشی، هنگامی که تبدیل از مجموعه داده‌های SVM واقعی به فضای ویژگی با ابعاد بالا رخ داده است، یک ابرصفحه ایجاد شده است. برای تمایز فضای خروجی واقعی، بهترین ابرصفحه خطی اعمال شده است. همچنین برای دسته‌بندی داده‌ها به دو کلاس، مانند حساسیت غیر سیلابی و سیلابی {۰، ۱}، استفاده می‌شود. قابلیت مدل SVM عمدتاً به توابع هسته مناسب، مانند هسته چندجمله‌ای (PL)، هسته سیگموئیدی (SIG)، تابع پایه شعاعی (RBF) و هسته خطی (LN) بستگی دارد. چندین مطالعه (Tehrany et al., 2015; Bui et al., 2012) در بین توابع هسته مختلف برای مدل‌سازی حساسیت به سیل، که به عنوان یک تابع هسته معیار انتخاب شده است، برتری دارد. با توجه به انعطاف‌پذیری هسته پایه شعاعی برای مقابله با ابعاد مختلف مجموعه داده‌ها و توانایی تعمیم بهتر آن، عمدتاً برای مدل‌سازی حساسیت به سیل استفاده شده است (Chen et al., 2020). محدودیت‌های قابل توجه مدل‌سازی با SVM معمولاً به دشواری آن در ثبت پارامترهای حیاتی مربوط می‌شود. تابع هدف SVM به صورت زیر تعریف می‌شود:



شکل (۲): نقشه معیارهای مورد استفاده در پژوهش

اساس مجموعه‌ای از متغیرهای توضیحی برآورد می‌کند. (Rifath et al., 2024)

در این مطالعه، متغیر وابسته به صورت دودویی (۰ = عدم وقوع سیل، ۱ = وقوع سیل) تعریف شد و متغیرهای مستقل شامل عوامل توپوگرافی (ارتفاع، شیب، جهت)، هیدرولوژیکی (فاصله از رودخانه، چگالی زهکشی)، اقلیمی (بارش تجمعی)، و کاربری/پوشش زمین بودند که همگی به صورت متغیرهای پیوسته وارد مدل شدند (Ho et al., 2024). به دلیل سادگی، قابلیت تفسیر و کارایی آن در مدل‌سازی روابط بین متغیرها، به طور گسترده مورد استفاده قرار می‌گیرد. این مدل از تابع لجستیک برای نگاشت خروجی به بازه [0,1] استفاده می‌کند. معادله مدل به صورت زیر است:

$$P(y = 1|x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}}$$

که در آن ضرایب β_i و x_i متغیرهای ورودی هستند.

این مدل ساده، قابل تفسیر و مناسب برای تحلیل‌های ابتدایی و مقایسه‌ای است.

ارزیابی عملکرد و دقت مدل

دقت نقشه‌برداری حساسیت به سیل و عملکرد مدل‌های انتخاب‌شده از طریق یک فرآیند دقیق بررسی داده‌های موجودی سیل، که نشان‌دهنده سوابق سیل است، با نقشه‌های حساسیت به سیل تازه توسعه‌یافته‌ی ارزیابی شد (Rahman et al., 2023). عملکرد و دقت مدل‌های انتخاب‌شده با استفاده از تکنیک‌های Recall، F1-score، دقت و مشخصه عملکرد گیرنده (ROC) ارزیابی شد.

Recall

Recall، نسبت مناطق مستعد سیل که به طور دقیق شناسایی شده‌اند به تعداد کل مناطق مستعد سیل واقعی است (Halder et al., 2024).

معادله (۱):

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i$$

معادله تابع محدودیت (۲):

$$y_i(w \cdot \phi(x_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0$$

$$y_i(w \cdot \phi(x_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0$$

که در آن: $\phi(x)$ تابع کرنل (در اینجا RBF)؛ پارامتر جریمه برای خطاهای طبقه‌بندی؛ ξ_i متغیرهای خطا؛ y_i برچسب کلاس برای نمونه x_i هستند.

مدل گرادیان بوستینگ (GB)

الگوریتم گرادیان بوستینگ یکی از روش‌های یادگیری تقویتی است که با ایجاد مجموعه‌ای از درخت‌های تصمیم به صورت متوالی، دقت پیش‌بینی را افزایش می‌دهد. (Friedman, 2001) ایده اصلی این الگوریتم بر این اساس است که هر مدل جدید خطاهای باقی‌مانده مدل‌های قبلی را اصلاح کند.

معادله (۳):

$$h_m(x) \cdot v + F_{m-1}(x) =$$

$F_m(x)$

که در آن $F_m(x)$ مدل نهایی بعد از m تکرار، v نرخ یادگیری و $h_m(x)$ درخت تصمیم جدیدی است که خطاهای مرحله قبل را اصلاح می‌کند.

ویژگی مهم GB توانایی آن در مدل‌سازی روابط غیرخطی و شناسایی متغیرهای کلیدی مؤثر است. در این پژوهش، مدل گرادیان بوستینگ برای ترکیب داده‌های مکانی (بارندگی، شیب، ارتفاع، فاصله از رودخانه و ...) و تولید نقشه حساسیت به سیل در حوزه آبخیز نکارود به کار گرفته شد.

رگرسیون لجستیک (Logistic Regression – LR)

رگرسیون لجستیک (Logistic Regression) یک روش آماری برای طبقه‌بندی دودویی است که احتمال وقوع یک پدیده (مانند وقوع یا عدم وقوع سیل) را بر

معادله (۹): ROC-AUC =

$$\text{ROC-AUC} = \int_0^1 \text{TPR}(\text{FPR})d(\text{fpr})$$

که در آن، TP تعداد پیش‌بینی‌های مثبت صحیح (True Positives)، TN تعداد پیش‌بینی‌های منفی صحیح (True Negatives)، FP تعداد پیش‌بینی‌های مثبت نادرست (False Positives) و FN تعداد پیش‌بینی‌های منفی نادرست (False Negatives) هستند. $TPTNFPFN$ مقادیر این تکنیک‌ها، از ۰ تا ۱، نشان دهنده توانایی مدل در تمایز بین مناطق سیل‌زده و غیر سیل‌زده است و مقادیر نزدیک‌تر به ۱ نشان دهنده عملکرد پیش‌بینی برتر است (Waleed and Sajjad 2025). مجموعه داده‌ها به دو بخش آموزش و آزمون تقسیم شدند. داده‌های آموزشی برای برازش مدل و ارزیابی عملکرد آن از طریق منحنی موفقیت (success rate curve) استفاده شد، در حالی که داده‌های آزمایشی مستقل برای اعتبارسنجی توان پیش‌بینی مدل از طریق منحنی پیش‌بینی (prediction rate curve) در نقشه‌های حساسیت به سیل به کار رفتند.

بحث و نتایج

در این مطالعه، با هدف پهنه‌بندی خطر سیل در حوزه آبخیز نکارود، سه الگوریتم یادگیری ماشین شامل SVM، LR و مدل GB مورد استفاده و مقایسه قرار گرفتند. مدل‌ها با استفاده از مجموعه داده‌ای شامل ۱۵ متغیر محیطی و ژئومورفولوژیکی آموزش دیده و سپس ارزیابی شدند.

نتایج ماشین بردار پشتیبان:

داده‌های ژئومورفولوژیکی، هیدرولوژیکی، اقلیمی، زمین‌شناسی و کاربری اراضی پس از پیش‌پردازش و مقیاس‌بندی با استفاده از روش‌های StandardScaler و MinMaxScaler، در سطح نمونه‌ها (پیکسل‌ها/نقاط رخداد سیل) به صورت تصادفی به دو مجموعه آموزش (۷۰٪) و آزمون

معادلهمعادله (۵): Recall =

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

F1-score

F1-score تعادل یک مدل را بر اساس دقت و یادآوری اندازه‌گیری می‌کند و با به حداقل رساندن مثبت‌ها و منفی‌های کاذب، به شناسایی دقیق مناطق مستعد سیل کمک می‌کند (Seydi et al., 2022).

معادلهمعادله (۶): F1 score =

$$\text{AF1 score} = \frac{2 * \text{TP}}{2 * \text{TP} + \text{FP} + \text{FN}}$$

دقت کلی (Overall Accuracy)

دقت، نسبت پیش‌بینی‌های صحیح را برای همه مناطق اندازه‌گیری می‌کند و نشان می‌دهد که چقدر خوب بین مناطق با حساسیت بالا و پایین تمایز قائل می‌شود. این، ارزیابی سطحی از دقت پیش‌بینی مدل و توانایی آن در طبقه‌بندی صحیح مناطق را ارائه می‌دهد (Waleed and Sajjad., 2025). دقت کلی با استفاده از معادله زیر محاسبه می‌شود:

معادله (۷):

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

مشخصه عملیاتی گیرنده (ROC)

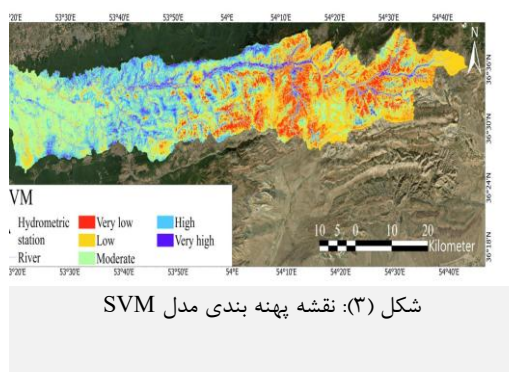
از تکنیک ROC (Receiver Operating Characteristic) برای ارزیابی عملکرد و دقت مدل با محاسبه AUC استفاده شد. تکنیک ROC یکی از پرکاربردترین تکنیک‌ها برای ارزیابی عملکرد و دقت مدل‌های پیش‌بینی است (Rahman et al., 2023). منحنی ROC، نرخ مثبت واقعی (حساسیت) را در برابر نرخ مثبت کاذب (۱-ویژگی) در مقادیر آستانه مختلف رسم می‌کند و ارزیابی جامعی از دقت مدل ارائه می‌دهد.

معادله (۸): FPR/1-specificity =

$$\frac{\text{FP}}{\text{FP} + \text{TN}}$$

$$1 - \text{specificity} = \frac{\text{FP}}{\text{FP} + \text{TN}}$$

سطح کل منطقه (حدود ۸/۹۶۱ هکتار) را در بر می‌گیرد. با وجود مساحت محدود، این نواحی باید به‌عنوان کانون‌های بحرانی در برنامه‌ریزی و اقدامات پیشگیرانه مورد توجه ویژه قرار گیرند. به‌طور معمول، این نواحی همپوشانی بالایی با مناطق پرشیب، کاربری‌های حساس یا مسیرهای هیدرولوژیک اصلی دارند.



جدول (۱): مساحت طبقه‌های مختلف پهنه بندی سیل
مدل SVM

کلاس	تعداد پیکسل	مساحت	درصد
خیلی کم	۲۶۳۸۸۳	۱۹۴۶۹/۶۱	۱۰/۳۰
کم	۷۸۶۳۹۲	۵۸۰۲۰/۹۷	۳۰/۷۰
متوسط	۹۰۶۸۵۱	۶۶۹۰۸/۵۸	۳۵/۴۱
زیاد	۴۸۲۸۹۷	۳۵۶۲۸/۷۳	۱۸/۸۵
خیلی زیاد	۱۲۱۴۵۸	۸۹۶۱/۳۲	۴/۷۴

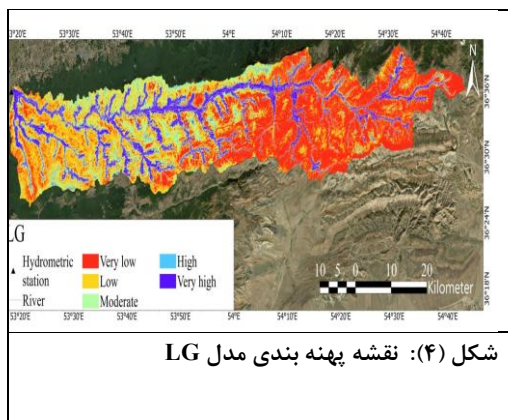
مدل رگرسیون لجستیک

در این مدل داده‌های ژئومورفولوژیکی، اقلیمی، هیدرولوژیکی، زمین‌شناسی و کاربری اراضی پس از پیش‌پردازش و استانداردسازی با StandardScaler به دو بخش آموزشی (۷۰٪) و آزمایشی (۳۰٪) تقسیم شدند. برای آموزش مدل، ابرپارامترهای کلیدی شامل $\text{solver}=\text{lbfgs}$ ، $C=1.0$ ، $\text{penalty}=\text{L2}$ و $\text{multi_class}=\text{multinomial}$ ، $\text{max_iter}=500$

(۳۰٪) تقسیم شدند. این تفکیک بدون جداسازی فضایی حوزه آبخیز انجام شد و هر دو مجموعه کل محدوده مطالعاتی را پوشش می‌دهند. مدل با هسته‌های مختلف بررسی شد و در نهایت هسته RBF با مقادیر پیش‌فرض ابرپارامترها شامل $C=1.0$ ، $\gamma=1.0$ ، $\text{scale}=\text{scale}$ ، $\text{gamma}=\text{gamma}$

و حداکثر تکرار نامحدود بهترین عملکرد را ارائه داد. نتایج نقشه‌های پهنه‌بندی نشان داد که SVM توانسته است الگوهای پیچیده داده‌ها را به خوبی شناسایی کرده و مناطق با خطر بالای سیل را با دقت مناسبی طبقه‌بندی کند.

بر اساس نتایج مدل SVM، طبقه‌بندی منطقه مورد مطالعه به پنج سطح «خیلی کم»، «کم»، «متوسط»، «زیاد» و «خیلی زیاد» انجام شده است (شکل ۳) (جدول ۱). یافته‌ها نشان می‌دهد که گسترده‌ترین سطح مربوط به کلاس متوسط است که با مساحت حدود ۶۶/۹۰۸ هکتار، حدود ۳۵/۴۱ درصد از کل منطقه را پوشش می‌دهد. این امر نشان می‌دهد که بخش قابل توجهی از منطقه در وضعیت نیمه‌حساس یا دارای خطر متوسط قرار دارد که می‌تواند از نظر مدیریتی به عنوان پهنه‌های نیازمند پایش مستمر در نظر گرفته شود. پس از آن، کلاس «کم» با ۳۰/۷۰ درصد (حدود ۵۸/۰۲۱ هکتار) در رتبه دوم قرار دارد. این گستره زیاد در طبقه کم نشان می‌دهد که بخش‌های قابل ملاحظه‌ای از منطقه دارای حساسیت پایین بوده و احتمال وقوع پدیده مورد بررسی در آن کمتر است. در مقابل، پهنه‌های با خطر یا حساسیت زیاد و خیلی زیاد مجموعاً حدود ۲۳/۵۹ درصد (بیش از ۴۴/۵۹۰ هکتار) از منطقه را تشکیل می‌دهند که اگرچه از نظر سطحی کمتر از دو کلاس اول هستند، اما به دلیل اثرگذاری بالقوه بر سکونتگاه‌ها، زیرساخت‌ها و منابع طبیعی، دارای اهمیت راهبردی در مدیریت ریسک هستند. کمترین سهم به طبقه خیلی زیاد اختصاص دارد که تنها ۴/۷۴ درصد از



جدول (۲): مساحت طبقه های مختلف پهنه بندی سیل
مدل LG

کلاس	تعداد پیکسل	مساحت	درصد
خیلی کم	۹۹۰۶۵۵	۷۳۰۹۱/۷۴	۳۸/۶۸
کم	۷۵۴۲۰۹	۵۵۶۴۶/۴۷	۲۹/۴۵
متوسط	۳۸۸۹۹۸	۲۸۷۰۰/۷۵	۱۵/۱۹
زیاد	۹۰۸۱۴	۶۷۰۰/۳۷	۳/۵۵
خیلی زیاد	۳۳۶۸۰۵	۲۴۸۴۹/۸۹	۱۳/۱۵
جمع کل	۲۵۶۱۴۸۱	۱۸۸۹۸۹/۲۱	۱۰۰

مدل گرادیان بوستینگ

این مدل با ساخت متوالی درخت های تصمیم گیری و یادگیری از خطاهای مدل های قبلی، توانست الگوهای پیچیده مرتبط با وقوع سیل را شناسایی کند. داده های محیطی شامل متغیرهای ژئومورفولوژیکی، اقلیمی، هیدرولوژیکی، زمین شناسی و کاربری اراضی پس از پیش پردازش به دو مجموعه آموزشی (۷۰٪) و آزمایشی (۳۰٪) تقسیم شدند. برای آموزش مدل، ابر پارامترهای کلیدی شامل $n_estimators=1000$ ، $learning_rate=0.01$ ، $max_depth=20$ ، $min_samples_split=5$ ، $min_samples_leaf=2$ و $subsample=0.8$ و $max_features=sqrt$ این پیکربندی به مدل امکان داد تا با مدیریت روابط غیرخطی و کاهش بیش برآزش، پهنه بندی دقیقی از خطر سیل ارائه دهد و مبنای مناسبی برای تحلیل الگوهای مکانی مناطق پرخطر فراهم آورد.

این پیکربندی باعث شد مدل با کاهش بیش برآزش و حفظ تعادل بین دقت و سادگی، رابطه ای خطی میان متغیرهای ورودی و احتمال وقوع سیل ایجاد کند. با وجود محدودیت در مقایسه با مدل های پیچیده تر، LR به عنوان روشی ساده، قابل تفسیر و کارآمد توانست نتایج مناسبی در طبقه بندی چندکلاسه خطر سیل ارائه دهد.

خروجی مدل LG نشان می دهد که (شکل ۴) (جدول ۲)، بخش اعظم منطقه در طبقه خیلی کم قرار گرفته است؛ به طوری که با ۹۹۰/۶۵۵ پیکسل و مساحت حدود ۷۳/۰۹۲ هکتار، معادل ۳۸/۶۸ درصد از کل محدوده را پوشش می دهد. پس از آن، بیشترین سهم به کلاس کم با ۷۵۴/۲۰۹ درصد (حدود ۵۵/۶۴۶ هکتار) تعلق دارد. این دو کلاس در مجموع حدود ۶۸ درصد مساحت منطقه را شامل می شوند که نشان دهنده غلبه پهنه های با خطر پایین است. در مقابل، طبقات زیاد و خیلی زیاد در مجموع حدود ۱۶/۷ درصد مساحت منطقه را تشکیل داده اند. در این میان، کلاس خیلی زیاد با ۳۳۶/۸۰۵ پیکسل و مساحت ۲۴/۸۵۰ هکتار (۱۳/۱۵ درصد)، بیش از سه برابر کلاس زیاد (۳/۵۵ درصد) سهم دارد و به عنوان کانون های اصلی وقوع سیل در نظر گرفته می شود. این امر بیانگر آن است که اگرچه گستره پهنه های با خطر بالا محدودتر از پهنه های کم خطر است، اما شدت تهدید در این نواحی به مراتب بیشتر بوده و باید در اولویت اقدامات مدیریتی قرار گیرند. طبقه متوسط نیز با سهم ۱۵/۱۹ درصد (۲۸/۷۰۱ هکتار) نشان دهنده مناطقی با حساسیت میان سطحی است که بسته به شرایط بارش و ویژگی های هیدرولوژیکی می تواند به نواحی پرخطر تبدیل شود.

جدول (۳): مساحت طبقه های مختلف پهنه بندی سیل

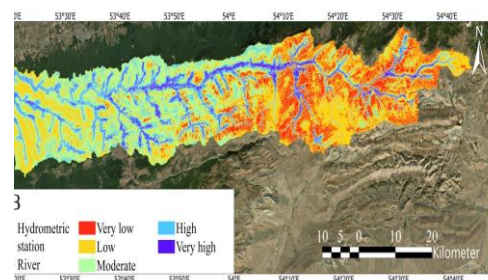
درصد	مساحت	تعداد پیکسل	کلاس
۱۳/۸۳	۲۶۱۴۲/۸۳	۳۵۴۳۲۹	خیلی کم
۳۷/۲۳	۷۰۳۴۹/۲۲	۹۵۳۴۸۴	کم
۲۹/۵۳	۵۵۸۱۶/۶۱	۷۵۶۵۱۵	متوسط
۱۴/۴۰	۲۷۲۱۸/۸۵	۳۶۸۹۱۳	زیاد
۵/۰۱	۹۴۶۱/۷۰	۱۲۸۲۴۰	خیلی زیاد
۱۰۰	۱۸۸۹۸۹/۲۱	۲۵۶۱۴۸۱	جمع کل

مدل GB

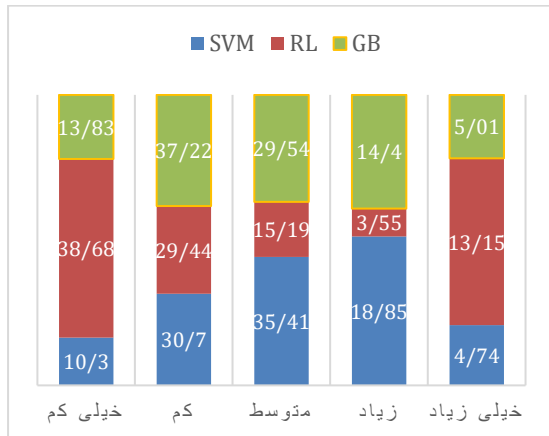
مقایسه نتایج

بررسی نتایج سه مدل نشان می‌دهد که هر یک از الگوریتم‌ها در شناسایی و توزیع پهنه‌های خطر سیل الگوهای متفاوتی ارائه کرده‌اند (شکل ۶) (جدول ۴). مدل SVM تمرکز بیشتری بر طبقات میانی دارد؛ به طوری که بیشترین سهم را به کلاس متوسط (۳۵/۴۱ درصد) اختصاص داده و در کنار آن، کلاس‌های کم (۳۰/۷۰ درصد) و زیاد (۱۸/۸۵ درصد) نیز سهم بالایی دارند. این الگو بیانگر آن است که SVM منطقه را با نگاه متوازن تری میان طبقات مختلف توزیع کرده و حساسیت بالاتری به نواحی با خطر متوسط تا زیاد دارد. مدل RL رفتاری متضاد با SVM نشان می‌دهد؛ بیشترین سطح در کلاس خیلی کم (۳۸/۶۸ درصد) قرار گرفته و پس از آن کلاس کم (۲۹/۴۴ درصد) در رتبه دوم است. این دو طبقه در مجموع حدود ۶۸ درصد مساحت را شامل می‌شوند که نشان‌دهنده غلبه مناطق کم‌خطر در خروجی این مدل است. در مقابل، سهم کلاس خیلی زیاد در RL نسبتاً بالاتر (۱۳/۱۵ درصد) است که بیانگر تمایل این مدل به برجسته‌سازی نواحی بحرانی و تمرکز بر تشخیص کانون‌های اصلی خطر است. مدل GB رویکردی میانه‌رو بین دو مدل دیگر اتخاذ کرده است. بیشترین سهم مربوط به کلاس کم (۳۷/۲۲ درصد) و پس از آن متوسط (۲۹/۵۴ درصد) است و در مجموع نزدیک به ۶۷ درصد منطقه را در بر می‌گیرند. سهم

بر اساس نتایج مدل GB (شکل ۵) (جدول ۳)، توزیع مکانی خطر سیل در پنج طبقه مشخص شد. بیشترین سهم مربوط به طبقه کم است که با مساحت حدود ۷۰/۳۴۹ هکتار، معادل ۳۷/۲۳ درصد از کل منطقه را پوشش می‌دهد. این موضوع نشان می‌دهد که بخش وسیعی از منطقه در معرض خطر پایین قرار دارد، اما همچنان در صورت وقوع بارش‌های شدید می‌تواند تحت تأثیر قرار گیرد. پس از آن، طبقه با مساحت ۵۵/۸۱۷ هکتار، حدود ۲۹/۵۴ درصد از کل منطقه را شامل می‌شود که بیانگر شرایط نسبتاً متعادل ولی پریسکتر از کلاس کم است. از سوی دیگر، حدود ۱۹/۴۱ درصد از منطقه (شامل کلاس‌های زیاد و خیلی زیاد) در محدوده‌های پرخطر واقع شده‌اند. کلاس زیاد با مساحت ۲۷/۲۱۹ هکتار، سهم ۱۴/۴۰ درصدی دارد، در حالی که طبقه خیلی زیاد با مساحت ۹/۴۶۲ هکتار، حدود ۵/۰۱ درصد از کل منطقه را تشکیل می‌دهد. هرچند این درصد نسبتاً کوچک به نظر می‌رسد، اما همین نواحی کانون‌های بحرانی و اصلی وقوع سیل بوده و بیشترین تهدید را برای سکونتگاه‌ها، زیرساخت‌ها و منابع طبیعی ایجاد می‌کنند. کمترین سهم مربوط به طبقه خیلی کم است که با مساحت ۲۶/۱۴۳ هکتار، حدود ۱۳/۸۳ درصد از منطقه را در بر می‌گیرد. این پهنه‌ها عموماً مناطق با شیب ملایم یا دارای شرایط هیدرولوژیکی پایدار هستند که احتمال وقوع سیل در آنها اندک است.



شکل (۵): نقشه پهنه بندی مدل GB



شکل (۶): مساحت طبقه های مختلف پهنه بندی سیل مدل ها (به درصد)

ارزیابی عملکرد مدل ها

نتایج حاصل از مقایسه سه مدل رگرسیون لجستیک (LR)، گرادیان بوستینگ (GB) و ماشین بردار پشتیبان (SVM) در پهنه بندی خطر سیل نشان داد که عملکرد این مدل ها در تفکیک مناطق پرخطر و کم خطر متفاوت است. در میان مدل های مورد بررسی، مدل گرادیان بوستینگ با دقت کلی (Accuracy) معادل ۹۶.۴۵٪، بالاترین عملکرد را نشان داد. همچنین، مقادیر بالای ۹۴.۹۳٪ Precision، ۹۳.۲۰٪ Recall (و ۹۴.۰۵٪ F1-Score) (در این مدل بیانگر تعادل مناسب میان شناسایی صحیح مناطق پرخطر و کاهش خطاهای طبقه بندی است. مدل SVM با دقت کلی ۸۷.۲۰٪ عملکرد متوسطی از خود نشان داد. اگرچه مقادیر ۸۷.۲۰٪ Precision و ۸۶.۲۰٪ Recall (در این مدل قابل قبول است، اما کاهش مقدار ۸۶.۶۹٪ F1-Score (نسبت به مدل GB نشان دهنده محدودیت این مدل در تفکیک هم زمان و متوازن مناطق پرخطر و کم خطر می باشد.

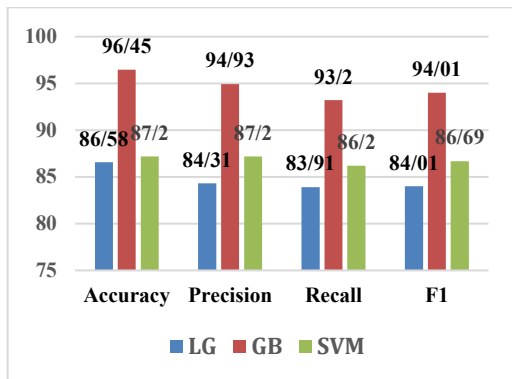
مدل رگرسیون لجستیک (LR) با مقادیر Accuracy (۸۶.۵۸٪)، Precision (۸۴.۳۱٪)، Recall (۸۳.۹۱٪) و F1-Score (۸۴.۰۱٪) (ضعیف ترین عملکرد را در میان سه مدل مورد بررسی داشت. این

طبقات زیاد و خیلی زیاد در این مدل به ترتیب ۱۴/۱۴، ۴۰/۴۰ درصد و ۵/۰۱ درصد است که کمتر از SVM اما بیشتر از RL در طبقه زیاد قرار می گیرد. این موضوع نشان می دهد که GB نسبت به SVM محافظه کارانه تر عمل کرده اما در مقایسه با RL برآورد متعادل تری از طبقات پرخطر دارد.

به طور کلی، نتایج بیانگر آن است که SVM در شناسایی مناطق با خطر متوسط تا زیاد عملکرد برجسته تری داشته، RL بر نواحی با خطر خیلی کم و خیلی زیاد تأکید بیشتری دارد، و GB بینابینی ترین الگو را ارائه کرده است. بنابراین، استفاده از یک رویکرد تلفیقی یا وزنی میان این سه مدل می تواند به کاهش عدم قطعیت و افزایش قابلیت اطمینان در نقشه های نهایی خطر سیل کمک کند.

جدول (۴): مساحت طبقه های مختلف پهنه بندی سیل مدل ها (به درصد)

کلاس	مدل SVM	مدل RL	مدل GB
خیلی کم	۱۰/۳۰	۳۸/۶۸	۱۳/۸۳
کم	۳۰/۷۰	۲۹/۴۴	۳۷/۲۲
متوسط	۳۵/۴۱	۱۵/۱۹	۲۹/۵۴
زیاد	۱۸/۸۵	۳/۵۵	۱۴/۴۰
خیلی زیاد	۴/۷۴	۱۳/۱۵	۵/۰۱
جمع کل	۱۰۰	۱۰۰	۱۰۰



شکل (۷): دقت مختلف مدل های مورد استفاده در پهنه بندی سیل

منحنی ROC

در این پژوهش، دقت مدل های یادگیری ماشین مورد استفاده در پهنه بندی خطر سیل با بهره گیری از منحنی ROC و شاخص AUC مورد ارزیابی قرار گرفت. منحنی ROC با ترسیم نرخ مثبت واقعی (TPR) در برابر نرخ مثبت کاذب (FPR)، توانایی مدل ها در تفکیک مناطق پرخطر از کم خطر را نمایش می دهد. نتایج نشان داد که الگوریتم GB با مقدار AUC برابر با ۸۹/۶۳٪ بهترین عملکرد را در میان مدل های مورد بررسی داشته و توانسته است با دقت بالاتری نسبت به سایر مدل ها مناطق مستعد سیل را شناسایی نماید. پس از آن، مدل SVM با AUC معادل ۸۲/۹۲٪ عملکرد نسبتاً قابل قبولی ارائه داد، هرچند که در مقایسه با GB از قدرت تفکیک پایین تری برخوردار بود. در مقابل، مدل LR با AUC برابر با ۸۱/۷۹٪ کمترین دقت را در میان سه مدل نشان داد. به طور کلی، یافته ها حاکی از آن است که استفاده از الگوریتم های پیشرفته ای همچون گرادیان بوستینگ می تواند نقش مؤثری در بهبود کارایی پهنه بندی خطر سیل ایفا کند، در حالی که مدل های ساده تر مانند رگرسیون لجستیک در شرایط مشابه از دقت پایین تری برخوردارند. این نتایج اهمیت انتخاب مدل های یادگیری ماشین مناسب را در مطالعات مرتبط با مدیریت بحران و ارزیابی ریسک سیل برجسته می سازد.

نتایج حاکی از آن است که مدل های خطی کلاسیک در مقایسه با الگوریتم های یادگیری ماشین پیشرفته تر، توانایی محدودتری در بازنمایی الگوهای پیچیده مرتبط با وقوع سیل دارند.

برتری عملکرد مدل GB نسبت به LR و SVM را می توان به ماهیت ساختاری این مدل نسبت داد. در حالی که رگرسیون لجستیک مبتنی بر فرض روابط خطی میان متغیرهای مستقل و متغیر هدف است و عملکرد SVM به انتخاب کرنل و تنظیم پارامترها وابسته می باشد، گرادیان بوستینگ با بهره گیری از مجموعه ای از درخت های تصمیم تقویتی قادر است روابط غیرخطی و برهم کنش های پیچیده میان عوامل مؤثر بر سیل را بهتر بازنمایی کند. از این رو، عملکرد برتر این مدل می تواند ناشی از توان بالاتر آن در مدل سازی ساختار پیچیده داده های محیطی باشد. به طور کلی، نتایج نشان می دهد که در میان مدل های مورد بررسی، گرادیان بوستینگ گزینه ای کارآمدتر برای پهنه بندی خطر سیل در منطقه مطالعاتی محسوب می شود، در حالی که SVM و به ویژه رگرسیون لجستیک به دلیل محدودیت در بازنمایی روابط غیرخطی، عملکرد ضعیف تری از خود نشان داده اند.

جدول (۵): دقت مدل های مورد استفاده در پهنه بندی سیل حوزه آبخیز نکارود

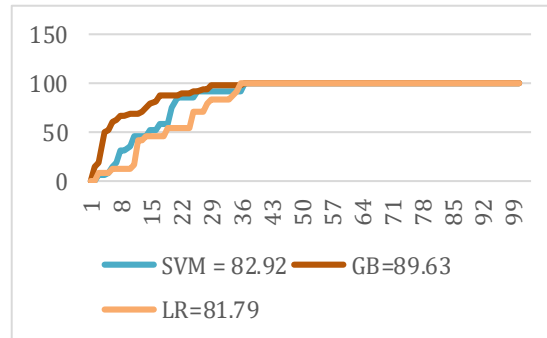
مدل	LG	GB	SVM
Accuracy	۸۶/۵۸	۹۶/۴۵	۸۷/۲۰
Precision	۸۴/۳۱	۹۴/۹۳	۸۷/۲۰
Recall	۸۳/۹۱	۹۳/۲۰	۸۶/۲۰
f1	۸۴/۰۱	۹۴/۰۵	۸۶/۶۹

نه تنها در شناسایی مناطق پرخطر بلکه در تحلیل کمی سهم هر متغیر مؤثر نیز به کار گرفته شود.

نتایج این پژوهش با یافته‌های بسیاری از مطالعات پیشین در زمینه پهنه‌بندی خطر سیل هم‌راستا است. برای مثال، مطالعات متعددی گزارش کرده‌اند که الگوریتم‌های مبتنی بر درخت تصمیم و تقویت یادگیری، نظیر گرادیان بوستینگ، در مقایسه با مدل‌های خطی و روش‌های کلاسیک، عملکرد بهتری در پیش‌بینی مناطق سیلابی ارائه می‌دهند (e.g., Tehrany et al., 2015; Pham et al., 2020; Arabameri et al., 2021) و کرمی و همکاران، ۱۴۲۰ و حنیفی نیا، عبقری ۱۴۰۰ مقادیر AUC گزارش شده برای مدل GB در این تحقیق (۰.۸۹۶) با دامنه مقادیر ارائه شده در مطالعات مشابه (۰.۸۵-۰.۹۳) قابل مقایسه بوده و بیانگر کارایی مناسب این مدل در محیط‌های کوهستانی و نیمه‌مرطوب است.

از نظر اهمیت متغیرها، نتایج این پژوهش نشان داد که بارندگی مهم‌ترین عامل مؤثر بر وقوع سیل است که این یافته با نتایج مطالعات پیشین در حوزه‌های آبخیز شمال ایران و مناطق با اقلیم مرطوب هم‌خوانی دارد (e.g., Rahmati et al., 2016; Khosravi et al., 2018). همچنین، نقش قابل توجه شیب، فاصله از رودخانه و ارتفاع در افزایش حساسیت سیلابی، در پژوهش‌های متعدد به‌عنوان عوامل کلیدی کنترل‌کننده فرآیندهای رواناب و تمرکز جریان گزارش شده است. این هم‌راستایی نشان می‌دهد که مدل GB توانسته است الگوهای فیزیکی حاکم بر رخداد سیل را به‌درستی بازنمایی کند.

در مجموع، نتایج این تحقیق در حوزه آبخیز نکارود استان مازندران نشان می‌دهد که نقشه‌های حساسیت تولیدشده می‌توانند به‌عنوان ابزار تصمیم‌یار برای سازمان‌های مدیریت بحران، مدیران منابع آب و برنامه‌ریزان شهری به کار گرفته شوند. این نقشه‌ها می‌توانند در جلوگیری از توسعه در مناطق پرخطر، بهبود سامانه‌های هشدار سریع و کاهش خسارات



شکل (۸): نمودار منحنی ROC

نتیجه‌گیری

نتایج این پژوهش نشان داد که استفاده از الگوریتم‌های یادگیری ماشین می‌تواند دقت پهنه‌بندی خطر سیل را به‌طور چشمگیری ارتقا دهد. ارزیابی عملکرد مدل‌ها نشان داد که الگوریتم GB با مقدار AUC برابر با ۰/۸۹۶، ضریب کاپا ۰/۸۴ و دقت کلی ۰/۸۷ بهترین نتایج را در پیش‌بینی مناطق سیلابی و غیرسیلابی به‌دست آورد. مدل SVM با AUC برابر ۰/۸۷۲، کاپا ۰/۷۹ و دقت ۰/۸۳ در رتبه دوم قرار گرفت، در حالی که مدل LR با AUC برابر ۰/۸۵۳، کاپا ۰/۷۴ و دقت ۰/۸۰ ضعیف‌ترین عملکرد را داشت. این اختلاف عملکرد نشان می‌دهد که مدل‌های مبتنی بر تقویت یادگیری (Boosting) توانایی بیشتری در بازنمایی روابط غیرخطی و پیچیده میان عوامل محیطی و هیدرولوژیکی دارند. بررسی اهمیت متغیرها نیز نشان داد که بارندگی (وزن نسبی ۰/۲۸) بیشترین تأثیر را بر وقوع سیل دارد و پس از آن شیب (۰/۲۱)، فاصله از رودخانه (۰/۱۹) و ارتفاع (۰/۱۵) به‌ترتیب در رتبه‌های بعدی قرار گرفتند. متغیرهای کاربری اراضی و نفوذپذیری خاک نیز نقش تکمیلی اما معناداری در افزایش یا کاهش حساسیت منطقه داشتند. این یافته‌ها بیانگر آن است که ترکیب داده‌های مکانی با مدل‌های یادگیری ماشین می‌تواند

عمیق استفاده شود تا هم دقت و هم قابلیت تعمیم نتایج افزایش یابد.
قدردانی: این مقاله مستخرج از رساله دکترا بوده لذا نویسندگان بر خود لازم می دانند تا از معاونت پژوهشی دانشگاه علوم کشاورزی و منابع طبیعی ساری سپاسگزاری نمایند.

انسانی و اقتصادی ناشی از سیل نقش بسزایی ایفا کنند. با وجود این، محدودیت‌هایی همچون وابستگی به دقت داده‌های ورودی و نیاز به اعتبارسنجی در مناطق دیگر وجود دارد. بنابراین، پیشنهاد می‌شود در مطالعات آینده از داده‌های چندزمانه، تصاویر ماهواره‌ای با قدرت تفکیک بالاتر و مدل‌های یادگیری

منابع:

- قره داغی، ب و ا. قاسم زاده. (۱۴۰۳). رویکرد پیش‌بینی جدید با استفاده از ترکیب یادگیری ماشین برای پیش بینی مناطق حساس به وقوع سیل (مطالعه موردی: حوضه آبریز کارون). نشریه سنجش از دور و GIS ایران، دوره ۱۶، شماره ۲، ص ۱-۱۸
- حنیفی نیا، ع و ه. عبقری. (۱۴۰۴). پیش‌بینی مناطق مستعد سیل با استفاده از مدل‌های یادگیری ماشین خطی تعمیم یافته و بیشینه آنتروپی. مخاطرات محیط طبیعی، دوره ۱۴، شماره ۴۳، ص ۳۴-۱۹.
- کرمی، پ، اسلامی نژاد، س. ا، افتخاری، م، اکبری، م و م. راستگو. (۱۴۰۲). پهنه‌بندی حساسیت سیل با استفاده از روش‌های یادگیری ماشین بهبود یافته توسط الگوریتم ژنتیک. نشریه محیط زیست طبیعی، دوره ۷۶، شماره ۱، ص ۴۳-۶۰.
- سلیمانی، ک و ب. محسنی. (۱۴۰۳). تعیین اولویت پهنه بندی خطر وقوع سیل با استفاده از سیستم اطلاعات جغرافیایی در حوزه آبخیز بابلرود (استان مازندران). نشریه سنجش از دور و سیستم اطلاعات جغرافیایی در منابع طبیعی، دوره ۲، شماره ۱۵، ص ۲۸-۵۳.
- شادمانی، م، معروفی، ص، سبزی پرور، ع. ا و ک. محمدی. (۱۳۸۷). مدل ژئومورفولوژیکی سیلاب و پهنه بندی خطر سیل با استفاده از شبکه عصبی مصنوعی و GIS. کنفرانس بین المللی بحران آب. SID. <https://sid.ir/paper/814344/fa>
- Ali, R., Kuriqi, A., Abubaker, S., & Kisi, O. (2019). Long-term trends and seasonality detection of the observed flow in Yangtze River using Mann-Kendall and Sen's innovative trend method. *Water*, 11(9), 1855.
- Arabameri, A., Pradhan, B., Rezaei, K., et al. (2021). Comparison of multi-criteria decision making and machine learning methods for flood susceptibility mapping. *Journal of Hydrology*, 597, 126164.
- Botzen, W. W., de Boer, J., & Terpstra, T. (2013). Framing of risk and preferences for annual and multi-year flood insurance. *Journal of economic psychology*, 39, 357-375.
- Bui, D. T., Hoang, N. D., Martínez-Álvarez, F., Ngo, P. T. T., Hoa, P. V., Pham, T. D., ... & Costache, R. (2020). A novel deep learning neural network approach for predicting flash flood susceptibility: A case study at a high frequency tropical storm area. *Science of The Total Environment*, 701, 134413.
- Bui, D. T., Pradhan, B., & Lofman, O. Revhaug. I., 2012. Landslide susceptibility assessment in vietnam using support vector machines, decision tree, and Naive Bayes Models. *Mathematical Problems in Engineering*, 2012.
- Chen, J., Li, Q., Wang, H., & Deng, M. (2020). A machine learning ensemble approach based on random forest and radial basis function neural network for risk evaluation of regional

flood disaster: a case study of the Yangtze River Delta, China. *International journal of environmental research and public health*, 17(1), 49.

Costache, R. (2019). Flash-flood Potential Index mapping using weights of evidence, decision Trees models and their novel hybrid integration. *Stochastic Environmental Research and Risk Assessment*, 33(7), 1375-1402.

Costache, R., Hong, H., & Pham, Q. B. (2020). Comparative assessment of the flash-flood potential within small mountain catchments using bivariate statistics and their novel hybrid integration with machine learning models. *Science of the Total Environment*, 711, 134514.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.

Hoa, P. V., Binh, N. A., Hong, P. V., An, N. N., Thao, G. T. P., Hanh, N. C., ... & Bui, D. T. (2024). One-dimensional deep learning driven geospatial analysis for flash flood susceptibility mapping: a case study in North Central Vietnam. *Earth Science Informatics*, 17(5), 4419-4440.

Islam, A. R. M. T., Talukdar, S., Mahato, S., Kundu, S., Eibek, K. U., Pham, Q. B., ... & Linh, N. T. T. (2021). Flood susceptibility modelling using advanced ensemble machine learning models. *Geoscience Frontiers*, 12(3), 101075.

Kaur, H., Gupta, S., Parkash, S., Thapa, R., & Mandal, R. (2017). Geospatial modelling of flood susceptibility pattern in a subtropical area of West Bengal, India. *Environmental Earth Sciences*, 76(9), 339.

Kia, M. B., Pirasteh, S., Pradhan, B., Mahmud, A. R., Sulaiman, W. N. A., & Moradi, A. (2012). An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia. *Environmental earth sciences*, 67(1), 251-264.

Kubra, A., Lokesh, K. J., Shakthi, G., Surya, M., Mohammad Jawad, A., & Melquier, A. (2024, October). Enhanced Flood Prediction and Management for Urban Underpass Using Machine Learning. In *International Conference on Computer & Communication Technologies* (pp. 385-395). Singapore: Springer Nature Singapore.

Meliho, M., Khatlaji, A., & Asinyo, J. (2021). Spatial modeling of flood susceptibility using machine learning algorithms. *Arabian Journal of Geosciences*, 14(21), 2243.

Parizi, E., Khojeh, S., Hosseini, S. M., & Moghadam, Y. J. (2022). Application of Unmanned Aerial Vehicle DEM in flood modeling and comparison with global DEMs: Case study of Atrak River Basin, Iran. *Journal of environmental management*, 317, 115492.

Parsian, S., Amani, M., Moghimi, A., Ghorbanian, A., & Mahdavi, S. (2021). Flood hazard mapping using fuzzy logic, analytical hierarchy process, and multi-source geospatial datasets. *Remote Sensing*, 13(23), 4761.

Pham, B. T., Prakash, I., Khosravi, K., et al. (2020). A comparison of machine learning methods for flood susceptibility mapping: A case study of India. *Science of the Total Environment*, 701, 134413.

Pollack, A. B., Sue Wing, I., & Nolte, C. (2022). Aggregation bias and its drivers in large-scale flood loss estimation: A Massachusetts case study. *Journal of Flood Risk Management*, 15(4), e12851.



- Pourghasemi, H. R., Razavi-Termeh, S. V., Kariminejad, N., Hong, H., & Chen, W. (2020). An assessment of metaheuristic approaches for flood assessment. *Journal of Hydrology*, 582, 124536.
- Rahman, Z. U., Ullah, W., Bai, S., Ullah, S., Jan, M. A., Khan, M., & Tayyab, M. (2023). GIS-based flood susceptibility mapping using bivariate statistical model in Swat River Basin, Eastern Hindukush region, Pakistan. *Frontiers in Environmental Science*, 11, 1178540.
- Rahmati, O., Pourghasemi, H. R., & Melesse, A. M. (2016). Application of GIS-based data-driven random forest and maximum entropy models for flood susceptibility mapping. *Geomatics, Natural Hazards and Risk*, 7(4), 1309–1331.
- Rifath, A. R., Muktadir, M. G., Hasan, M., & Islam, M. A. (2024). Flash flood prediction modeling in the hilly regions of Southeastern Bangladesh: A machine learning attempt on present and future climate scenarios. *Environmental Challenges*, 17, 101029.
- Saleh, A., Yuzir, A., & Sabtu, N. (2022). Flash flood susceptibility mapping of sungai pinang catchment using frequency ratio. *Sains Malaysiana*, 51(1), 51-65.
- Samanta, S., Pal, D. K., & Palsamanta, B. (2018). Flood susceptibility analysis through remote sensing, GIS and frequency ratio model. *Applied Water Science*, 8(2), 66.
- Serraj, R., & Pingali, P. (Eds.). (2018). *Agriculture & food systems to 2050: global trends, challenges and opportunities*.
- Seydi, S. T., Kanani-Sadat, Y., Hasanlou, M., Sahraei, R., Chanussot, J., & Amani, M. (2022). Comparison of machine learning algorithms for flood susceptibility mapping. *Remote Sensing*, 15(1), 192.
- Tehrany, M. S., Pradhan, B., & Jebur, M. N. (2015). Flood susceptibility mapping using a novel ensemble approach of evidential belief function and support vector machine. *Journal of Hydrology*, 512, 332–343.
- Tehrany, M. S., Pradhan, B., Mansor, S., & Ahmad, N. (2015). Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. *Catena*, 125, 91-101.
- Vapnik, V. (2013). *The nature of statistical learning theory*. Springer science & business media.
- Waleed, M., & Sajjad, M. (2025). Advancing flood susceptibility prediction: A comparative assessment and scalability analysis of machine learning algorithms via artificial intelligence in high-risk regions of Pakistan. *Journal of Flood Risk Management*, 18(1), e13047.
- Zeng, Z., Tang, G., Hong, Y., Zeng, C., & Yang, Y. (2017). Development of an NRCS curve number global dataset using the latest geospatial remote sensing data for worldwide hydrologic applications. *Remote Sensing Letters*, 8(6), 528-536.